

# Leveraging the Power of Mass Spectrometry-Based Proteomics to Reveal Novel Biological Insights

Adam Ashley Dowle

PhD by Publication

University of York

Chemistry

July 2016

# Abstract

Mass spectrometry-based proteomics has the potential to offer new qualitative and quantitative insights into a wide array of biological questions. The work presented in this thesis, including the associated submitted papers, evidences my contribution to this field. Paper 1 demonstrates the successful identification of post translational modifications imparted upon histones following DNA damage. Papers 2-4 employ a label-free quantification approach to study the immunomodulation molecules endemic parasites use to impair host resistance. Paper 5 uses an isobaric tagging approach to quantify changes in immune response when dendritic cells are exposed to medically important antigens. Paper 6 displays the power of mass spectrometry in identifying a novel co-factor, which could not be identified by X-ray crystallography. In paper 7 mass spectrometry is used to show that the remaining unidentified enzyme in the morphinan biosynthetic pathway is expressed as a fusion protein composed of two distinct enzymes. These studies are all linked in their reliance on my use of mass spectrometry-based proteomics to make original contributions to knowledge and understanding within their fields.



# List of Contents

Abstract	- Page 2
List of Contents	- Page 3
List of Tables	- Page 6
List of Figures	- Page 7
Acknowledgements	- Page 9
Declarations	- Page 10
Integrative Chapter	- Page 12
Introduction	- Page 12
Proteomics background	- Page 13
Paper 1 - Dynamics of plant histone modifications in response to DNA damage	- Page 16
Paper 2 - The secretome of the filarial parasite, <i>Brugia malayi</i> : proteomic profile of adult excretory-secretory products	- Page 28
Paper 3 – Proteomic analysis of secretory products from the model gastrointestinal nematode <i>Heligmosomoides polygyrus</i> reveals dominance of Venom Allergen-Like (VAL) proteins	- Page 28
Paper 4 – Secretion of Protective Antigens by Tissue-Stage Nematode Larvae Revealed by Proteomic Analysis and Vaccination-Induced Sterile Immunity	- Page 28
Paper 5 – Plasma membrane proteomes of differentially matured dendritic cells identified by LC-MS/MS combined with iTRAQ labelling	- Page 35
Paper 6 – S-adenosyl-S-carboxymethyl-L-homocysteine: a novel co-factor found in the putative tRNA modifying enzyme CmoA	- Page 46
Paper 7 – Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein	- Page 53
Conclusion	- Page 62
Appendices	- Page 63
Appendix 1A. Submitted paper 1 – Dynamics of plant histone modifications in response to DNA damage	- Page 63

Appendix 1B. Declaration from Georgina Drury attesting to my contribution to paper 1	- Page 73
Appendix 1C. Declaration from Jerry Thomas attesting to my contribution to paper 1	- Page 75
Appendix 2A. Submitted paper 2 – The secretome of the filarial parasite, <i>Brugia malayi</i> : proteomic profile of adult excretory-secretory products	- Page 77
Appendix 2B. Declaration from James Hewitson attesting my contribution to paper 2	- Page 92
Appendix 3A. Submitted paper 3 – Proteomic analysis of secretory products from the model gastrointestinal nematode <i>Heligmosomoides polygyrus</i> reveals dominance of venom allergen-like (VAL) proteins.	- Page 94
Appendix 3B. Declaration from James Hewitson attesting to my contribution to paper 3	- Page 117
Appendix 4A. Submitted paper 4 – Secretion of protective antigens by tissue-stage nematode larvae revealed by proteomic analysis and vaccination-induced sterile immunity	- Page 119
Appendix 4B. Declaration from James Hewitson attesting to my contribution to paper 4	- Page 135
Appendix 5A. Submitted paper 5 – Plasma membrane proteomes of differentially matured dendritic cells identified by LC-MS/MS combined with iTRAQ labelling	- Page 137
Appendix 5B. Declaration from Stephanie Ferret-Bernard attesting to my contribution to paper 5	- Page 149
Appendix 5C. Declaration from Jerry Thomas attesting my contribution to paper 5	- Page 151
Appendix 6A. Submitted paper 6 – S-Adenosyl-S-carboxymethyl-L-homocysteine: a novel cofactor found in the putative tRNA-modifying enzyme CmoA	- Page 153
Appendix 6B. Declaration from Robert Byrne attesting to my contribution to paper 6	- Page 163
Appendix 7A. Submitted paper 7 - Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein	- Page 165

Appendix 7B. Declaration from Thilo Winzer attesting my contribution to paper 7	- Page 169
Appendix 7C. Declaration from Jerry Thomas attesting to my contribution to paper 7	- Page 171
Abbreviations	- Page 173
References	- Page 177

# List of Tables

Table 1. Table of acetylated and methylated peptides identified in histones H3 and H4 after plants received 160 Gy dose of ionising radiation.	- Page 20
Table 2. Relative quantification of proteins common to HES and HEx determined by relative empAI values.	- Page 31
Table 3. MED vs LPS or SEA labelling design. Each experiment number derived from a separate biological preparation.	- Page 37
Table 4. Proteins upregulated in SEA-DCs vs MED-DCs. Bold indicates $\log_{10}$ ratios greater than 0.1.	- Page 39
Table 5. Proteins down regulated in SEA-DCs vs MED-DCs. Bold indicates $\log_{10}$ ratios greater than 0.1.	- Page 40
Table 6. Proteins upregulated in LPS-DCs vs MED-DCs. Bold indicates $\log_{10}$ ratios greater than 0.1.	- Page 41
Table 7. Proteins down regulated in LPS-DCs vs MED-DCs. Bold indicates $\log_{10}$ ratios greater than 0.1.	- Page 41
Table 8. Proteins differently regulated in SEA-DCs vs LPS-DCs. Bold indicates $\log_{10}$ ratios greater than 0.1.	- Page 42
Table 9. Validation of iTRAQ results using flow cytometry.	- Page 44
Table 10. SmartFormula3D™ results table generated from $m/z$ values observed upon qCID fragmentation of ion at $m/z$ 433.1333.	- Page 49
Table 11. Comparison of sequence coverage, spectral counts and unique peptide sequences obtained in the 15 fractions excised from in-gel fractionated high-morphine cultivar protein extracts.	- Page 58

# List of Figures

- Figure 1. Sequence coverage achieved following trypsin digestion of Coomassie-stained bands. - Page 16
- Figure 2. Theoretical peptides generated on Arg-C digestion of histones H3 (top) and H4 (bottom). - Page 17
- Figure 3. Sequence coverage achieved after Arg-C digestion of histones H3 (top) and H4 (bottom). - Page 18
- Figure 4. Abundance of positively charged residues in histones H3 (top) and H4 (bottom). - Page 18
- Figure 5. Sequence coverage achieved after Arg-C digestion and propionylation of histones H3 (top) and H4 (bottom). - Page 19
- Figure 6A-D. Spectral annotations showing identified mono- and di-acetylations in histone H3 peptides - Page 21
- Figure 7A-D. Spectral annotations showing identified acetylations in histone H4 peptides - Page 23
- Figure 7E-H. Spectral annotations showing identified acetylations in histone H4 peptides - Page 24
- Figure 8A-B. Spectral annotations showing identified methylations in histone H3 peptides. - Page 25
- Figure 9. Representation of the acetylation forms of the H4 peptide GKGGKGLGKGGAKR - Page 26
- Figure 10. Venn diagram comparing the count of unique and overlapping proteins identified from somatic (*BmA*) and secreted-excreted (*BES*) fraction of *Brugia malayi*. - Page 33
- Figure 11. Venn diagram comparing the count of unique and overlapping proteins identified from secreted-excreted L4 larval (*L4*), adult (*HES*) and egg released material (*ERM*) from *Heligmosomoides polygyrus*. - Page 34
- Figure 12. Classification of identified proteins by GO terms, showing proteins assigned to the plasma membrane to be the most commonly identified in all samples. - Page 36

Figure 13. Classification of identified proteins by GO terms after iTRAQ labelling.	- Page 37
Figure 14. Whetton plot, showing protein level $\log_{10}$ fold change vs number of peptides identified for all technical replicates in the six iTRAQ analyses.	- Page 38
Figure 15. Biosynthetic pathway of uridine to uridine-5-oxyacetic acid conversion in bacterial tRNA.	- Page 46
Figure 16. FTICR-MS spectrum of purified CmoA.	- Page 47
Figure 17. FTICR-mass spectrum of purified CmoA, with instrument tuning to favour observation of low $m/z$ signal.	- Page 48
Figure 18. qCID FTICR-mass spectrum of precursor $m/z$ 443.1333.	- Page 49
Figure 19. Illustration of fragmentation positions in <i>S</i> -adenosyl- <i>S</i> -carboxymethyl-L-homocysteine, showing masses and formulae consistent with values measured by FTICR-MS.	- Page 50
Figure 20. Hypothesised roles of cofactor <i>S</i> -adenosyl- <i>S</i> -carboxymethyl-L-homocysteine in the modification of 5-methoxyuridine.	- Page 51
Figure 21. Coomassie-stained gel image resulting from 2D-PAGE separation of proteins from <i>Papaver somniferum</i> stems.	- Page 54
Figure 22. 1D-PAGE of high-morphine cultivar protein extracts.	- Page 55
Figure 23. Sequence coverage for endogenous STORR fusion protein excised from between 95-105 kDa in gel (fraction 4, figure 22).	- Page 56
Figure 24. Annotated product ion spectrum for peptide identification IKPCVQSAASER (linker peptide pos 559-579).	- Page 56
Figure 25. Annotated product ion spectrum for peptide identification DMESSGVPVITLGSGK (linker peptide pos 580-595).	- Page 57
Figure 26. Quantification of STORR peptides through 1D gel.	- Page 59
Figure 27. 1D-PAGE of replicate high-morphine cultivar protein extracts (a-c), and $^{15}\text{N}$ -labeled recombinant STORR (d).	- Page 60
Figure 28. Quantification of STORR peptides through gel.	- Page 60

# Acknowledgements

I thank Prof. Jane Thomas-Oates who has supervised me through the writing of the thesis for her much valued guidance and support. I thank the co-authors on the papers submitted without whom the thesis would not be possible. I thank my former line manager Dr. Jerry Thomas and long standing colleague Dr. David Ashford, both of whom I have learnt a great deal from. I also thank Jerry and David for their encouragement to begin the process of submitting for a PhD. I thank my current line manager Dr. Tony Larson for his ongoing support. I thank all the members of Jane's lab and the Bioscience Technology facility for their exceptional collegiality and the knowledge that they have kindly shared. I thank the Department of Chemistry Graduate office and the University of York for allowing me the opportunity to submit this work. I also thank my wife Clare for her support and encouragement throughout.

# Declaration

This thesis is presented for the award of PhD by publication. I declare that I meet the requirements to submit for the award of the degree of PhD by Publication. Specifically: I have held a Master's degree (MChem) for over three years; I am a current member of staff of the University of York and have been in continuous employment with the University for over three years; all work submitted has been carried out whilst employed by the University and published during the past 8 years.

The thesis comprises an integrative chapter and the seven published papers listed below.

1. Drury GE, Dowle AA, Ashford DA, Waterworth WM, Thomas J, West CE. Dynamics of plant histone modifications in response to DNA damage. *Biochem J.* **2012** Aug 1; 445(3): 393-401
2. Hewitson JP, Harcus YM, Curwen RS, Dowle AA, Atmadja AK, Ashton PD, Wilson A, Maizels RM. The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory-secretory products. *Mol Biochem Parasitol.* **2008** Jul; 160(1): 8-21.
3. Hewitson JP, Harcus Y, Murray J, van Agtmaal M, Filbey KJ, Grainger JR, Bridgett S, Blaxter ML, Ashton PD, Ashford DA, Curwen RS, Wilson RA, Dowle AA, Maizels RM. Proteomic analysis of secretory products from the model gastrointestinal nematode *Heligmosomoides polygyrus* reveals dominance of venom allergen-like (VAL) proteins. *J Proteomics.* **2011** Aug 24; 74(9): 1573-94.
4. Hewitson JP, Ivens AC, Harcus Y, Filbey KJ, McSorley HJ, Murray J, Bridgett S, Ashford D, Dowle AA, Maizels RM. Secretion of protective antigens by tissue-stage nematode larvae revealed by proteomic analysis and vaccination-induced sterile immunity. *PLoS Pathog.* **2013** Aug; 9(8): e1003492
5. Ferret-Bernard S, Castro-Borges W, Dowle AA, Sanin DE, Cook PC, Turner JD, MacDonald AS, Thomas JR, Mountford AP. Plasma membrane proteomes of differentially matured dendritic cells identified by LC-MS/MS combined with iTRAQ labelling. *J Proteomics.* **2012** Jan 4; 75(3): 938-48.
6. Byrne RT, Whelan F, Aller P, Bird LE, Dowle A, Lobley CM, Reddivari Y, Nettleship JE, Owens RJ, Antson AA, Waterman DG. S-Adenosyl-S-carboxymethyl-L-homocysteine: a novel cofactor found in the putative tRNA-modifying enzyme CmoA. *Acta Crystallogr D Biol Crystallogr.* **2013** Jun; 69(Pt6): 1090-8.
7. Winzer T, Kern M, King AJ, Larson TR, Teodor RI, Donniger SL, Li Y, Dowle AA, Cartwright J, Bates R, Ashford D, Thomas J, Walker C, Bowser TA, Graham IA. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science.* **2015** Jul 17; 349(6245): 309-12.

The submitted papers are presented in appendices 1A-7A.



All papers submitted have been published and are co-authored. Signed declarations are provided from all first authors specifying my contribution and the circumstances in which the work was carried out. Declarations are also provided by my line manager at the time (Jerry Thomas), for papers on which he was a co-author. The signed declarations are in the appendix.

This work has not previously been presented for an award at this or any other University. All sources are acknowledged as references.

# Integrative Chapter

## Introduction

The seven publications presented cover a range of biological investigations yet all are bound by their reliance on mass spectrometry-based proteomics – the aspect that I performed. In each case my proteomic analyses were vital to the study.

In paper 1 I identified novel acetylated proteoforms of *Arabidopsis* histones using LC-MS/MS, database searching and manual validation of product ion spectra.<sup>1</sup>

In papers 2-4 I employed emPAI quantification to provide qualitative differences between somatic and excreted proteomes of medically important parasites.<sup>2-4</sup>

Paper 5 describes my use of iTRAQ quantification to compare proteomes of dendritic cells exposed to antigens known to illicit different immunological responses.<sup>5</sup>

Paper 6 presents my identification of the novel protein co-factor S-adenosyl-S-carboxymethyl-L-homocysteine using Fourier transform ion cyclotron resonance mass spectrometry.<sup>6</sup>

In paper 7 I used a combination of <sup>15</sup>N metabolic labelling and 1D-PAGE to prove that the final enzyme in the morphinan biosynthetic pathway is naturally expressed as a fusion of two distinct enzymes.<sup>7</sup>

This integrative chapter details my contribution to each study, setting the background and incorporating additional material that was not included in the publications for reasons of focus or brevity. Collectively the work evidences the power of mass spectrometry based-proteomics to reveal novel biological insights.

Proteomics has progressed significantly in the eight years spanning the publications. I therefore present my work in the context of the knowledge and technology of the time but also reflect in the present day, discussing limitations and detailing how the studies could be improved. This chapter offers discussion of key concepts, technologies, analysis regimes and limitations of proteomics.

## Proteomics background

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), where proteins are separated in gel by both pI and physical length<sup>8,9</sup> was the key enabling technology for proteomics. The technique was widely applied following first use of the term 'proteomics' in Wilcox's 1994 PhD thesis and subsequent 1996 paper.<sup>10</sup> Resolving by pI is achieved by applying proteins to an immobilized pH gradient in the presence of electric potential; proteins migrate across the gradient until they reach their pI and become uncharged. In the second dimension, proteins coated with SDS are loaded onto a polyacrylamide gel. Electric potential is applied causing proteins to migrate, with speed dictated by molecular length: shorter sequences have lower frictional forces and so migrate further per unit time.<sup>11</sup> Protein length is considered proportional to molecular mass. Proteins can be visualised by staining to produce gel spots.<sup>12,13</sup> If quantitative stains are used, intensity is proportional to protein amount.<sup>14</sup> Although quantification is implicit, identification is not; migration of standards can be compared for simple mixtures or known targets probed by western blotting<sup>15</sup> but for unknowns direct identification is intractable.

Mass spectrometry coupled with genomic sequencing proved pivotal to delivering identification with biologically amenable protein amounts.<sup>16</sup> Peptide mass fingerprinting (PMF) was developed as a rapid and robust protein identification technique. Protein spots excised from the gel are treated with endoprotease of defined specificity to generate peptides.<sup>17</sup> For instance, trypsin cleaves amide bonds C-terminal to arginine or lysine residues, provided they are not N-terminal to proline.<sup>18</sup> Expected cleavage positions are used to calculate theoretical peptide sequences and so masses for proteins of known sequence. Predicted masses can be compared *in silico* for protein identification based on numbers of spectrum features matching theoretical values.<sup>19</sup> Software packages automate searches and provide estimations of significance based on these features.<sup>20-22</sup>

If precursors are isolated and fragmented in the mass spectrometer, information on the constituent amino acid sequence can be revealed.<sup>23,24</sup> CID fragmentation, which is most frequently used for peptides,<sup>25,26</sup> produces *b*- and *y*-ions as dominant features, resulting from cleavage of the amide bond<sup>27,28</sup> (ion-type nomenclature was proposed by Roepstorff<sup>29</sup> and developed by Biemann).<sup>30</sup> Fragmentation position is explained by the mobile proton hypothesis,<sup>31-33</sup> which states protonation occurs on basic residues before transferring to the amide carbonyl, driving fragmentation. Analogous to PMF, peptide product ion spectra can be searched against calculated *m/z* values of fragment ions to provide peptide matches, from which peptide matches and thus protein identification can be inferred.<sup>34</sup> Unlike PMF where a single diagnostic value is obtained per peptide (its *m/z*), fragmentation produces multiple data points increasing identification specificity.

If complex mixtures of peptides are analysed by mass spectrometry only the most abundant/amenable components are sampled. Less abundant or responsive analytes can be

missed due to limitations in mass resolution, isolation resolution, acquisition speed/duty cycle and ion response/suppression. Fractionation ameliorates these effects, affording deeper penetration into proteome space.<sup>35</sup> Peptide-level fractionation can remove the necessity for 2D-PAGE, which is limited by irreproducibility and incompatibility with extreme pH or low solubility proteins.<sup>36</sup>

LC-MS/MS has overtaken 2D-PAGE as the method of choice in most proteomic laboratories. Liquid chromatography can be performed on- or off-line with mass spectrometry. In off-line approaches collected fractions are analysed sequentially making them amenable to MALDI, where sample is mixed with matrix prior to ionisation.<sup>37</sup> On-line approaches couple elution to the mass spectrometer increasing throughput.<sup>38</sup> Electrospray ionisation is the typical on-line ionisation mode.<sup>39</sup> Chromatography for proteomics mostly uses reversed-phase C<sub>18</sub> columns, separating peptides by relative affinities for stationary and mobile phases resulting from different hydrophobicities. Peptides are loaded in aqueous-dominated mobile phase before elution over a gradient of increasing organic content.<sup>38,40</sup> Orthogonal multi-dimensional chromatography can provide further separation, with SCX, HILIC and high pH reversed phase C<sub>18</sub> all used.<sup>41-43</sup>

Mass spectrometry is not implicitly quantitative; analyte response is dependent on amount but also less predictable factors including physico-chemical analyte properties, ion suppression effects and solvent composition.<sup>44</sup> These factors must be compensated for quantification. For absolute quantification, an internal standard (IS) is added in known amount and analyte response relative to the IS measured.<sup>45</sup> The IS should be as similar as possible to the analyte and yet distinguishable, either chromatographically or in mass; isotopically heavy analogues are ideal. Isotopically labeled samples are combined before LC-MS/MS, with imparted mass differences providing the link from spectra to provenance. Although isotopically different, labeled peptides are otherwise chemically identical, promoting equal response.<sup>45</sup> *In vivo* labeling strategies such as SILAC<sup>46</sup> and <sup>15</sup>N metabolic labeling<sup>47</sup> offer the least technical variance because samples are combined earliest.<sup>48</sup> For individual peptides, heavy variants can be synthesised.<sup>49</sup> Absolute quantification of a range of proteins is possible using a QconCAT approach, whereby quantotypic peptides are selected as representatives for multiple proteins and combined to generate a concatamer gene. The gene is transfected into an expression system to produce a concatamer protein containing the quantotypic peptides.<sup>50</sup> Expression in heavy growth medium can produce heavy labelled peptides. QconCAT protein is added to sample before enzymatic digestion, allowing potential quantification of all proteins with peptides common to the QconCAT.

While *in vivo* labeling techniques offer lowest technical variance, they have reduced applicability than *in vitro* approaches.<sup>51</sup> A variety of *in vitro* labelling techniques is available including methods employing well characterized, low cost chemical reactions<sup>52,53</sup> and more elaborate commercial approaches facilitating higher multiplexing.<sup>54-60</sup> Two of the most frequently used are SCIEX's

iTRAQ<sup>54,55</sup> and Thermo's TMT,<sup>56-60</sup> each of which add isobaric tags to free amines at peptide N-termini and lysine residues. Mass spectra of differentially-labelled samples appear identical as tags are isobaric, but unique  $m/z$  reporter ions are released upon fragmentation, enabling quantification.<sup>54-60</sup>

Quantification is also possible without labelling. Label-free approaches do not combine samples before LC-MS/MS acquisition; instead samples are analysed successively and relative responses extracted from datasets.<sup>61</sup> Response can be measured by precursor and/or product ion chromatogram peak areas,<sup>62</sup> which necessitates software to align chromatograms and overlay peptide identifications.<sup>63</sup> Alternatively, response can be inferred from spectral counts;<sup>64</sup> for product ion spectra triggered by intensity, the greater the relative abundance of analyte, the more frequently it is selected for fragmentation.<sup>63-67</sup>

## Paper 1 - Dynamics of plant histone modifications in response to DNA damage

Paper 1 presents my identification of acetylated and methylated forms of *Arabidopsis thaliana* histones after plant exposure to ionising radiation. The study began in 2010 when I was contacted by University of Leeds postdoctoral researcher Georgina Drury (GD), who had purified a set of wild type *Arabidopsis thaliana* proteins and was aiming to sequence these. GD provided the samples as two Coomassie stained bands and I performed in-gel digestion with trypsin following reduction with dithioerythritol and S-carbamidomethylation with iodoacetamide. I analysed the resulting peptide mixtures by LC-MS/MS using a maXis qTOF mass spectrometer and peak lists produced were searched against the *Arabidopsis thaliana* subset of the UniProt database using the Mascot search program. Both bands matched histone sequences, one H3 and the other H4. Coverage for H3 was reasonable at 52% but only two peptides were identified in H4 giving a sequence coverage of just 24% (figure 1).

Band 1 – H3, 52% sequence coverage

```
1  MARTKQTARK STGGKAPRKQ LATKAARKSA PATGGVKKPH RFRPGTVALR
51  EIRKYQKSTE LLIRKLPFQR LVREIAQDFK TDLRFQSSAV AALQEAAEAY
101 LVGLFEDTNL CAIHAKRVTI MPKDIQLARR IRGERA
```

Band 2 – H4, 24% sequence coverage

```
1  MSGRGKGGKG LGKGGAKRHR KVLRDNIQGI TKPAIRRLAR RGGVKRISGL
51  IYEETRGLVK IFLENVIRDA VTYTEHARRK TVTAMDVVYA LKRQGRITLYG
101 FGG
```

**Figure 1. Sequence coverage achieved following trypsin digestion of Coomassie-stained bands. Bands 1 (top) and 2 (bottom) were identified as histones H3 and H4, respectively. Amino acid sequence from peptides identified through Mascot database searching of product ion spectra highlighted in red.**

Examining the sequences, I observed the high frequency of R and K (trypsin cleavage sites) would produce many peptides below optimum mass for analysis. I advised GD that more could be done to improve coverage. GD explained that histone proteins were her intended target and the aim was to characterise the post translational modifications (PTMs) of these proteins following treatment with ionising radiation.

Post translational modifications (PTMs) are chemical alterations to amino acids in proteins that occur during or after protein synthesis. PTMs can play crucial roles in biological processes such as cell signalling and disease response.<sup>68-70</sup> Mass spectrometry is well placed to identify PTMs by measuring mass difference caused by modification, ideally at the individual amino acid level in

product ion mass spectra.<sup>71</sup> PTM identification can be built into database searching workflows by appending the mass increment to a dataset against which the experimental data are searched. PTM searches are usually run as variable modifications, so that both modified and unmodified variants are considered.<sup>72</sup>

Characterising post translational modifications of histones is important because of their effects on the structure and function of chromatin.<sup>73</sup> The chromatin complex packages DNA in eukaryotic nuclei and is vital for DNA stability, mitosis and genome maintenance.<sup>74</sup> Ionising radiation is a known effector of DNA damage, causing double strand breaks. Acetylation of histones H3 and H4 has been demonstrated as an epigenetic marker of chromatin remodelling in response to DNA damage in both mammals and yeast<sup>75,76</sup>. This was the target of GD's work. Acetylation of the N-terminal region of histones is especially crucial as the point of interaction with DNA. Unfortunately, in my initial LC-MS/MS analysis the N-terminal region of H4 was not observed.

I explained that sequence coverage may be improved by changing to an alternative protease that would produce more N-terminal peptides of amenable mass. *In-silico* digestions suggested that endoprotease Arg-C should produce suitable peptides (figure 2).

H3, 88% theoretical coverage

```
1  MARTKQTARK STGGKAPRKQ LATKAARKSA PATGGVKKPH RFRPGTVALR
51  EIRKYQKSTE LLIRKLPFQR LVREIAQDFK TDLRFQSSAV AALQEAAEAY
101 LVGLFEDTNL CAIHAKRVTI MPKDIQLARR IRGERA
```

H4, 85% theoretical coverage

```
1  MSGRGKGGKG LGKGGAKRHR KVLRDNIQGI TKPAIRRLAR RGVVKRISGL
51  IYEETRGVLK IFLENVIRDA VTYTEHARRK TVTAMDVVYA LKRQGRTLYG
101  FGG
```

**Figure 2. Theoretical peptides generated on Arg-C digestion of histones H3 (top) and H4 (bottom). Amino acid sequence from in-silico digest products between 500 and 4,000 Da highlighted in blue.**

GD provided further samples and I repeated LC-MS/MS analysis following Arg-C digestion. Knowing now the proteins contained no cysteine residues, the reduction and alkylation steps were omitted. Close to 50% coverage was achieved for both histones (figure 3). Unfortunately, I still did not observe the N-terminus of H3.

### H3, 54% sequence coverage

1 MARTKQTARK STGGKAPRKQ LATKAARKSA PATGGVKKPH RFRPGTVALR  
51 EIRKYQKSTE LLIRKLPFQR LVREIAQDFK TDLRFQSSAV AALQEAAEAY  
101 LVGLFEDTNL CAIHAKRVTI MPKDIQLARR IRGERA

### H4, 49% sequence coverage

1 MSGRGKGGKG LGKGGAKRHR KVLRDNIQGI TKPAIRRLAR RGGVKRISGL  
51 IYEETRGLVK IFLENVIRDA VTYTEHARRK TVTAMDVVYA LKRQGRITLYG  
101 FGG

**Figure 3. Sequence coverage achieved after Arg-C digestion of histones H3 (top) and H4 (bottom). Amino acid sequence from peptides identified through Mascot database searching of product ion spectra highlighted in red.**

Arg-C digestion improved peptide sizes, but many contained a large number of positively charged amino acids (figure 4) imparting hydrophilicity, and reducing binding to C<sub>18</sub> media used in the analytical and trap columns. If peptides do not bind to the trap they flow to waste and are not sampled.

### H3

1 MARTKQTARK STGGKAPRKQ LATKAARKSA PATGGVKKPH RFRPGTVALR  
51 EIRKYQKSTE LLIRKLPFQR LVREIAQDFK TDLRFQSSAV AALQEAAEAY  
101 LVGLFEDTNL CAIHAKRVTI MPKDIQLARR IRGERA

### H4

1 MSGRGKGGKG LGKGGAKRHR KVLRDNIQGI TKPAIRRLAR RGGVKRISGL  
51 IYEETRGLVK IFLENVIRDA VTYTEHARRK TVTAMDVVYA LKRQGRITLYG  
101 FGG

**Figure 4. Abundance of positively charged residues in histones H3 (top) and H4 (bottom). Positive amino acids arginine, histidine and lysine highlighted in green.**

I considered loading directly onto the analytical column. However, in my literature research, I found reports that chemical modification of peptides by propionylation can reduce hydrophilicity, leading to increased coverage in histones.<sup>77</sup> I thus propionylated the remainder of the Arg-C digests using the protocol developed by Syka.<sup>77</sup> LC-MS/MS analysis of the propionylated peptides produced very good N-terminal coverage (figure 5).



H3, 47% sequence coverage

```
1  MARTKQTARK  STGGKAPRKQ  LATKAARKSA  PATGGVKKPH  RFRPGTVALR
51  EIRKYQKSTE  LLIRKLPFQR  LVREIAQDFK  TDLRFQSSAV  AALQEAAEAY
101 LVGLFEDTNL  CAIHAKRVTI  MPKDIQLARR  IRGERA
```

H4, 69% sequence coverage

```
1  MSGRGKGGKG  LGKGGAKRHR  KVLRDNIQGI  TKPAIRRLAR  RGGVKRISGL
51  IYEETRGLK  IFLENVIRDA  VTYTEHARRK  TVTAMDVVYA  LKRQGRPLYG
101  FGG
```

**Figure 5. Sequence coverage achieved after Arg-C digestion and propionylation of histones H3 (top) and H4 (bottom). Amino acid sequence from peptides identified through Mascot database searching of product ion spectra highlighted in red.**

Knowing suitable N-terminal coverage could be obtained, GD then extracted histones from plants exposed to 160 Gy of ionising radiation.

My analysis of histones from treated plants was conducted using Arg-C and propionylation (paper 1). Database searching was performed to consider five variable modifications of lysine: acetylation, methylation, di-methylation, tri-methylation and propionylation. I assumed tri-methylation would be difficult to distinguish from acetylation because mass additions imparted are separated by just 36 mDa: acetylation 42.0106 Da; tri-methylation 42.0470 Da. However, accuracy of the maXis instrument was sufficient to exclude tri-methylation for all identified peptides, with maximum precursor mass error being 8 mDa. I additionally observed a number of lysine residues were not propionylated.

Although Mascot is capable of identifying modified peptides and provides expect scores showing the confidence of the match, validation of site assignment is crucial. Without interrogation, incorrect site assignment may be accepted. For instance, in a peptide with two potentially modifiable amino acids located close together, ion series from residues preceding and succeeding these may localise the modification to those positions but the data may not allow distinction between them. In such cases the Mascot score will be equal for both possibilities but one will be listed top and displayed in the search result. Incorrect assignments can also occur where ions from different series coincidentally have the same  $m/z$  as potentially diagnostic ions. It is important to consider the possibility of chimeric spectra containing product ions from two or more co-selected precursors – positional isomers are especially prone to co-selection as their hydrophobicities and hence retention times are often similar. For large datasets, setting global stringencies for site assignment, such as delta score<sup>78-80</sup> can be valuable but, where tractable, few

methodologies beat manual validation. For these data I used Mascot for initial identifications then located the corresponding MS<sup>2</sup> spectrum using Bruker's data analysis software. I looked for crucial ions for expected sequences that could distinguish positional assignments. I identified 14 different acetylated or methylated peptides (table 1). Post validation, spectra were annotated for publication using ProPhoSI,<sup>81</sup> which provides more aesthetic images than manually annotated DataAnalysis spectra and better labelling than Mascot for less frequently observed ion types, including loss of water (0-ions) and loss of ammonia (\*-ions). My summary of the site assignments is in figures 6-8. Suffixes ac, me and me<sub>2</sub> are used to denote acetylated, methylated and dimethylated lysines respectively.

Histone	Peptide	Modified sites	Expect Value	Mass error (mDa)
H3	K <sup>9</sup> STGGKAPR <sup>17</sup>	K14	2.0×10 <sup>-7</sup>	5
H3	K <sup>9</sup> STGGKAPR <sup>17</sup>	K9, K14	1.6×10 <sup>-4</sup>	6
H3	K <sup>18</sup> QLATKAAR <sup>26</sup>	K23	3.3×10 <sup>-5</sup>	4
H3	K <sup>18</sup> QLATKAAR <sup>26</sup>	K18, K23	9.8×10 <sup>-6</sup>	8
H3	K <sup>9</sup> STGGKAPR <sup>17</sup>	K9	2.8×10 <sup>-4</sup>	5
H3	K <sup>9</sup> STGGKAPR <sup>17</sup>	K9, K14	7.8×10 <sup>-3</sup>	5
H4	G <sup>4</sup> KGGKGLGKGGAKR <sup>17</sup>	K8	5.0×10 <sup>-5</sup>	1
H4	G <sup>4</sup> KGGKGLGKGGAKR <sup>17</sup>	K16	1.5×10 <sup>-7</sup>	6
H4	G <sup>4</sup> KGGKGLGKGGAKR <sup>17</sup>	K5, K16	1.2×10 <sup>-5</sup>	2
H4	G <sup>4</sup> KGGKGLGKGGAKR <sup>17</sup>	K8, K16	1.3×10 <sup>-6</sup>	7
H4	G <sup>4</sup> KGGKGLGKGGAKR <sup>17</sup>	K12, K16	6.4×10 <sup>-6</sup>	5
H4	G <sup>4</sup> KGGKGLGKGGAKR <sup>17</sup>	K5, K12, K16	5.2×10 <sup>-8</sup>	<1
H4	G <sup>4</sup> KGGKGLGKGGAKR <sup>17</sup>	K8, K12, K16	3.0×10 <sup>-4</sup>	2
H4	G <sup>4</sup> KGGKGLGKGGAKR <sup>17</sup>	K5, K8, K12, K16	2.7×10 <sup>-6</sup>	3

**Table 1. Table of acetylated and methylated peptides identified in histones H3 and H4 after plants received 160 Gy dose of ionising radiation. Acetylated lysine residues coloured red, methylated lysine residues are coloured green and dimethylated lysine residues are orange. Identification performed using the Mascot search engine following LC-MS/MS acquisition of propionylated endoprotease Arg-C-derived peptides. Expect values taken from Mascot results.**

H3 Acetylated

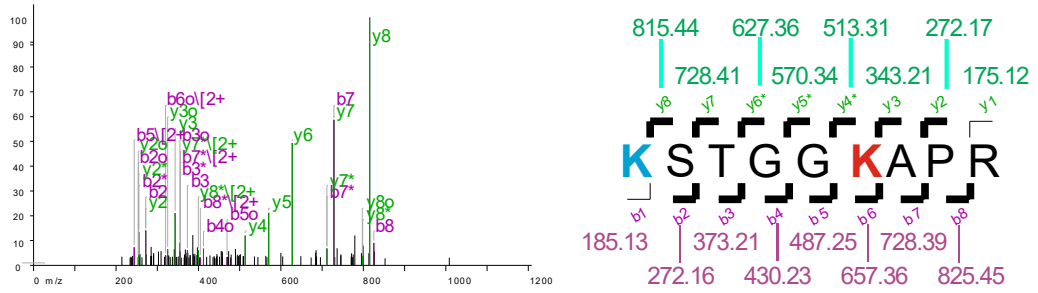
K<sup>9</sup>STGGKacAPR<sup>17</sup> – strong *b*- and *y*-ion series observed, lacking only the lowest *m/z* ion in each series (Figure 6A).

K<sup>9</sup>acSTGGKacAPR<sup>17</sup> – with only two lysine residues in the peptide and two acetylations determined from the precursor mass no positional ambiguity exists. Identification is strengthened by strong *b*- and *y*-ion series, with the latter missing only the *y*<sub>1</sub>-ion (Figure 6B).

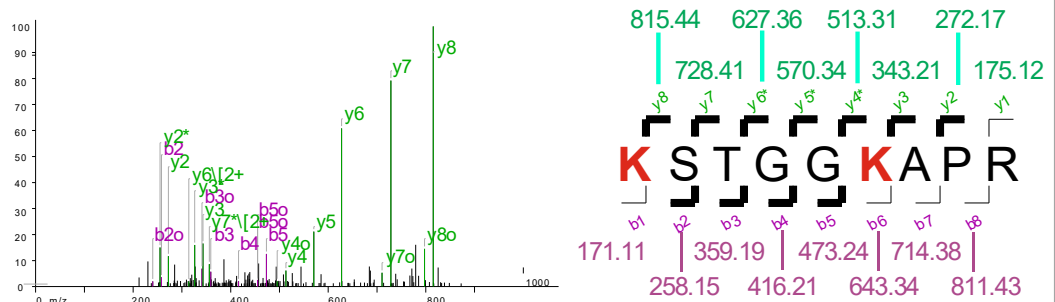
K<sup>18</sup>QLATKacAAR<sup>26</sup> – continuous *y*-ion series lacking only *y*<sub>1</sub>. *b*<sub>2</sub> and *b*<sub>3</sub> also demonstrate acetylation is not on the N-terminal lysine K<sup>18</sup> (Figure 6C).

K<sup>18</sup>acQLATKacAAR<sup>26</sup> – precursor mass shows acetylation of both lysine residues. Intense and continuous *y*-ion series supports identification (Figure 6D).

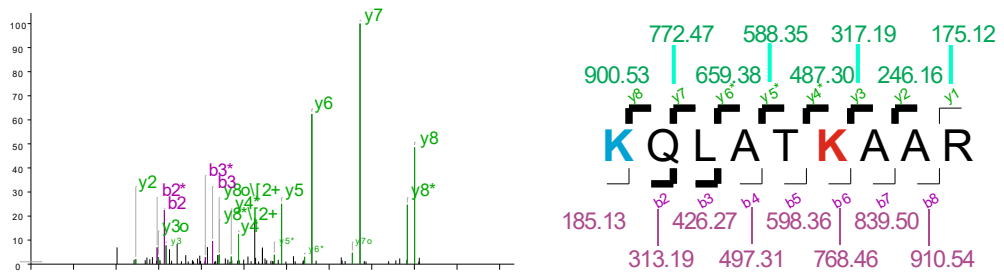
### A. K<sup>9</sup>STGGKacAPR<sup>17</sup>



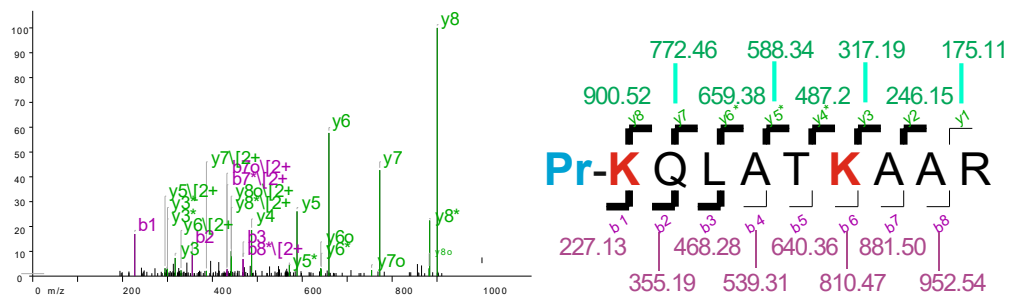
### B. K<sup>9</sup>acSTGGKacAPR<sup>17</sup>



### C. K<sup>18</sup>QLATKacAAR<sup>26</sup>



### D. K<sup>18</sup>acQLATKacAAR<sup>26</sup>



**Figure 6A-D. Spectral annotations showing identified mono- and di-acetylations in histone H3 peptides. Tandem mass spectra were re-drawn and annotated using ProPhoSI to show matches to predicted b- and y-ions. Bold black lines in amino acid sequences indicate that a b- (purple) or y-ion (green) is observable in the spectrum. Acetylated lysines are coloured red, propionylated lysines blue and N-terminal propionylation labelled Pr and coloured blue.**

#### H4 Acetylated

G<sup>4</sup>KGGKacGLGKGGAKR<sup>17</sup> – acetylation at K<sup>8</sup> shown by mass difference of 170.1 *m/z* units between ions *b*<sub>4</sub> and *b*<sub>5</sub>, and between *y*<sub>9</sub> and *y*<sub>10</sub> (Figure 7A).

G<sup>4</sup>KGGKGLGKGGAKacR<sup>17</sup> – *y*<sub>2</sub> localises mono-acetylation to K<sup>16</sup>. Further evidence is provided by *b*<sub>10</sub> and *b*<sub>13</sub> that constrain addition of acetyl to residues G<sup>14</sup>AK<sup>16</sup>, in which only modification of lysine is rational (Figure 7B).

G<sup>4</sup>KacGGKGLGKGGAKacR<sup>17</sup> – evidence of acetylation at K<sub>5</sub> derived from *b*<sub>2</sub>, *b*<sub>3</sub>, *y*<sub>10</sub>, *y*<sub>11</sub> and *y*<sub>12</sub>, which all require acetylated K<sup>5</sup> to reconcile either their own mass (*b*-ions) or the delta-mass to the precursor (*y*-ions). Acetylation at K<sup>16</sup> produced *y*<sub>2</sub>, *y*<sub>4</sub> and *y*<sub>5</sub> ions consistent with mono-acetylation C-terminal to A<sup>15</sup> (Figure 7C).

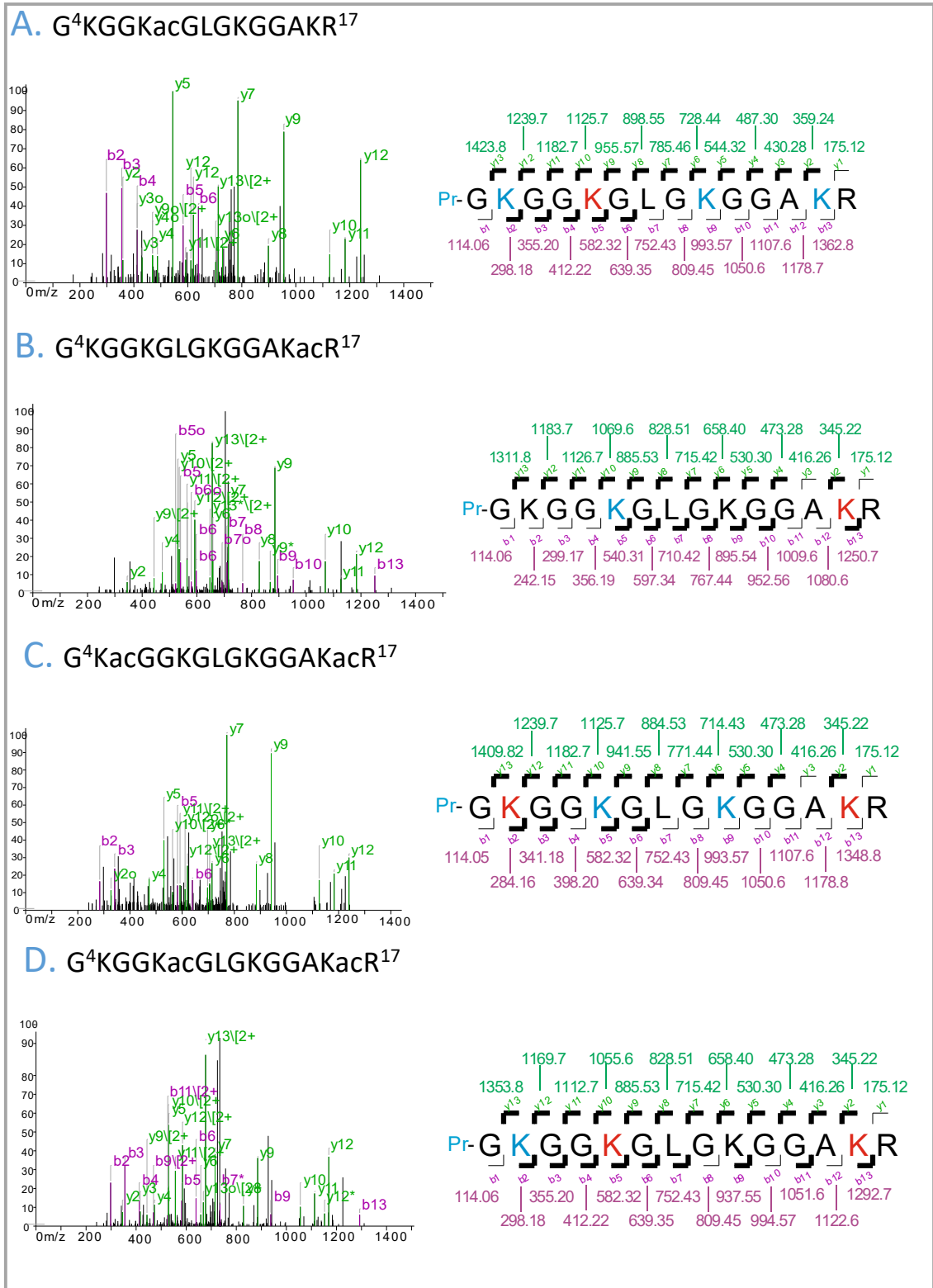
G<sup>4</sup>KGGKacGLGKGGAKacR<sup>17</sup> – strong *b*- and *y*-ion series localises positions of acetylation. Specifically, *b*<sub>4</sub>, *b*<sub>5</sub>, *y*<sub>9</sub> and *y*<sub>10</sub> for K<sup>8</sup> and *y*<sub>2</sub>, *y*<sub>11</sub> and *y*<sub>13</sub> for K<sup>16</sup> (Figure 7D).

G<sup>4</sup>KGGKGLGKacGGAKacR<sup>17</sup> – intense *y*-ions at positions *y*<sub>5</sub> and *y*<sub>9</sub> localise acetylation to K<sup>12</sup> and K<sup>16</sup>, respectively, with additional lower intensity ions supporting the assignment (Figure 7E).

G<sup>4</sup>KacGGKGLGKacGGAKacR<sup>17</sup> – continuous ion series between *b*<sub>2</sub>-*b*<sub>7</sub> and *y*<sub>7</sub>-*y*<sub>12</sub> consistent with K<sup>8</sup> as the only non-acetylated lysine residue in the peptide. Although further evidence is provided by the product ions, the remaining three lysine residues must be acetylated to reconcile the precursor mass (Figure 7F).

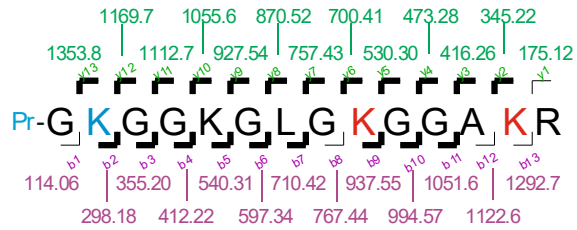
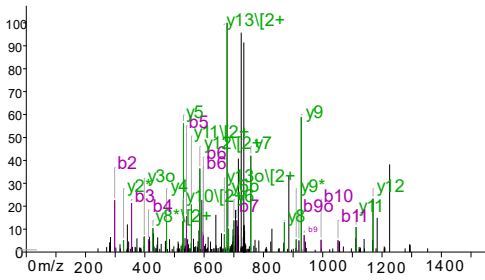
G<sup>4</sup>KGGKacGLGKacGGAKacR<sup>17</sup> – intense *y*<sub>5</sub>, *y*<sub>7</sub>, *y*<sub>9</sub> and *y*<sub>11</sub>-ions localise the three acetylation positions predicted by the precursor mass to the lysines at positions K<sup>8</sup>, K<sup>12</sup> and K<sup>16</sup>. The ions *b*<sub>2</sub>-*b*<sub>4</sub>, excludes the possibility of acetylation at K<sup>5</sup> (Figure 7G).

G<sup>4</sup>KacGGKacGLGKacGGAKacR<sup>17</sup> – precursor mass indicates tetra-acetylation, with no positions of ambiguity. Expected product ions validate identification (Figure 7H).

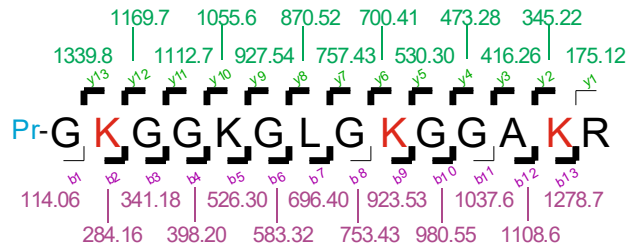
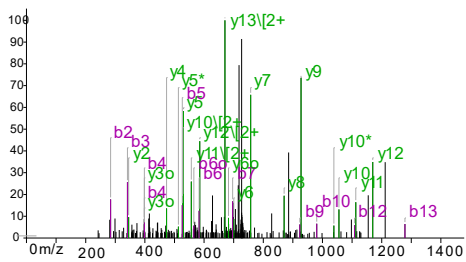


**Figure 7A-D. Spectral annotations showing identified acetylations in histone H4 peptides. Tandem mass spectra were re-drawn and annotated using ProPhoSI to show matches to predicted b- and y-ions. Bold black lines in amino acid sequences indicate that a b- (purple) or y-ion (green) is observable in the spectrum. Acetylated lysines are coloured red, propionylated lysines blue and N-terminal propionylation labelled Pr and coloured blue.**

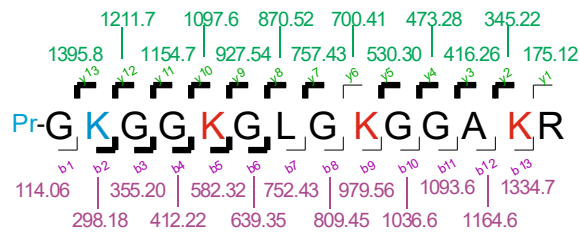
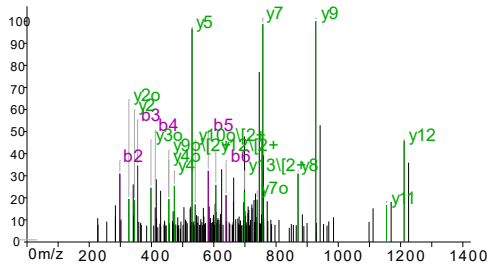
### E. G<sup>4</sup>KGGKGLGKacGGAKacR<sup>17</sup>



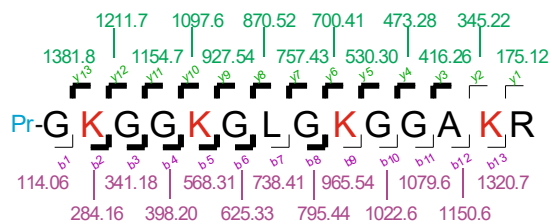
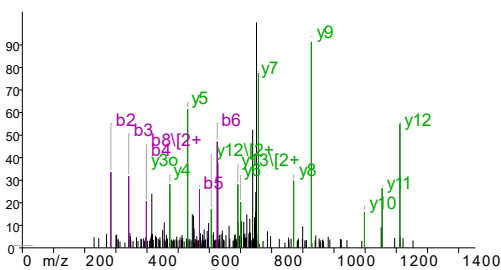
### F. G<sup>4</sup>KacGGKGLGKacGGAKacR<sup>17</sup>



### G. G<sup>4</sup>KGGKacGLGKacGGAKacR<sup>17</sup>



### H. G<sup>4</sup>KacGGKacGLGKacGGAKacR<sup>17</sup>

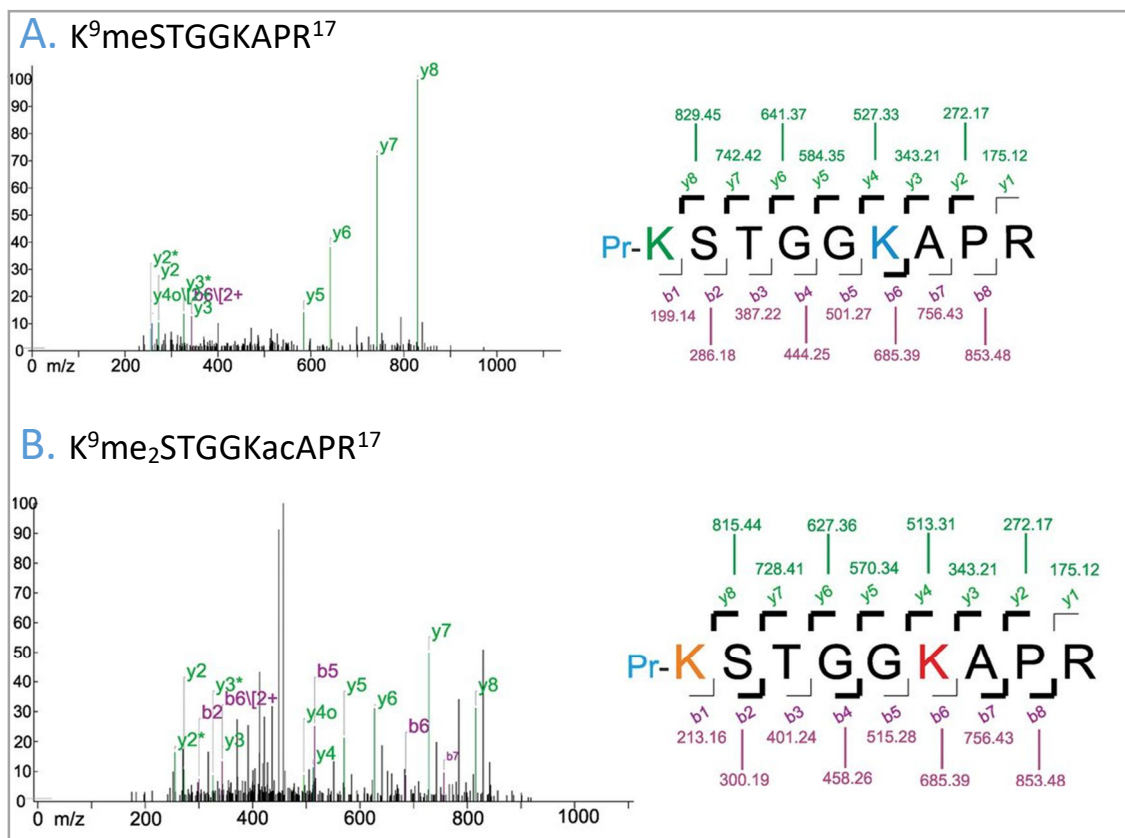


**Figure 7E-H. Spectral annotations showing identified acetylations in histone H4 peptides. Tandem mass spectra were re-drawn and annotated using ProPhoSI to show matches to predicted b- and y-ions. Bold black lines in amino acid sequences indicate that a b- (purple) or y-ion (green) is observable in the spectrum. Acetylated lysines are coloured red, propionylated lysines blue and N-terminal propionylation labelled Pr and coloured blue.**

### H3 Methylated

$K^9\text{meSTGGKAPR}^{17}$  – continuous  $y$ -ion series from  $y_3$  demonstrates both propionylation and methylation of the N-terminal lysine residue  $K^9$ , when compared with precursor mass (Figure 8A).

$K^9\text{me}_2\text{STGGKacAPR}^{17}$  – mass difference between precursor and  $y$ -ion series N-terminal to  $y_3$  infers propionylation and di-methylation at N-terminal lysine,  $K^9$ . There is a significant amount of unannotated ion intensity in the spectrum; although some compromise in fragment ion intensity can be attributed to the extent of the unfragmented precursor ion, additional positional forms cannot be excluded (Figure 8B). However, no additional forms could be inferred from unassigned signals, which are likely background noise resulting from selection of low intensity precursor.



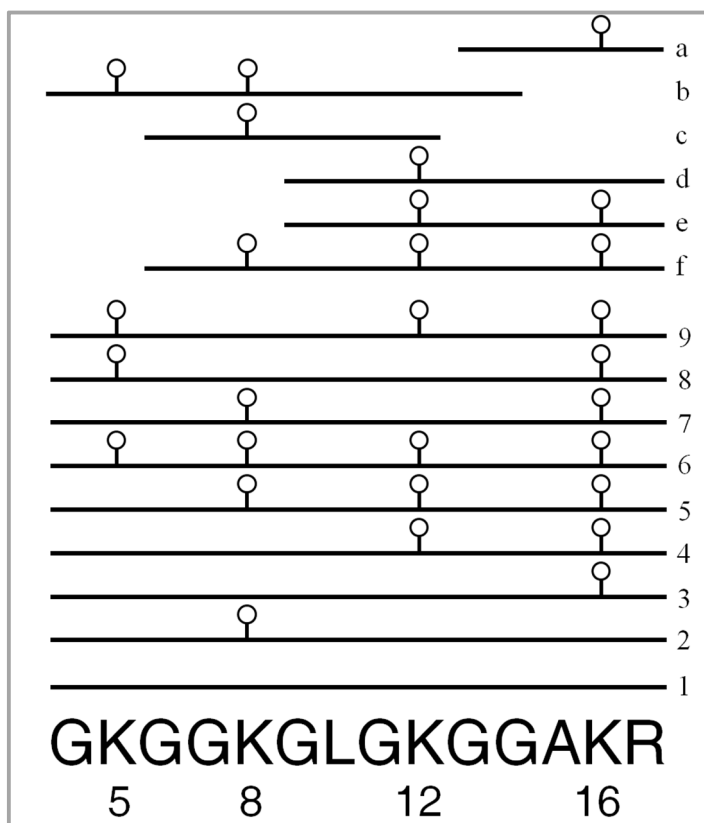
**Figure 8A-B. Spectral annotations showing identified methylations in histone H3 peptides. Tandem mass spectra were re-drawn and annotated using ProPhoSI to show matches to predicted b- and y-ions. Bold black lines in amino acid sequences indicate that a b- (purple) or y-ion (green) is observable in the spectrum. Acetylated lysines are coloured red, mono-methylated lysines green, dimethylated lysines orange, propionylated lysines blue and N-terminal propionylation labelled Pr and coloured blue.**

For the two acetylated peptides identified from H3,  $K^9\text{STGGKAPR}^{17}$  and  $K^{18}\text{QLATKAAR}^{26}$ , mono-acetylated versions are observed acetylated exclusively at positions  $K^{14}$  or  $K^{23}$ . Di-acetylation was observed for each peptide but no evidence was obtained for acetylation of  $K^9$  in the absence of  $K^{14}$  acetylation or modification of  $K^{18}$  without  $K^{23}$  acetylation, suggesting that acetylation occurs at  $K^{14}$  prior to  $K^9$ , and at  $K^{23}$  before  $K^{18}$ . Although absence of observation does not exclude presence, it is difficult to explain why the MS response of either positional mono-acetylated isoform would change significantly compared with its isoform. Inspection of product spectra from the mono-acetylated forms of  $K^9\text{STGGKAPR}^{17}$  (figure 7A) and  $K^{18}\text{QLATKAAR}^{26}$  (figure 7C) contain good

evidence for the positions assigned, with little ion intensity remaining that could suggest a second position of mono-acetylation from a co-selected precursor.

My results correlate with previous work on untreated *Arabidopsis* H3 that reported a relatively high frequency of detection for acetylation at K<sup>14</sup> and no observation of the mono-acetylated peptide K<sup>9</sup>acSTGGKAPR<sup>17,82</sup>. That study also showed the di-acetylated peptides K<sup>9</sup>acSTGGKacAPR<sup>17</sup> and K<sup>18</sup>acQLATKacAAR<sup>26</sup> to be some of the least observed forms in untreated plants, potentially suggesting these forms may be enriched by ionising radiation.

I identified the H4 peptide G<sup>4</sup>KGGKGLGKGGAKR<sup>17</sup> in nine different proteoforms, from unmodified to tetra-acetylated, with two positional isoforms for mono- and tri-acetylation and three distinct positional isoforms for di-acetylation (figure 8A-H). Previous studies have identified only the tetra-acetylated form in untreated higher plants, although work by Zhang has identified modification at each of the positions in shorter peptide sequences.<sup>83</sup> Figure 5 presents the H4 acetylation positions identified in my data compared with those of Zhang.



**Figure 9. Representation of the acetylation forms of the H4 peptide GKGGKGLGKGGAKR. Forms observed following treatment of plants with ionizing radiation (1-9) compared with forms observed by Zhang et al (a-f).<sup>83</sup> Acetylation of lysine residues 5, 8, 12 and 16 is indicated by open circles.**

In Zhang's analysis, shorter trypsin-generated peptide sequences yield less information on the concomitant nature of acetylation at positions K<sup>5</sup>, K<sup>8</sup>, K<sup>12</sup> and K<sup>16</sup>, although all bar one of their identifications can be considered consistent with my results. The exception is identification of di-acetylated peptide form G<sup>4</sup>KacGGKacGLGKGG<sup>14</sup> (figure 9). In my analysis, the variant where K<sup>5</sup> and K<sup>8</sup> are acetylated while K<sup>12</sup> remains unmodified is not observed. Other studies have



suggested that acetylation of H4 occurs first at position K<sup>16</sup> and proceeds sequentially towards the N-terminus.<sup>84</sup> My data show that, although acetylation at K<sup>16</sup> is frequently observed, multiple acetylation forms of K<sup>5</sup>, K<sup>8</sup> and K<sup>12</sup> are present, which are inconsistent with the suggested sequential addition from the C-terminal direction, demonstrating identification of novel proteoforms.

With the acetylation pattern characterised from irradiated plants, the next step would have been to quantitatively compare untreated and treated plants but this was not pursued because of limited grant funds. As a lower cost option GD used antibodies to probe relative abundances of identified acetylation forms from control and irradiated plants as described in paper 1.

Had further funds been available, either intensity-based label-free or isotopic labelling strategies could be considered. However, as positional isoforms are known and likely to elute at similar retention times, I would avoid isobaric labelling options such as iTRAQ or TMT where quantification is derived from common MS<sup>2</sup> reporter ions. Deconvolution of overlapping mass spectral peak areas from closely eluting peptides is more tractable than deconvoluting reporter ion intensities from chimeric product ion spectra, especially where dynamic exclusion is used and so very few MS<sup>2</sup> spectra would be sampled across the elution profile.

Any approach should also determine the relative amounts of common unmodified peptides to distinguish changes in protein expression from site occupancy. It would also be valuable to include samples exposed to different levels of irradiation to track trends of de/acetylation with increasing exposure.

Ultimately, any bottom-up approach where protease-derived peptides are analysed is limited by peptide length. True proteoform analysis would require a top-down alternative where intact proteins are analysed mass spectrometrically.<sup>85</sup> For example, I have demonstrated the peptide G<sup>4</sup>KGGKGLGKGGAKR<sup>17</sup> exists in a minimum of nine proteoforms, but this amino acid sequence makes up only 14% of the protein sequence. Using a bottom-up approach, how the different proteoforms relate to the remaining 86% of the sequence is unknown. Top-down approaches have recently been demonstrated capable of identifying over 1,000 proteoforms from complex mixtures using on-line HPLC separation.<sup>86,87</sup>

Although potential exists for expansion, my proteomics analysis successfully identified novel acetylation forms and paved the way for the first observation of H4 de-acetylation at positions K<sup>12</sup> and K<sup>16</sup> following double strand breakage in irradiated plants. The paper added to the public knowledge base on histone modifications and has been well received with six citations since publication,<sup>88-93</sup> including a PLOS Biology review by Karen T. Smith and Jerry L. Workman entitled Chromatin Proteins: Key Responders to Stress.<sup>88</sup>

Paper 2 – The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory-secretory products

Paper 3 – Proteomic analysis of secretory products from the model gastrointestinal nematode *Heligmosomoides polygyrus* reveals dominance of Venom Allergen-Like (VAL) proteins

Paper 4 – Secretion of protective antigens by tissue-stage nematode larvae revealed by proteomic analysis and vaccination-induced sterile immunity

Papers 2-4 detail my mass spectrometry-based analysis of helminth excretory-secretory proteomes (ES). The work is also unified by collaboration with University of Edinburgh postdoctoral researcher James Hewitson (JH). In each study, extracted proteins were prepared by JH for me to analyse by mass spectrometry.

Helminths impose a significant burden upon humanity – around one billion people are estimated to be infected.<sup>94</sup> These parasites are remarkably tolerant of the immune response and known to persist for long periods in immunocompetent patients.<sup>95</sup> It is rational that parasite excretions are significant to immunomodulation<sup>96</sup> and it has been demonstrated that such excreted products can elicit sterilizing immunity in hosts exposed prior to infection.<sup>97</sup> Elucidation of ES proteins can therefore present a valuable list of targets for therapeutic intervention. The work I present studies the murine intestinal nematode parasite *Heligmosomoides polygyrus*<sup>98</sup> and human filarial nematode *Brugia malayi*.<sup>99</sup>

My collaboration with JH began in 2007 when he asked me to identify proteins from 2D-PAGE spots derived from a comparison of *Brugia malayi* ES (BES) and somatic protein extract (BmA). On discussion, JH had hoped to see deeper into the proteome than he was achieving by 2D-PAGE and I suggested LC-MS/MS may facilitate this. JH provided me with extracted BES and BmA proteins and I analysed these by LC-MS/MS as detailed in paper 2. As hoped, LC-MS/MS gave deeper penetration, increasing the number of proteins identified in BES from 23 to 82 (paper 2, table 1 and figure 2D). Crucially, all 23 proteins identified using 2D-PAGE were present in my LC-MS/MS dataset. By modern standards 82 proteins is modest, and can be attributed to the LC-MALDI-MS/MS technique employed. The key drawback is that fractionation is limited to the number of positions on the MALDI target plate. In contrast, on-line LC-ESI-MS/MS is only limited by the resolution of the column and the speed of the mass spectrometer.

I took the same approach in paper 3 where *Heligmosomoides polygyrus* ES was analysed by both 2D-PAGE and LC-MS/MS. Here, the number of identifications increased from 53 to 374 with LC-ESI-MS/MS using a maXis qTOF mass spectrometer (paper 3, tables 1 and 2). Again, all 2D-PAGE identified proteins were present in my LC-MS/MS dataset.

Although the mass spectrometry-safe silver staining used for the 2D-PAGE work is not truly quantitative, relative intensity between spots can provide an approximation of inter-protein amount.<sup>100</sup> I hoped a similar quantitative estimation could be obtained from the LC-MS/MS data and advised using Mascot-calculated emPAI values to achieve this.<sup>101</sup> emPAI values are calculated from the relative proportion of observed ( $N_{obs}$ ) to theoretically observable ( $N_{theo}$ ) peptides (equation 1).

$$(1) \quad emPAI = 10^{(N_{obs}/N_{theo})} - 1$$

**Equation 1. Calculation of emPAI**

Considering spectral count as a proportion of theoretically observable peptides with respect to mass range and hydrophobicity is intended to compensate for bias towards longer proteins with unmodified spectral counting.

It should be noted there are imperfections in Mascot's calculation of emPAI values. Unlike Ishihama's implementation, individual peptides are not classed as observable based on measured masses and hydrophobicities. Instead  $N_{theo}$  is approximated from protein molecular mass and the total number of enzymatic cleavage positions in the database. Mascot also makes no estimation of hydrophobic retention compatibility.

Even disregarding Mascot's approximations, spectral counting is now considered less accurate than intensity based approaches. A better alternative is the more recently developed iBAQ, which calculates protein abundance by dividing total protein MS<sup>1</sup> response by  $N_{theo}$ .<sup>102</sup> Another viable alternative is the 'top three' method,<sup>103</sup> which compensates for varying peptide number and unpredictable response by considering only intensity of the three most highly responding peptides per protein.

Intensity-based approaches outperform spectral counting because they are less reliant on stochastic product ion selection. The amplitude of measurement for intensity-based measurement is also less discrete; i.e. for spectral count to increase, an additional peptide must be sampled. In contrast, continuity of amplitude in intensity-based approaches is limited only by detector measurement. Such fine measurements are especially crucial where small numbers of peptides are identified for a given protein.

In addition to using emPAI for intra-sample, inter-protein quantification I also used it for inter-sample relative protein quantification of the same protein between ES and somatic extracts. The same limitations exist here vs intensity-based approaches, although when comparing the same protein inter-sample, the imperfections of Mascot's derivation apply consistently between the two comparators and effectively cancel out.

Compositional differences measured in my LC-MS/MS datasets were highly divergent for ES vs somatic extracts. For instance, in paper 3 where HES was compared with soluble somatic protein extract (HEx), only 28% (104) of the proteins I identified in HES were also seen in HEx.<sup>3</sup> The relative abundance of the 104 common proteins as estimated by relative emPAI values was also highly divergent (table 2).

Protein	HES/HEX emPAI	Protein	HES/HEX emPAI
VAL-2.1 (isotig03106)	70	Nucleoside diphosphate kinase (isotig05236)	0.80
TTR-1 Transthyretin-related (isotig04612)	34	TTR-6 Transthyretin related (isotig05212)	0.70
VAL-1.3 (isotig06456)	33	Rab GDP dissociation inhibitor (isotig06129)	0.70
VAL-7.5 (isotig01526)	32	Chondroitin family member (isotig06559)	0.70
VAL-1.1 (isotig06320)	23	Vitellogenin (contig00199)	0.70
NAS-3.1 Astacin protease (isotig01799)	17	Myoglobin (isotig11742)	0.70
VAL-3.1 (isotig05790)	15	Enolase (isotig06965)	0.70
VAL-7.1 (isotig01524)	14	Transaldolase (isotig07176)	0.70
NSN-1 Novel secreted protein, No SP (isotig18405)	13	Vitellogenin (contig00203)	0.70
Astacin protease family member (isotig01791)	12	Aldolase (isotig04899)	0.70
CSP-4 Conserved secreted protein with SP (isotig13290)	11	Vitellogenin (fragment) (sing00711)	0.60
VAL-4 (isotig10387)	9.7	Myoglobin (isotig05169)	0.60
Aspartyl protease (necepsin) (isotig06497)	6.5	Myoglobin (isotig04274)	0.60
Kunitz inhibitor (isotig01217)	6.3	TTR-11 Transthyretin related (isotig04258)	0.60
VAL-12 (isotig06637)	5.4	Ribosomal protein 60S L40 / Ubiquitin (isotig13199)	0.60
Astacin protease family member (fragment) (isotig01219)	5.2	Serpin (isotig02225)	0.60
Superoxide dismutase (isotig09104)	4.7	Vitellogenin (contig00212)	0.50
C-type lectin mannose receptor-like (isotig00496)	4.2	Vitellogenin (contig00211)	0.50
MFH-1 Ascaris MFP2b homologue (isotig07283)	4.0	Chondroitin family member (isotig04525)	0.50
Cystatin (isotig02700)	3.4	Fatty acid/retinol binding protein (isotig11475)	0.50
CSP-2 Conserved secreted protein with SP (isotig09016)	2.9	Fatty acid/retinol binding protein (fragment) (sing12372)	0.40
Motile sperm domain containing protein (isotig16097)	2.3	Phosphatidylethanolamine-binding protein (isotig05126)	0.40
Galectin (isotig09082)	2.2	Independent phosphoglycerate mutase (isotig01699)	0.40
TIL domain-containing protein (isotig00779)	2.1	Vitellogenin (Hp_ADY_001G11)	0.40
Arginine kinase (isotig05009)	2.0	Cyclophilin (isotig11124)	0.40
Motile sperm domain containing protein (isotig16375)	1.7	Thioredoxin peroxidase (isotig05188)	0.30
ERM family member (isotig02815)	1.6	Enolase (fragment) (isotig18996)	0.30
Cyclophilin (isotig09088)	1.6	Glutathione S-transferase (isotig11588)	0.30
Actin depolymerising factor (isotig07522)	1.5	HSP70 (fragment) (isotig13132)	0.30
Macrophage migration inhibitory factor (isotig14093)	1.5	Myoglobin (isotig11356)	0.30
TTR-5 Transthyretin-like (isotig13207)	1.5	Calreticulin (Hpb-CRT)	0.30
MFH-3 Ascaris MFP2b homologue (isotig07980)	1.4	Lipocalin domain-containing protein (isotig05316)	0.30
TTR-10 Transthyretin-like (isotig11792)	1.4	Myoglobin (isotig11666)	0.30
NSP Novel isotig12701 (isotig12701)	1.4	CSN-5 Conserved secreted protein, No SP (isotig03664)	0.30
Chondroitin family member (isotig01493)	1.3	Aldolase (isotig04915)	0.30
Ferritin (isotig15727)	1.0	Major sperm protein (isotig01565)	0.20
NSP-39 Novel secreted protein with SP (isotig16333)	1.0	HSP60 (isotig06093)	0.20
Vitellogenin (contig00207)	1.0	Piwi domain-containing protein (isotig04875)	0.20
Myoglobin (fragment) (isotig19571)	1.0	14-3-3 family member (isotig07975)	0.20
Calmodulin (isotig07710)	1.0	CSN-2 Conserved secreted protein, No SP (isotig04423)	0.20
Cathepsin-B like cysteine protease (isotig07017)	1.0	Actin (isotig08340)	0.20
Galectin (isotig09713)	1.0	Cytochrome C (isotig11144)	0.20
Kunitz inhibitor (isotig02721)	1.0	Actin (isotig01640)	0.10
Kunitz inhibitor (isotig02378)	1.0	Eukaryotic translation elongation factor 2 (isotig05363)	0.10
ML-domain containing protein (isotig10098)	1.0	Protein disulfide isomerase (isotig01916)	0.10
Motile sperm domain containing protein (isotig16212)	1.0	Eukaryotic translation elongation factor 1A (isotig02523)	0.10
NSP-59 Novel secreted protein with SP (isotig13319)	1.0	Protein disulfide isomerase (isotig05787)	0.10
Triose phosphate isomerase (isotig05016)	1.0	HSP70 (isotig07134)	0.10
VAL-15 (isotig02149)	1.0	Glutathione S-transferase (isotig10292)	0.10
Chondroitin family member (isotig08745)	1.0	Malate dehydrogenase (isotig05076)	0.10
NPA-1 Nematode polyprotein allergen (isotig02438)	0.90	Ribosomal protein 60S P0 (isotig01358)	0.10
Vitellogenin (contig00471)	0.80	Retinol binding protein (Hpb-RBP)	0.00

**Table 2. Relative quantification of proteins common to HES and HEX determined by relative emPAI values. Proteins measured as upregulated in HES annotated with red emPAI values and upregulation in HEX in blue. Intensity of colour is proportional to magnitude of change.**

Biological replicates were not analysed due to limited protein availability from parasitic systems. Lack of replication was commonplace at the time but would be virtually unpublishable today. In this respect, a greater understanding of the statistical nature of data and resulting robustness is possibly the most impactful change proteomics has made in the past decade.

Without replication it is impossible to determine if changes are significant or symptomatic of underlying variance. Technical replication (i.e. repeated protein extraction or LC-MS/MS acquisitions from a common sample) only measures variance in the step replicated and is typically lower than biological variation.<sup>104</sup> Technical replication can be useful for assessment of variance at different procedural positions, especially during method development.<sup>105</sup> Biological replication is key to showing differences as biologically consistent. Biological replicates encompass underlying technical variation meaning technical replication can often be omitted. Biological analyses must also contain appropriate controls to ensure that observations result from the stimulus applied rather than more generic pleotropic or stress responses.<sup>106</sup>

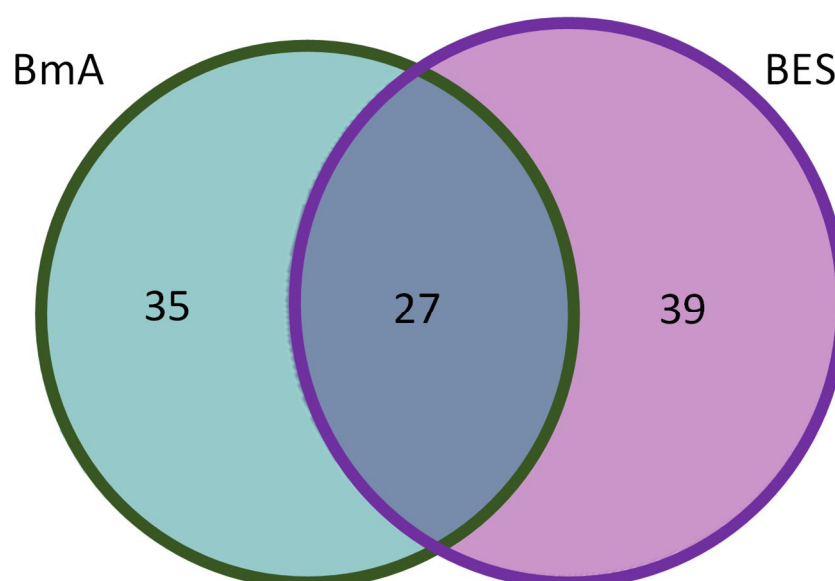
By current standards, the work presented in papers 2-4 should be viewed as qualitative rather than quantitative. However, my 'qualitative' mass spectrometry analysis was ultimately successful in suggesting gross differences in ES protein extracts that have been subsequently validated by orthogonal techniques or through repetition in similar studies.

HES was dominated by venom allergen/ancylostoma secreted (VAL) proteins – 14 of the top 20 and 26 of the top 100 proteins in HES were VALs when ranked by emPAI (paper 3). In contrast, only VAL 3.1 was present in the top 100 HEx emPAI-ranked proteins. GO analysis indicated translation as the most common biological process term for HEx, and structural component of ribosome the most common molecular function. In HES, proteolysis and metalloendopeptidase activity were the most common biological process and molecular function terms respectively. Proteolysis and peptidase activity may protect against host immune response by reducing hydrolytic activity towards the nematodes and responding with their own hydrolytic attack against mediators of immune response.

In paper 2 relative emPAI comparison of BES with somatic extract (BmA) estimates >10-fold enrichment of triose phosphate isomerase, leucyl aminopeptidase,  $\gamma$ -glutamyltranspeptidase and N-acetylglucosaminyltransferase in BES. Enrichment of peptidases is consistent with the HES results from paper 3. Several glycolytic enzymes were measured as highly abundant in BmA but only two are identified in BES: enolase and triose phosphate isomerase, the latter showing a 17-fold enrichment in BES (paper 2). At time of publication the role of triose phosphate isomerase at the helminth–host interface was unknown. However, my identification of triose phosphate isomerase in the LC-MS/MS data led to work by Hewitson *et. al.*<sup>107</sup> demonstrating that adult female *Brugia malayi* helminths release triose phosphate isomerase, promoting creation of an

ovum increasing the number offspring. Hence, antibody targeting and neutralisation of triose phosphate isomerase inhibits reproduction and reduces the microfilarial burden. Such an observation makes triose phosphate isomerase a target for immunological and pharmacological intervention against filarial infections.

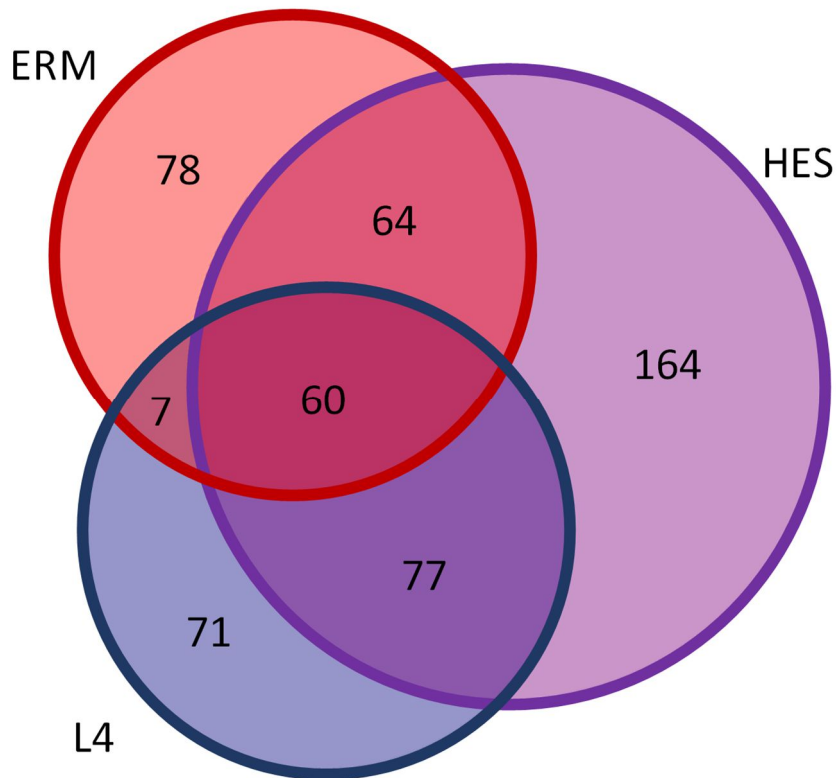
Further proteins, ranked highly by emPAI score in BES, were at undetectable levels in BmA. An indication of the disparity is presented in figure 10. It was noted that galectins were highly prominent in the list of BES protein identifications when ranked by emPAI. Again, the observation is common between BES and *Heligmosomoides polygyrus* ES studied in paper 2. A study by Kim *et. al.* published subsequent to the presented work has shown that exposure to an expressed homolog of galectin-9 prior to *Toxascaris leonina* infection can inhibit intestinal inflammation.<sup>108</sup> Such findings add further weight to targeting galectins for further study in parasitic systems.



**Figure 10. Venn diagram comparing the count of unique and overlapping proteins identified from somatic (BmA) and secreted-excreted (BES) fraction of *Brugia malayi*.**

Paper 4 is as an extension of paper 3 in which adult *Heligmosomoides polygyrus* ES is compared with larval stage four secreted analogue (L4), and egg-released material (ERM). L4 is a key infection stage where parasites migrate from submucosa to lumen. It would be advantageous to target antigens present at the larval stage with the hope of countering the infection before egg production and proliferation within the body.

Figure 11 illustrates the overlap of protein identifications between the HES, L4 and ERM extracts.



**Figure 11.** Venn diagram comparing the count of unique and overlapping proteins identified from secreted-excreted L4 larval (L4), adult (HES) and egg released material (ERM) from *Heligmosomoides polygyrus*.

Although VAL proteins are highly abundant in HES and L4, my emPAI comparison suggests there are differences in VAL composition. Notably, VAL-2 was measured as less abundant in L4 – VAL-2.1, 2.2 and 2.3 are ranked 11<sup>th</sup>, 1<sup>st</sup> and 41<sup>st</sup> respectively in adult ES but only VAL-2.3 was detected in L4 and was ranked 124<sup>th</sup>. JH validated VAL-2 quantification by ELISA, with the results showing consistent low abundance in L4 ES.

My LC-MS/MS data also indicated L4 ES contained high abundance acetylcholinesterases, lysozymes and apyrases. Subsequent work by JH, also presented in paper 4, showed antibodies extracted from mice inoculated with L4 and HES are antigenic to VAL-1 and acetylcholinesterase-1, suggesting that these extracts may be especially important in immune modulation.

Papers 2-4 have been extremely well received with over 250 citations collectively (PubMed). They led to deeper insights into helminth immunomodulation, including identification of new potential targets for therapeutic investigation, which have since been validated by subsequent investigations. In each case my LC-MS/MS identification and qualitative comparison was vital to the discovery of these targets and improving the understanding of the field.



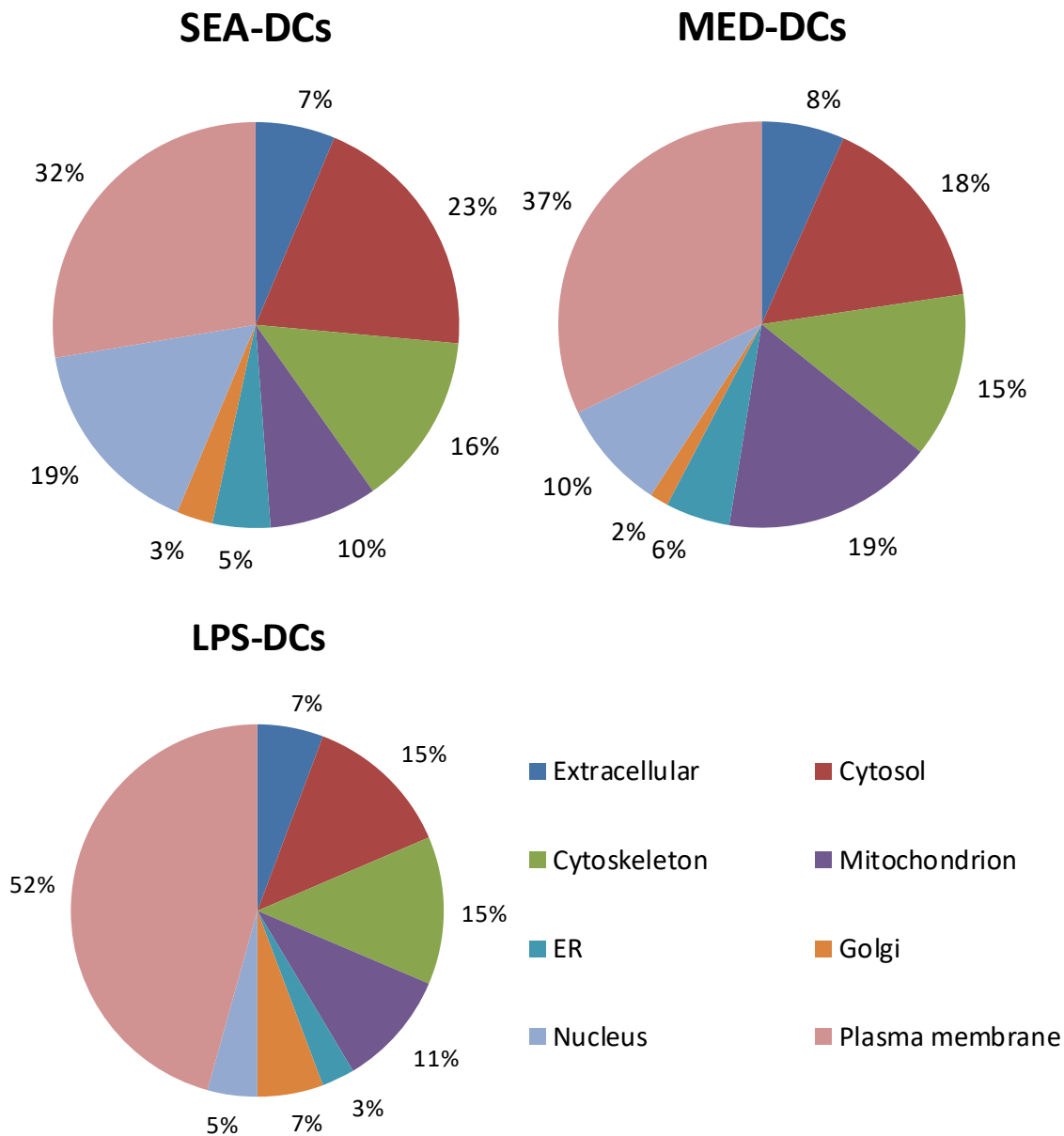
## Paper 5 – Plasma membrane proteomes of differentially matured dendritic cells identified by LC-MS/MS combined with iTRAQ labelling

Paper 5 highlights my use of isobaric tags for protein quantification applied in an immunological setting.

I was approached by University of York postdoctoral researcher Stéphanie Ferret-Bernard (S-FB), who wanted to compare relative proteomes of dendritic cells (DCs) exposed to bacterial LPS or schistosome egg antigen (SEA). LPS, which forms an outer layer on gram-negative bacteria<sup>109</sup> induces the inflammatory type 1 immune response (Th1).<sup>110</sup> In contrast, schistosomes, which are parasitic helminths, generate anti-inflammatory type 2 immune response (Th2). Th1 proceeds through recognition of antigens by DC surface proteins inducing a mature state where antigen is adhered to cell surface and trafficked for presentation to host immune system. At the time of study Th2 type response was much less understood than Th1 and although it had been shown to proceed via a non- or partially-matured DC profile, little was known of the molecular mediators.<sup>111</sup>

At time of publication, studies comparing proteomic changes in DCs upon exposure to helminth infection relied upon 2D-PAGE work-flows, limiting observation of membrane proteins.<sup>112</sup> Analysis of the less amenable hydrophobic membrane component was essential to provide full understanding of the system.

SF-B intended to prepare a centrifugal membrane preparation of DC cells exposed to either SEA (SEA-DCs) or LPS (LPS-DCs), with control samples exposed to culture medium (MED-DCs). Before proceeding to quantification we decided to test membrane enrichment. SF-B performed enrichment and I analysed the samples by LC-MALDI-MS/MS after tryptic digestion as detailed in the paper. The number of proteins identified varied between treatments with 119 proteins in MED-DCs, 149 proteins in SEA-treated and 61 proteins after LPS treatment. Gene ontology (GO) showed proteins attributed to plasma membrane were most frequently identified in each sample, although this was conspicuously different between the treatments; 32% of proteins were assigned as membrane in SEA-DCs and 52% in LPS-DCs (figure 12). Despite the differences, we decided to proceed to quantitative analysis.



**Figure 12. Classification of identified proteins by GO terms, showing proteins assigned to the plasma membrane to be the most commonly identified in all samples.**

At the time, the only mass spectrometry option available to me was the Applied Biosystems 4700 MALDI-TOF/TOF instrument. MALDI-MS response is strongly affected by crystallisation<sup>113</sup> making label-free comparison difficult, so I suggested an iTRAQ approach. iTRAQ relative peptide quantification is derived from reporter ions in product ion spectra of co-selected isobaric peptides, which are exposed to identical crystallisation and ionisation conditions. iTRAQ multiplexing also saves on instrument time.

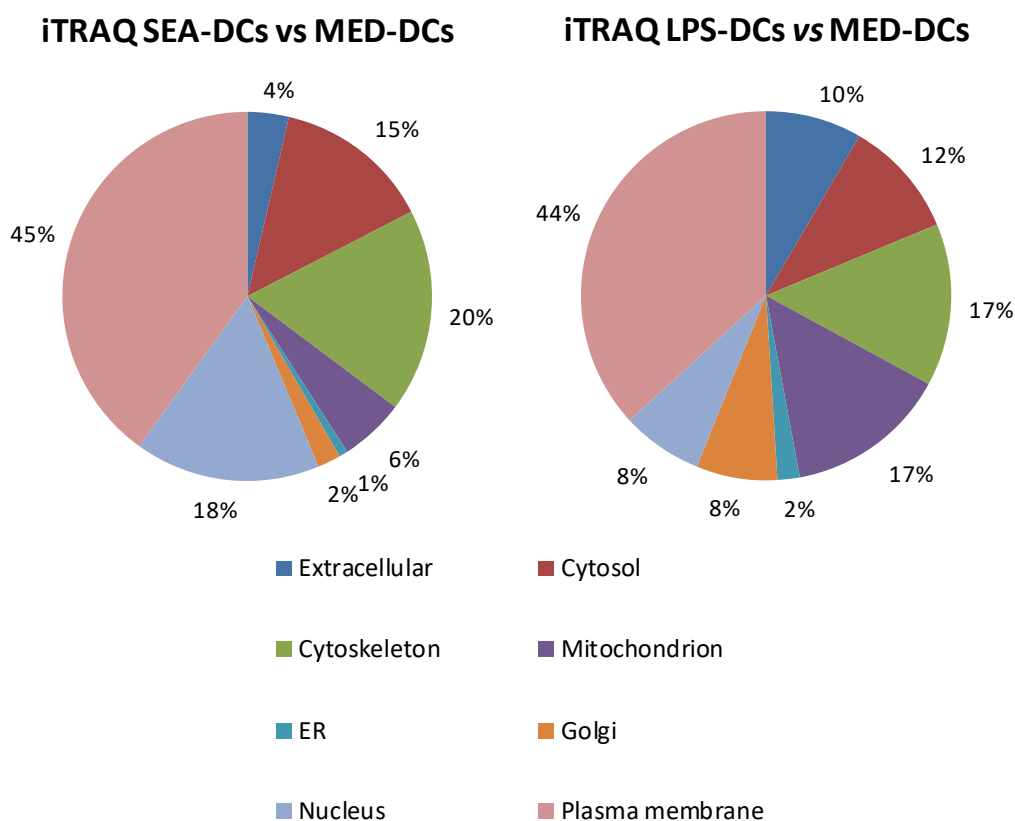
SF-B explained that LPS treatment would likely cause a more pronounced change than SEA and comparison of SEA vs MED control was equally if not more important than SEA vs LPS. I recommended SEA and LPS analyses be separated into individual comparisons vs MED control, reducing the chance of LPS high magnitude change swamping modest SEA change. We agreed on three biological replicates and to assess technical variance caused by digestion and labelling I

advised splitting the MED-DC samples into two aliquots before labelling. Table 3 details the labelling design.

Experiment \ Label	114	115	115
1	MED	LPS	MED
2	MED	SEA	MED
3	MED	LPS	MED
4	MED	SEA	MED
5	MED	LPS	MED
6	MED	SEA	MED

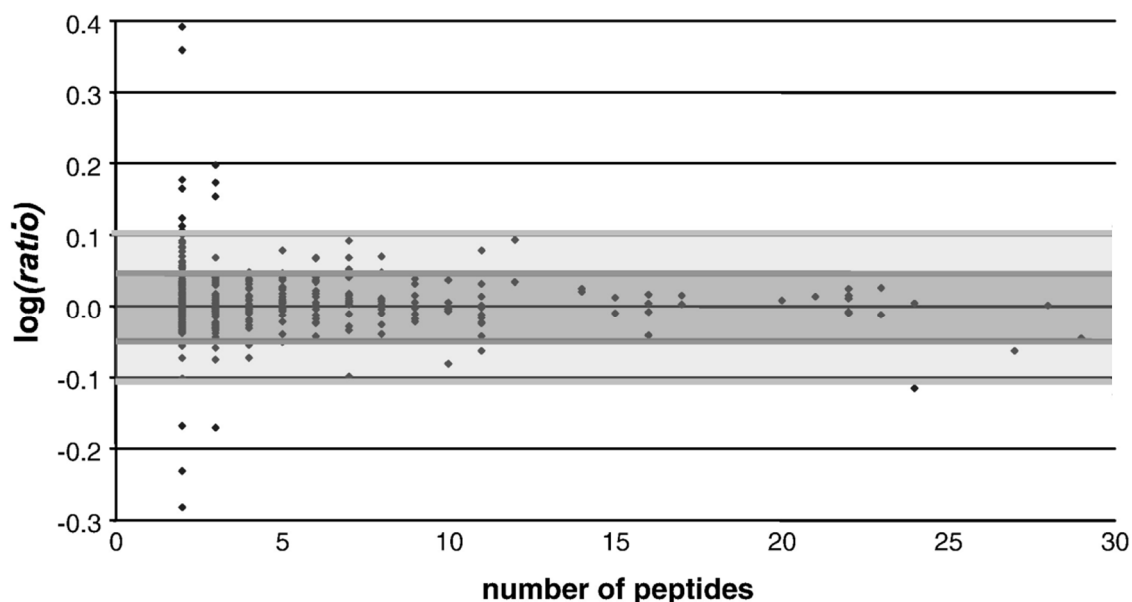
**Table 3. MED vs LPS or SEA labelling design. Each experiment number derived from a separate biological preparation.**

I performed LC-MALDI-MS/MS analysis on the combined iTRAQ samples as detailed in the paper. Repeating GO analysis on pooled iTRAQ results provided a similar distribution of cellular components (figure 13) to unlabelled samples, suggesting labelling was not changing the distribution of protein types identified. It should be noted the relative observation of nuclear proteins in SEA-DC samples is consistent with the unlabelled data, but remains suspicious from an enrichment viewpoint.



**Figure 13. Classification of identified proteins by GO terms after iTRAQ labelling.**

I used the ‘Whetton plot’ approach to determine differences in relative quantification based on variation in MED-DC controls.<sup>114</sup>  $\log_{10}$  ratios of MED-DC<sub>(iTRAQ114)</sub>/MED-DC<sub>(iTRAQ116)</sub> labelling replicates across all six biological replicates (n = 355) were plotted against number of identified peptides (figure 14). The Whetton plot provides an empirical estimation of the technical variance from which thresholds can be selected on the assumption that variance among experimental conditions will be approximately equal to variance in controls. Subjective values of 0.1 and 0.04 were chosen as high and low stringency thresholds respectively, based on the number of control protein measurements (false positives) exceeding these values. Proteins called as changing were required to have a 95% confidence interval for the mean ( $\pm 1.96$  SE) exceeding Whetton plot thresholds. Retrospectively applying the thresholds to the MED-DC/MED-DC controls gives a global 5% FDR at the higher threshold and a 20% FDR with the low stringency cut-off.



**Figure 14. Whetton plot, showing protein level  $\log_{10}$  fold change vs number of peptides identified for all technical replicates in the six iTRAQ analyses. Values of 0.1 (dark grey shaded region) and 0.04 (light grey shaded region) taken as high and low stringency thresholds respectively.**

Tables 4 to 7 list proteins designated as differently abundant. A total of 60 proteins exceeded the lower threshold in the SEA-DCs vs MED-DCs comparison, with 35 remaining at the higher-stringency cut-off. Filtering requiring mean  $\pm 1.96$  SE to exceed the fold-change threshold reduces the number of changes to 50 at lower stringency and 23 at the higher threshold.

Fewer changes were observed in LPS-DCs vs MED-DCs than in the SEA equivalent, with 18 values above the 0.04  $\log_{10}$  fold change threshold, dropping to 17 when the 95% confidence interval for SEM is applied. At the higher threshold 14 proteins exceed the 0.1 cut-off, reducing to nine applying  $\pm 1.96$  SE boundaries.

Protein	Mean log <sub>10</sub> (SEA-DCs/MED-DCs)	Mean - 1.96 SE (95% CI)
H2B histone family, member A	<b>0.34</b>	<b>0.25</b>
H3 histone, family 3A	<b>0.30</b>	<b>0.20</b>
Histone protein, family H2A	<b>0.30</b>	<b>0.23</b>
Histone H4	<b>0.27</b>	<b>0.23</b>
Histone 1, H4i	<b>0.26</b>	<b>0.18</b>
Histone H1.2	<b>0.25</b>	<b>0.20</b>
Heat shock protein HSP90-b	<b>0.23</b>	<b>0.17</b>
Histone H1.5	<b>0.20</b>	<b>0.13</b>
H2A histone family, member Z	<b>0.19</b>	<b>0.19</b>
Peptidyl-prolyl cis-trans isomerase	<b>0.18</b>	<b>0.15</b>
Vimentin	<b>0.18</b>	<b>0.15</b>
Tubulin b5	<b>0.17</b>	<b>0.17</b>
Histone H1.4	<b>0.16</b>	<b>0.16</b>
Heat shock protein HSP90-a	<b>0.15</b>	<b>0.13</b>
Protein disulfide isomerase related protein	<b>0.14</b>	<b>0.14</b>
Elongation factor 1-a1	<b>0.14</b>	<b>0.11</b>
S100 A9, calgranulin B, MRP-14	<b>0.14</b>	<b>0.10</b>
Heat shock cognate 71 kDa protein	<b>0.13</b>	<b>0.09</b>
Clathrin, heavy polypeptide	<b>0.13</b>	<b>0.12</b>
Pyruvate kinase 3	<b>0.12</b>	<b>0.10</b>
Non-musical heavy chain myosin II-A	<b>0.12</b>	<b>0.09</b>
Malate dehydrogenase	<b>0.11</b>	<b>0.10</b>
14-3-3 protein b/a	<b>0.11</b>	<b>0.10</b>
Talin-1	<b>0.10</b>	<b>0.09</b>
Lymphocyte cytosolic protein 1, 65 kDa macrophage protein	<b>0.10</b>	<b>0.07</b>
Peroxiredoxin 1	<b>0.10</b>	<b>0.07</b>
Heat shock cognate 71kDa protein	0.10	0.09
Coactosin-like protein	0.09	0.04
Elongation factor 2-b	0.09	0.08
Peptidyl-prolyl cis-trans isomerase	0.08	0.07
Ras-related protein Rab-7	0.07	0.03
Lysozyme C, type M precursor	0.06	0.06
Arginosuccinate synthase	0.06	0.02

Table 4. Proteins upregulated in SEA-DCs vs MED-DCs. Bold indicates log<sub>10</sub> ratios greater than 0.1. Yellow and green shading highlights values with mean – 1.96 SE >0.04 and >0.1 respectively.

Protein	Mean log <sub>10</sub> (SEA-DCs/MED-DCs)	Mean + 1.96 SE (95% CI)
Coronin, actin binding protein 1A	-0.05	-0.03
FcεRI	-0.05	-0.03
Cofilin 1, non-muscle	-0.05	-0.04
Ras GTPase-activating-like protein	-0.05	-0.03
Transgelin	-0.06	-0.03
Ras-related protein Rap-1B	-0.06	-0.04
Guanine nucleotide-binding protein, β-2 subunit	-0.06	-0.02
Murine Carbonic Anhydrase IV	-0.07	-0.04
Radixin	-0.08	-0.03
Hematopoietic cell specific Lyn substrate 1	-0.08	-0.05
Ezrin	-0.08	-0.06
CD44	-0.08	-0.05
Moesin (Membrane-organizing extension spike protein)	-0.08	-0.05
Guanine nucleotide-binding protein G(i), α-3 subunit	-0.09	-0.04
Myosin light chain, alkali, non-musical	-0.09	-0.05
Annexin A2, calpactin I heavy chain	-0.09	-0.06
Guanine nucleotide-binding protein G(i), α-2 subunit	-0.09	-0.07
Annexin A1, calpactin II	-0.10	-0.08
Actin, α, cardiac	-0.11	-0.07
H-2 class II histocompatibility antigen, A β chain precursor	-0.11	-0.11
Actin, β, cytoplasmic	-0.11	-0.08
Actin-related protein 2/3 complex subunit 2	-0.12	-0.12
β-actin	-0.12	-0.08
Calmodulin 2	-0.13	-0.07
Lymphocyte specific 1	-0.15	-0.11
S100 A10, calpactin I light chain	-0.15	-0.15
22 kDa neuronal tissue-enriched acidic protein, NAP-22, BASP-1, CAP-23	-0.21	-0.15

**Table 5. Proteins down regulated in SEA-DCs vs MED-DCs. Bold indicates log<sub>10</sub> ratios greater than 0.1. Yellow and green shading highlights values with mean + 1.96 SE <-0.04 and <-0.1 respectively.**

Protein	Mean log <sub>10</sub> (LPS-DCs/MED-DCs)	Mean - 1.96 SE (95% CI)
H-2 class II histocompatibility antigen, A-B a chain precursor	<b>0.32</b>	<b>0.31</b>
Transgelin 2	<b>0.27</b>	<b>0.11</b>
H-2 class II histocompatibility antigen, A b chain precursor	<b>0.16</b>	<b>0.07</b>
CD29, Integrin b-1 precursor (Integrin VLA-4 b-subunit)	<b>0.14</b>	<b>0.13</b>
22 kDa neuronal tissue-enriched acidic protein, NAP-22, BASP-1, CAP-23	<b>0.13</b>	<b>0.06</b>
CD98 heavy chain (4F2), lymphocyte activation antigen	0.10	0.09
Annexin A4	0.09	0.02
β-2-microglobulin precursor	0.07	0.06

**Table 6. Proteins upregulated in LPS-DCs vs MED-DCs. Bold indicates log<sub>10</sub> ratios greater than 0.1. Yellow and green shading highlights values with mean – 1.96 SE >0.04 and >0.1 respectively.**

Protein	Mean log <sub>10</sub> (LPS-DCs/MED-DCs)	Mean + 1.96 SE (95% CI)
S100 A10, calpactin I light chain	-0.05	-0.04
<b>CD18, Integrin β-2 precursor (LFA-1/CR3/P150,95 β-subunit)</b>	<b>-0.13</b>	<b>-0.09</b>
Prosaposin	<b>-0.15</b>	<b>-0.05</b>
ATP synthase a chain, mitochondrial precursor	<b>-0.18</b>	<b>-0.14</b>
Phosphate carrier protein, mitochondrial precursor	<b>-0.19</b>	<b>-0.07</b>
ATP synthase b chain, mitochondrial precursor	<b>-0.24</b>	<b>-0.14</b>
CD44	<b>-0.28</b>	<b>-0.26</b>
Cathepsin D precursor	<b>-0.29</b>	<b>-0.27</b>
Secretory protein YM-1 precursor	<b>-0.33</b>	<b>-0.31</b>
Moesin (Membrane-organizing extension spike protein)	<b>-0.33</b>	<b>-0.33</b>

**Table 7. Proteins down regulated in LPS-DCs vs MED-DCs. Bold indicates log<sub>10</sub> ratios greater than 0.1. Yellow and green shading highlights values with mean + 1.96 SE <-0.04 and <-0.1 respectively.**

The common MED-DCs sample used in both the SEA-DC and LPS-DCs analyses allowed relative SEA-DC vs LPS-DC protein ratios to be calculated (equation 2).

$$(2) \quad \log_{10}(\text{SEA-DCs/LPS-DCs}) = \log_{10}(\text{SEA-DCs/MED-DCs}) - \log_{10}(\text{LPS-DCs/MED-DCs})$$

**Equation 2. Calculation of SEA-DC vs LPS-DC protein ratios.**

Four proteins were upregulated in SEA-DCs at the higher Whetton plot-derived fold-difference threshold, while eight proteins were upregulated in LPS-DCs (table 8). Two further proteins showed greater expression in LPS-DCs at the lower stringency 0.04 threshold.

Protein	Mean log <sub>10</sub> (SEA-DCs/LPS-DCs)
<b>Moesin</b>	<b>0.26</b>
<b>CD44</b>	<b>0.21</b>
<b>CD18</b>	<b>0.11</b>
<b>Rab-7</b>	<b>0.11</b>
Rap-1B	-0.08
Actin, $\alpha$	-0.09
<b>Actin, <math>\beta</math></b>	<b>-0.10</b>
<b>S100 A10</b>	<b>-0.11</b>
<b><math>\beta</math>-2-microglobulin precursor</b>	<b>-0.11</b>
<b>Annexin A4</b>	<b>-0.12</b>
<b>Annexin A1</b>	<b>-0.13</b>
<b>H-2 class II histocompatibility antigen, A b chain precursor</b>	<b>-0.27</b>
<b>22 kDa neuronal tissue-enriched acidic protein, NAP-22, BASP-1, CAP-23</b>	<b>-0.29</b>
<b>Transgelin 2</b>	<b>-0.33</b>

**Table 8. Proteins differently regulated in SEA-DCs vs LPS-DCs. Bold indicates log<sub>10</sub> ratios greater than 0.1.**

Although published in 2010 the work for this paper was performed in 2007. The field has changed significantly in the intervening time and of the studies presented, this is the analysis I would change most if I were to repeat it today, almost a decade on.

Number of proteins identified was relatively modest, but as with paper 2 this is attributed to the LC-MALDI-MS/MS instrumentation available. MALDI-TOF/TOF mass spectrometers are also less suited to iTRAQ analyses because of the relatively wide precursor selection window mandated by the use of a precursor ion selector voltage gate.<sup>115</sup> Increasing precursor ion selection width increases propensity for co-selection and generation of chimeric spectra that have been shown to lead to attenuation of iTRAQ ratios. Modern ion-trap instruments can improve quantitative accuracy with isobaric tagging analyses by acquiring MS<sup>3</sup> spectra to sample reporter ion intensity, improving specificity of measurement.<sup>116,117</sup>

In retrospect, the design would be improved by including SEA and LPS treatments in the same labelling set, allowing direct comparison. Thermo now produce 10-plex isobaric labelling reagents that could compare three biological replicates of SEA-, MED- and LPS-treated DCs in a single acquisition. The number of biological replicates required could be estimated using power analysis once biological variation was measured in a scoping experiment.<sup>118</sup> If greater than three replicates were needed, successive 9-plex iterations could be performed, although optimally all samples should be produced simultaneously to minimise variance.

While well adopted at the time, the Whetton plot approach is now rarely used. The regime forces assumption that variance is equal between controls and comparator. In practice, variance differs protein-to-protein within sample, let alone inter-sample. Modern approaches assess variance for



individual proteins among biological replicates and either compare transformed ratios to a null hypothesis of normality or allow direct comparison of normalised reporter ion intensities.<sup>119-121</sup> Such methods calculate protein level p-values indicating statistically significant differences. Where data can be shown normally distributed, Welch's t-test can be applied (equation 3), to compare means ( $\bar{X}$ ) and variance ( $s$ ) for  $N$  observations. Unlike Student's t-test Welch's does not assume equal variance between comparators.

$$(3) \quad t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

**Equation 3. Welch's t-test.**

If the data are non-normal, non-parametric tests such as Wilcoxon signed-rank test can compare relative ranks  $R$  between populations (equation 4).

$$(4) \quad W = \sum_{i=1}^{N_r} [\sin(x_{2,i} - x_{1,i}) \cdot R_i]$$

**Equation 4. Wilcoxon signed-rank test.**

Whichever test is chosen, multiple test-correction should be used to ensure false significance is not obtained through virtue of testing repetition. The Bonferroni (equation 5) or less stringent Hochberg and Benjamini approaches (equation 6) can calculate multiple-test corrected q-values or FDRs respectively.

$$(5) \quad q = p \cdot \text{no. tests}$$

**Equation 5. Bonferroni multiple test correction.**

$$(6) \quad FDR = \frac{p \cdot \text{no. tests}}{\text{rank}}$$

**Equation 6. Hochberg and Benjamini calculation of false discovery rate.**

Considering that GO analysis of the SEA and LPS treated samples shows different proportions of membrane and nuclear proteins (figure 12) and the highest ranked proteins in SEA vs MED DCs were non-membrane proteins (table 4 – note prevalence of nuclear histones), the variation imparted by the membrane enrichment must be questioned. While replicates could be used to determine preparation variance, a better approach would be to move from iTRAQ or TMT to a SILAC approach. With SILAC, samples are combined prior to membrane enrichment, removing it and subsequent digestion steps as points of inter-sample variation.<sup>105</sup>

Validity of any enrichment should also be considered. For instance, if a protein is of lower abundance in the membrane preparation one cannot know whether expression of the protein has reduced globally or if the protein has been trafficked from the membrane into the soluble component. Using modern LC-MS/MS instrumentation in conjunction with protein or peptide fractionation it may be possible to sample deeply enough to obtain membrane penetration without enrichment. If enrichment is required, the soluble fraction should also be analysed to test for spatial differences. True spatial proteomics is only now becoming tractable. Kathryn Lilley is pioneering the field with development of LOPIT,<sup>122</sup> which uses a combination of isotopic labelling, density gradient fractionation and principle component analysis to assign organelle localisation by comparison with marker proteins of known organelle location.

A positive aspect of paper 5 was the orthogonal validation performed by SF-B using flow cytometry for 11 proteins with commercially available antibodies (table 9). Three proteins were not identified in my proteomic analysis. Of the remaining 19 comparisons, 12 were in agreement, either both showing significant differences or showing no change. Seven measurements showed significant change in only one technique. There were no cases where quantifications determined significant differences with the two approaches that were contradictory.

Protein	SEA vs MED iTRAQ	SEA vs MED Flow Cytometry	LPS vs MED iTRAQ	LPS vs MED Flow Cytometry
IA-b	MED	No change	LPS	LPS
CD98	No change	SEA	LPS	LPS
Galectin	No change	No change	No change	No change
IA/IE	MED	No change	LPS	LPS
MGL	No change	No change	No ID	No ID
CD29	No ID	No ID	LPS	LPS
CD18	No change	SEA	MED	No change
Rab7	No change	No change	No change	No change
S100 A10	MED	MED	No change	LPS
CD44	MED	No change	MED	MED
YM-1	No ID	No ID	MED	MED

**Table 9. Validation of iTRAQ results using flow cytometry. Shading indicates: green, upregulation in SEA-DCs; blue upregulation in MED-DCs; pink, upregulation in LPS-DCs; grey, no significant change; white, protein not identified in iTRAQ data set. iTRAQ significance taken as log mean ratio  $\pm 1.96 SE >0.04$  or  $<-0.04$ . Flow cytometry significance =  $p<0.05$ .**

In spite of the imperfections of the instrumentation and experimental design, biologically meaningful results were still obtained. LPS-DCs showed increase in H-2 class II histocompatibility antigen relative to both MED-DCs control and SEA-DCs, which is consistent with the maturation of DCs leading to Th1 response. Downregulation of the same protein upon SEA treatment is evidence of alternative Th2-type response. Surprisingly, NAP-22, a neuron-enriched protein localized mainly in the synaptic vesicles,<sup>123</sup> was shown to be the most down-regulated protein upon SEA treatment. NAP-22 regulates neurite outgrowth, synaptogenesis and growth-cone

guidance during development of the nervous system<sup>124</sup> but has not been reported in dendritic cells. Its identification in my data is confident with five unique peptides matched. Such a gross difference between treatments and unexpected observation calls for further analysis, although to date no immunological effect of NAP-22 has been reported in DCs.

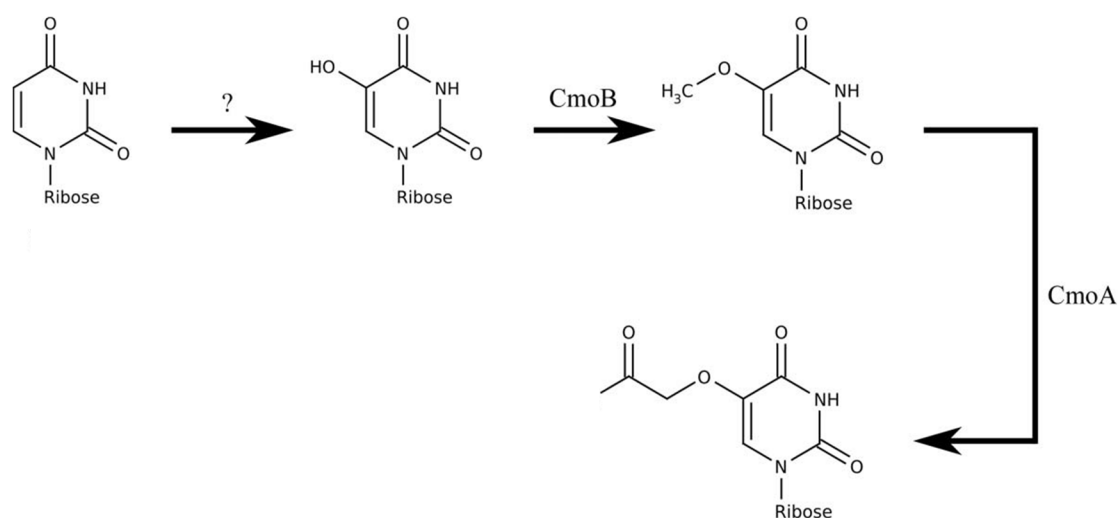
My proteomic data determined clathrin as upregulated in SEA-DCs. A publication two years later by Srivastava observed the same effect in a similar SEA-directed Th2-type immune response analysis.<sup>125</sup> Srivastava also measured upregulation of Rab7. Although increase of Rab7 was not significant in my SEA-DCs vs MED-DCs data set, Rab7 was calculated to have the fourth highest magnitude change between SEA-DCs vs LPS-DCs, suggesting a potentially important distinction between Th1-type and Th2-type response. Both clathrin and Rab7 are functional in endocytosis,<sup>125</sup> the incorporation of molecules within a cell.<sup>126</sup> Further evidence for upregulation of antigen uptake was evidenced by increases in heat shock protein 90 and heat shock conjugate 71, both of which are involved in protein folding.<sup>127</sup>

SEA-treated DCs showed upregulation of several enzymes (peptidyl-prolyl cis-trans isomerase, protein disulfide isomerase, pyruvate kinase 3, malate dehydrogenase, peroxiredoxin 1, peptidyl-prolyl cis-trans isomerase, lysozyme C and argininosuccinate synthase), suggesting greater metabolic activity. Increase in histones implies increased protein synthesis, which may aid cell proliferation.<sup>74</sup> No corresponding upregulation of enzymes or histones was observed for LPS-DCs. It should be noted that, as observed in paper 1, histones are commonly modified and the changes in regulation here may be confounded by PTM differences.

While the study appears naïve by today's standards, my proteomic dataset added to the expanding knowledge that Th2 response induced by SEA-DCs progresses through limited maturation of DC cells.<sup>112</sup> In addition, my proteomic analysis yielded deeper penetration into the membrane component than was possible with previous gel-based studies and has subsequently been cited ten times.<sup>128-137</sup>

## Paper 6 – *S*-adenosyl-*S*-carboxymethyl-L-homocysteine: a novel co-factor found in the putative tRNA modifying enzyme CmoA

Paper 6 centres on my mass spectrometric identification of a novel protein co-factor. The work began in 2010 when I was contacted by Fiona Whelan (FW) and Fred Antson (FA), a post-doctoral researcher and Professor respectively, at University of York who were collaborating with Robert Byrne (RB) of Oxford University. They were studying CmoA, an enzyme that catalyses 5-methoxyuridine to uridine-5-oxyacetic acid conversion (figure 10). The conversion is part of the uridine to uridine-5-oxyacetic acid biosynthetic pathway<sup>138</sup> in bacterial tRNA (figure 15). Nucleoside modification in tRNA transforms the macromolecule into its mature state and conversion to uridine-5-oxyacetic acid aids translation by altering anticodon stem loop conformation, lowering the entropic increase required for remodelling.<sup>139</sup>



**Figure 15. Biosynthetic pathway of uridine to uridine-5-oxyacetic acid conversion in bacterial tRNA.**

The researchers' studies had culminated in the resolution of CmoA crystal structure using X-ray crystallography but there remained unexplained intensity in the data. They hypothesised the intensity could derive from a co-factor, but it was incongruent with the expected ligands *S*-adenosylmethionine or *S*-adenosylhomocysteine involved in similar biosynthetic methylations.<sup>140-142</sup> The density instead favoured an *S*-adenosylmethionine derivative where the methyl group was substituted with a group of trigonal planar geometry. I explained that if the ligand could be observed by FTICR-MS the accuracy afforded should be sufficient for empirical formula determination.

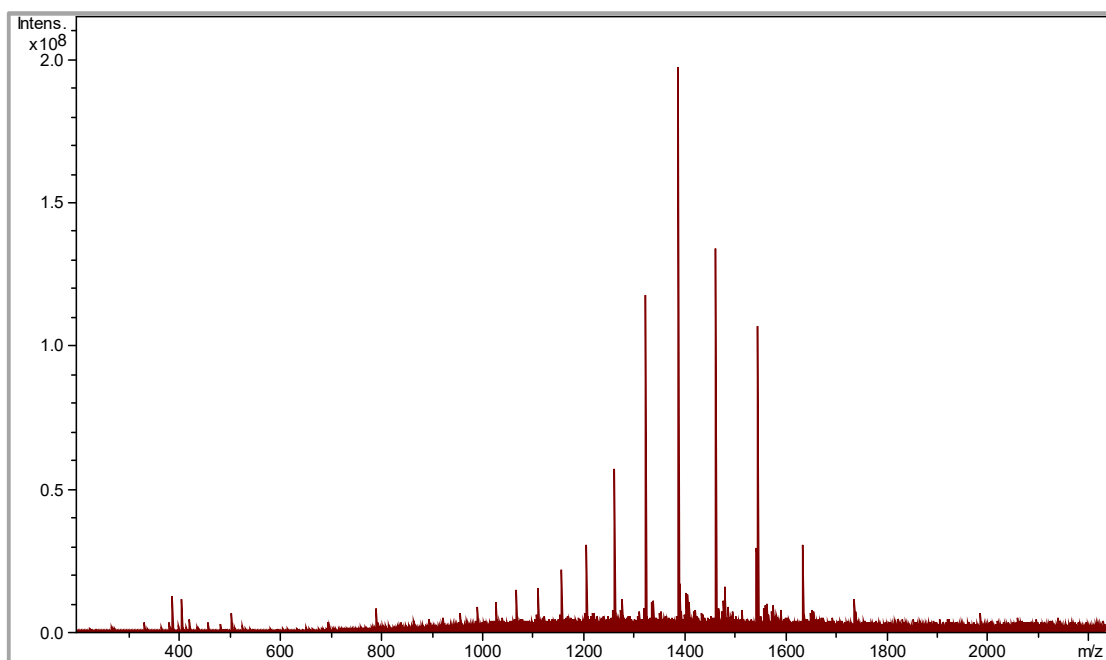
FTICR-MS uses the principle that ions exposed to a magnetic field undergo cyclotron motion with angular frequency ( $\omega_c$ ) proportional to magnetic field strength ( $B$ ) and inversely proportional to  $m/z$  (equation 7).

$$(1) \quad \omega_c = \frac{zB}{m}$$

**Equation 7. Calculation of cyclotron angular frequency.**

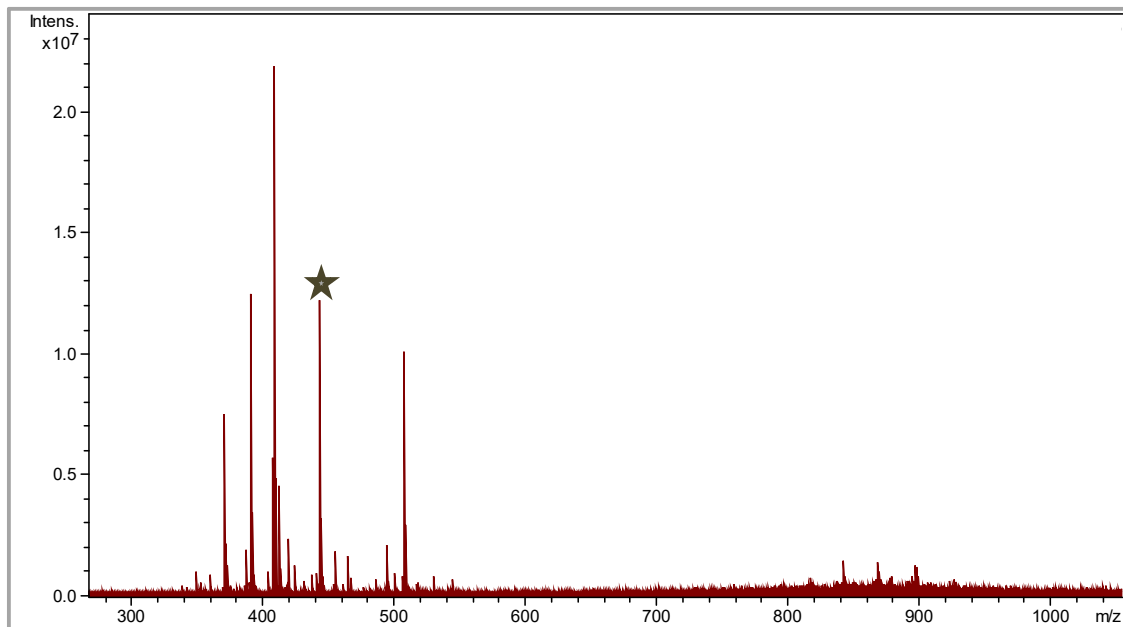
Although costly high strength magnets are required, angular frequency can be measured with very high precision providing unparalleled mass accuracy compared with other mass analysers. If the analyte could be fragmented, the product ion spectrum may provide information on the ligand's structure.

FW provided a purified extract of *E. coli*-expressed CmoA that I analysed using FT-ICR-MS as in paper 2. Under denaturing conditions (aqueous 50% (v/v) acetonitrile containing 1% (v/v) formic acid) signal was observed for the protein (figure 16). Deconvolution using Bruker's sophisticated numerical annotation procedure (SNAP) provided a mass of 27,763.2 Da, 0.3 Da mass error from the predicted mass of CmoA. SNAP uses an averagine algorithm to pick masses from isotopic patterns by comparing with predicted patterns for combinations of predefined atoms.



**Figure 16. FTICR-MS spectrum of purified CmoA. Deconvoluted mass measured = 27,763.2 Da.**

I repeated acquisition with parameters adjusted to facilitate observation of the lower mass ligand. Source accumulation was increased from 0.01 to 0.2 s and flight time to acquisition cell reduced from 1.2 to 0.6 ms. Figure 17 presents the resulting spectrum.



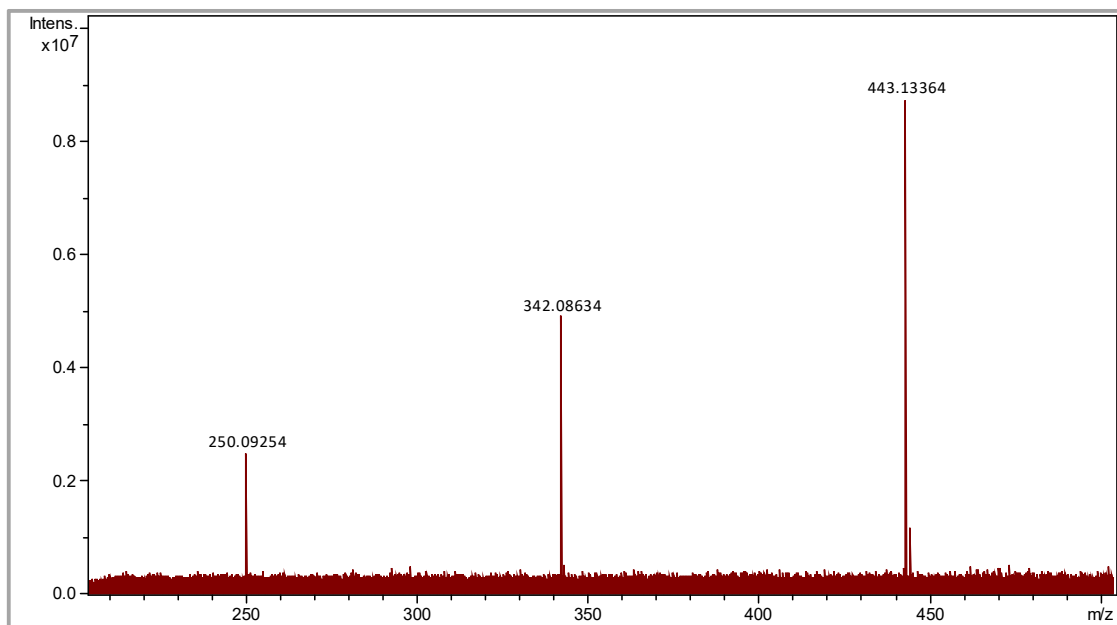
**Figure 17. FTICR-mass spectrum of purified CmoA, with instrument tuning to favour observation of low  $m/z$  signal. The ion at  $m/z$  443.1333 is starred.**

Multiple ions were observed between  $m/z$  300 and 600. I selected the most intense precursors for fragmentation using CID in the hexapole collision cell (qCID). Precursor and product ions were submitted to SmartFormula3D™ in Bruker DataAnalysis software to assign empirical formulae. Mass errors were restricted to 1 mDa and C, H, N, O and S atoms were considered.

Searching the PubChem database for compounds structurally similar to *S*-adenosylmethionine or *S*-adenosylhomocysteine within a 0.5  $m/z$  unit of the observed precursor ions provided a single match to a derivative in which the methyl group is replaced by a carboxymethyl group: *S*-adenosyl-*S*-carboxymethyl-L-homocysteine;  $m/z$  443.1343 [M+H]<sup>+</sup>.

*S*-adenosyl-*S*-carboxymethyl-L-homocysteine is within 1 mDa of the precursor ion at  $m/z$  443.1333 (starred figure 17). The qCID spectrum for this precursor produced product ions at  $m/z$  250.0925 and 342.0863 (figure 18). SmartFormula3D™ suggested two empirical formulae that could reconcile all three  $m/z$  values (table 10). When ranked by fit to isotopic pattern (mSigma) the top hit (C<sub>16</sub>H<sub>23</sub>N<sub>6</sub>O<sub>7</sub>S) is consistent with *S*-adenosyl-*S*-carboxymethyl-L-homocysteine.

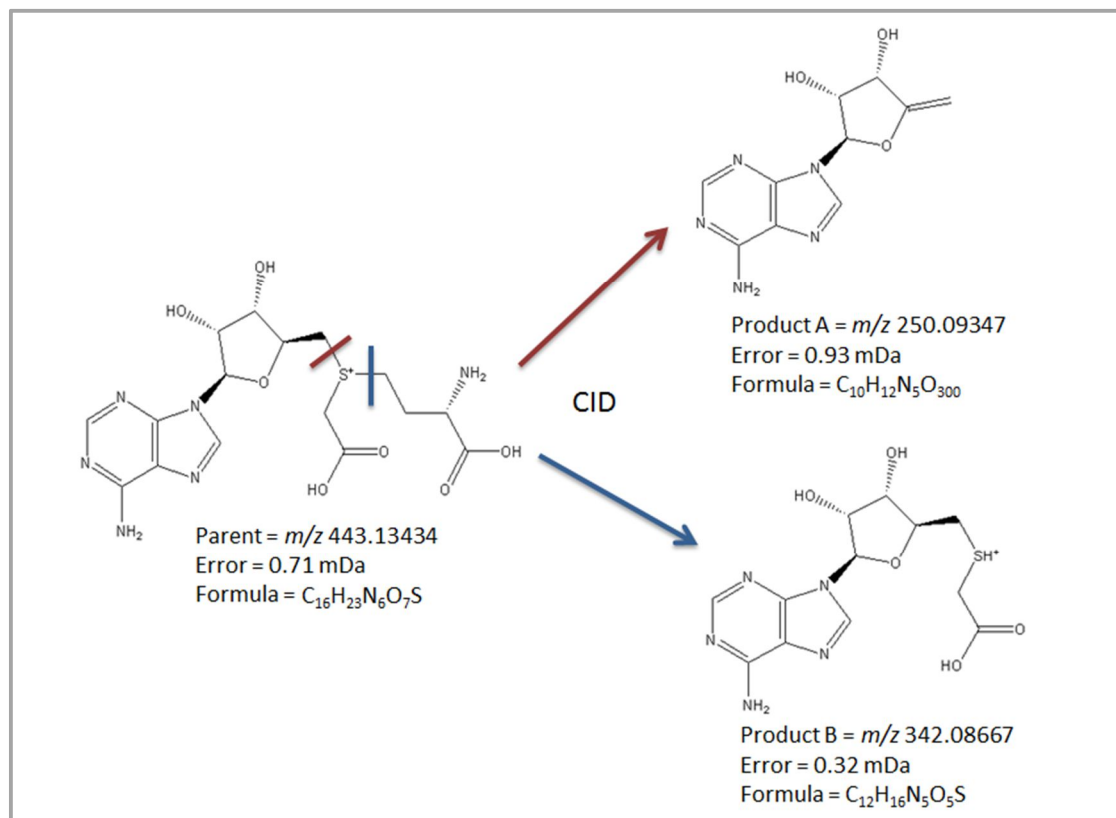
In addition to the precursor  $m/z$ , suggested empirical formulae for both product ions of the top ranked match are consistent with fragmentation of *S*-adenosyl-*S*-carboxymethyl-L-homocysteine at C-S bonds (figure 19).



**Figure 18.** qCID FTICR-mass spectrum of precursor m/z 443.1333. Product ion m/z measurements are annotated.

Result Rank	Precursor Formula	mSigma	m/z	Error mDa
1	C <sub>16</sub> H <sub>23</sub> N <sub>6</sub> O <sub>7</sub> S	25.9	443.13434	0.71
	Product Formula		m/z	Error mDa
	C <sub>10</sub> H <sub>12</sub> N <sub>5</sub> O <sub>3</sub>		250.09347	0.93
	C <sub>12</sub> H <sub>16</sub> N <sub>5</sub> O <sub>5</sub> S		342.08667	0.32
Result Rank	Precursor Formula	mSigma	m/z	Error mDa
2	C <sub>14</sub> H <sub>21</sub> N <sub>9</sub> O <sub>6</sub> S	22.7	433.13300	-0.64
	Product Formula		m/z	Error mDa
	C <sub>8</sub> H <sub>10</sub> N <sub>8</sub> O <sub>2</sub>		250.09212	-0.42
	C <sub>12</sub> H <sub>16</sub> N <sub>5</sub> O <sub>5</sub> S		342.08667	0.32

**Table 10.** SmartFormula3D™ results table generated from m/z values observed upon qCID fragmentation of ion at m/z 433.1333. Table suggests two possible empirical formulae consistent with the precursor ion and the two dominant product ions measured at m/z 342.08667 and 250.09347. The suggested empirical formulae must be within 1 mDa of the measured values, with product ion formulae subsets of the precursor formula. Results ranked by mSigma, a measure of how well the observed isotopic pattern matched the theoretical pattern for the suggested precursor.



**Figure 19. Illustration of fragmentation positions in *S*-adenosyl-*S*-carboxymethyl-*L*-homocysteine, showing masses and formulae consistent with values measured by FTICR-MS.**

Subsequent re-analysis of the X-ray crystallography data by RB and FA, modelling *S*-adenosyl-*S*-carboxymethyl-*L*-homocysteine as the co-factor produced good alignment with the observed densities, confirming my mass spectrometry identification.

I observed no signal for either *S*-adenosylmethionine ( $m/z$  399.1445) or *S*-adenosylhomocysteine ( $m/z$  385.1289) in the mass spectrum of CmoA (figure 17). Other signals in the low  $m/z$  region could not be rationalised against *S*-adenosylmethionine or *S*-adenosylhomocysteine derivatives or hybrids and no further components were suggested from the crystallography data. Additional signals in the mass spectrum are most likely environmental contaminants rather than additional cofactors.

It should be noted that my mass spectrometry data in isolation do not prove complexation of *S*-adenosyl-*S*-carboxymethyl-*L*-homocysteine with CmoA, rather their co-localisation in the extract. I hoped to observe the complex in the gas phase but this remained elusive. In addition to denaturing analysis I prepared the same sample in aqueous 35 mM ammonium acetate pH 7.5 at a range of concentrations from 1-100  $\mu$ M. I also added between 0.5-5 % acetonitrile to help spray stability but under none of the conditions was I able to observe the complex within the mass spectrometer. I could only maintain signal by increasing source accumulation and in-source decay to a level where the dissociated ligand and protein were observed. I hypothesise that maintaining

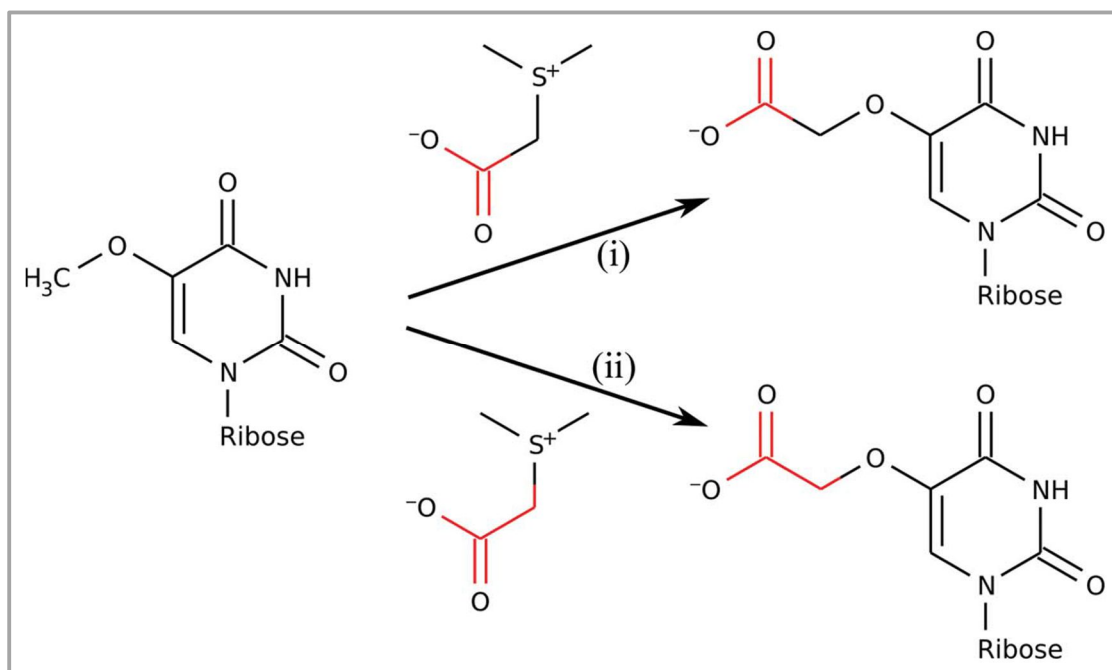


the complex may have been difficult because the native form is a homodimer of two CmoA proteins and without maintaining protein dimerization, ligand affinity may be lost.

If the complex had been maintained in the gas phase my plan was to isolate a single charge state ion and fragment the intact protein in the ICR cell to release the ligand, simultaneously proving complexation and generating  $m/z$  measurements for identification. As it stands, the combination of my mass spectrometry and RB's and FA's X-ray crystallography provide good evidence for *S*-adenosyl-*S*-carboxymethyl-L-homocysteine being the co-factor of CmoA.

Had X-ray data not been available alternative methodologies such as surface plasmon resonance,<sup>143,144</sup> analytical ultracentrifugation<sup>145</sup> or isothermal titration calorimetry<sup>144,146</sup> could be used to test CmoA and *S*-adenosyl-*S*-carboxymethyl-L-homocysteine binding affinity.

My identification of *S*-adenosyl-*S*-carboxymethyl-L-homocysteine as the cofactor of CmoA provides a novel and unexpected insight into tRNA modification. It is currently hypothesised that the cofactor may assist in uridine-5-oxyacetic acid formation through either (i) transfer of the carboxyl group onto the methoxy group of 5-methoxyuridine or (ii) substitution of the methyl group side chain in 5-methoxyuridine by the entire carboxymethyl group (figure 20). These hypotheses could be tested by feeding the system with isotopically labelled *S*-adenosyl-*S*-carboxymethyl-L-homocysteine and following incorporation of labelled carbons in the enzyme product.



**Figure 20.** Hypothesised roles of cofactor *S*-adenosyl-*S*-carboxymethyl-L-homocysteine in the modification of 5-methoxyuridine. Modification could involve either (i) the addition of a carboxyl group (red) onto the methoxy group of 5-methoxyuridine or (ii) substitution of the methyl group of 5-methoxyuridine for the entire carboxymethyl group (red).

Following *S*-adenosyl-*S*-carboxymethyl-L-homocysteine's identification in complex with CmoA, RB has shown homologous intensity in the X-ray structure of *H. influenza* YecO protein which on re-interrogation better fits with the same co-factor, suggesting the substrate may be more widely used in nature.

The paper has been cited five times since publication with the citing studies mostly referencing the identification of *S*-adenosyl-*S*-carboxymethyl-L-homocysteine, which was achieved through my mass spectrometry analysis.<sup>147-151</sup>

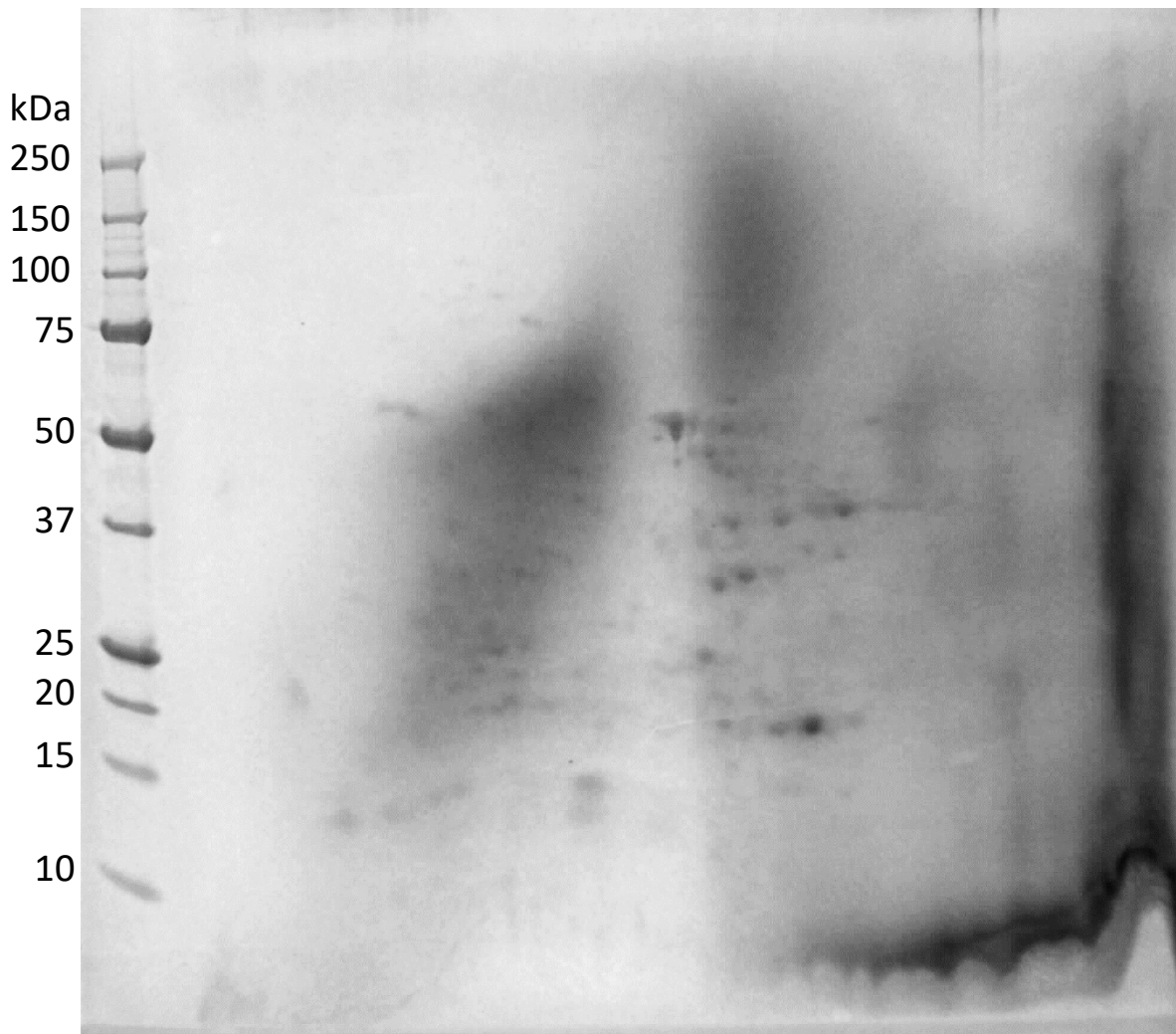
## Paper 7 – Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein

Paper 7 identifies the novel fusion protein STORR as the final non-elucidated step in the morphinan biosynthesis pathway. My role was to use mass spectrometry to prove the proposed fusion protein was naturally expressed as a single contiguous sequence.

At date of publication (2014) all but one of the genes responsible for morphinan synthesis were known,<sup>152-154</sup> leaving elucidation of the final component, which epimerizes (*S*)- to (*R*)-reticuline, critical to full understanding and potential exploitation of the biological system. The pathway is the basis of opiate production in the opium poppy (*Papaver somniferum*), which was first demonstrated in 1806<sup>155</sup> and remains the commercial source of production.

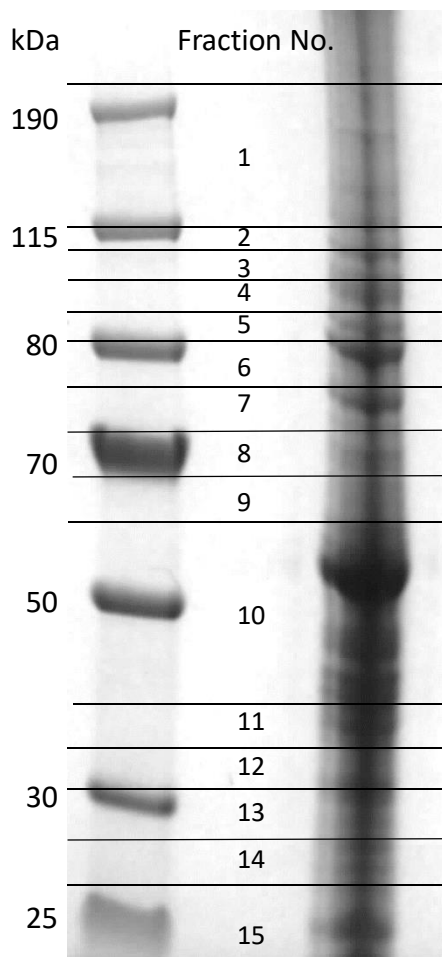
Investigation of an in-house expressed sequence tag (EST) library from opium poppies by University of York postdoctoral researchers Thilo Winzer (TW) and Marcello Kern (MK) suggested the final gene product in the pathway was a contiguous assembly comprising a cytochrome P450 monooxygenase 3'-linked to an oxidoreductase. The contiguous assembly was designated STORR ((*S*)- to (*R*)-reticuline). The EST sequence suggested constituent protein subunits are fused through the linker peptide KPCVQSAASERD. Although the genomic data indicated a novel fusion it was vital to show that the fusion was translated and expressed in wild type plants. TW had attempted to raise an antibody against the STORR fusion but it had proved unsuccessful for purification or western blotting. TW and MK asked whether I could demonstrate the contiguous fusion in wild type plants using a mass spectrometry approach.

I first attempted to identify the fusion using 2D-PAGE, which although declining in usage is still unrivalled for resolution of intact proteins. If a single spot could be isolated containing sequences from both protein subunits it would be strong evidence for the fusion. Protein extracts from *Papaver somniferum* stems were provided by MK and I separated the material using 2D-PAGE. The resulting Coomassie-stained gel is shown in figure 21. An *E. coli* expressed recombinant version of STORR had been produced by Jared Cartwright (JC) for the purpose of antibody production. My intention was to run the recombinant protein into an analogous 2D-PAGE gel to localise position of the protein in the wild type extract for excision. Unfortunately, the recombinant protein proved to be insoluble in isoelectric focusing loading buffer. In the absence of a control, localisation of STORR was difficult. The theoretical molecular mass and pI are 100.5 kDa and 6.48 respectively, and although I considered speculatively excising spots around these values my focus instead turned to a 1D-PAGE approach.



**Figure 21. Coomassie-stained gel image resulting from 2D-PAGE separation of proteins from *Papaver somniferum* stems.**

Although lower resolution than 2D-PAGE, if the protein subunits could be co-localised within a 1D-PAGE segment at the expected mass of STORR it could still provide strong indication of the fusion. I asked MK to provide another protein extract, which I ran into a 1D-PAGE gel (figure 22). I fractionated the gel unevenly to create finer gradation around the expected position of STORR at 100 kDa (fraction 4). To look for the presence of unfused cytochrome P450 monooxygenase and oxidoreductase, narrow fractions were also taken around 65 kDa and 35 kDa, consistent with the theoretical masses of these two enzymes. I subjected proteins in the 15 excised fractions to in-gel trypsin digestion after reduction with dithioerythritol and S-carbamidomethylation with iodoacetamide, prior to LC-MS/MS acquisition using a maXis qTOF mass spectrometer. Searching peak lists against an in-house *Papaver somniferum* database using the Mascot search engine I identified 2,853 proteins at 1% FDR. FDR was estimated by comparing relative proportions of matches obtained in the forward database against a randomised database with equal distribution of amino acids, protein lengths and cleavage site frequency.



**Figure 22. 1D-PAGE of high-morphine cultivar protein extracts. Fractions shown were excised for tryptic digestion.**

Fraction 4, at the expected mass of the fusion, yielded 76% sequence coverage for the fusion with peptides spanning the length of the protein (figure 23). Crucially, coverage included the linker peptide KPCVQSAASERD, fused at the N-terminus to P450 monooxygenase (IKPCVQSAASER positions 559-579) and C-terminally to oxidoreductase (DMESSGVPVITLGSGK positions 580-595). The linker peptide is unique in the *Papaver somniferum* and NCBI nr databases, meaning identification cannot be dismissed as homology with an incidental protein. With termini flanked by supporting sequences it is strong evidence for the expression of the STORR fusion protein. The two peptide matches spanning the linker peptide are made with good expect scores and show good correlation with predicted sequence (figures 24 and 25).

1 MELQYISYFQ PTSSVVALLL ALVSILSSVV VLR**KTFLNNY** **SSSPASSTKT**  
 51 AVLSHQ**RQQS** **CALPISGLLH** **IFMNKGLIH** **VTLGNMADKY** **GPIFSFPTGS**  
 101 **HRTLIVSSWE** **MVKECFTGNN** **DTAFSNRPIP** **LAFKTIFYAC** **GGIDSYGLSS**  
 151 **VPYGYWREL** **RKVCVHNLLS** **NQQLLFRHL** **IISQVDTSFN** **KLYELCKNSE**  
 201 **DNHGNYYYYT** **TTAAGMVRID** **DWLAELSFNV** **IGRIVCGFQS** **GPKTGAPSRV**  
 251 EQFK**EAINEA** **SYFMSTSPVS** **DNVPMLGWID** **QLTGLTRNMK** **HCGK**KLDLVV****  
 301 **ESIINDHRQK** **RRFSRTKGGD** **EKDDEQDDFI** **DICLSIMEQP** **QLPGNNNPSQ**  
 351 IPIK**SIVLDM** **IGGGTDTTKL** **TTIWTLSLLL** **NNPHVLDKAK** **QEVD**AHFR**TK**  
 401 **RRSTNDAAAA** **VVDFDDIRNL** **VYIQAIKES** **MRLYPASPVV** **ERLSGEDCVV**  
 451 **GGFHV**PAGTR**** **LWANVWKMQR** **DPKVVDDPLV** **FRPDR**FLSDE**** **QKMVDVR**GQN****  
 501 **YELL**PF**GAGR** **RVCPGV**SFSL**** **DLMQ**LV**TRL** **ILE**FEM**KSPS** **GK**VDM**TATPG**  
 551 **LMSYK**VI**PLD** **ILL**THR**RIKP** **CV**Q**SAASERD** **MES**S**SGVPVIT** **LG**S**GK**VMP**VIL**  
 601 **GMGT**FEK**VGK** **GSE**RER**LAIL** **KAIE**V**GYRYF** **DTAA**AY**ETEE** **VL**GEA**IAEAL**  
 651 **QLGLV**KSR**DE** **LFIS**S**MLWCT** **DAH**ADR**VLLA** **LQNSLRNLKL** **EYVDLYMLPF**  
 701 **PASL**KPG**KIT** **MDI**PEE**DICR** **MDY**RS**VWAAM** **EECQNLG**FTK**** **SIGVSNF**SCK****  
 751 **KLQEL**MAT**AN** **IPPA**VN**QVEM** **SPA**FQ**QKLR** **EYCN**ANN**ILV** **SAISV**LGS**NG**  
 801 **TPWGS**NAV**LG** **SEVL**KK**IAMA** **KGK**SVA**QVSM** **RWV**YE**QGASL** **VVK**SF**SEERL**  
 851 **REN**LN**IFDWE** **LTKED**HEK**IG** **EIP**Q**CRILSA** **YFLV**SP**NGPF** **KSQ**EEL**W**DDE****  
 901 A

Figure 23. Sequence coverage for endogenous STORR fusion protein excised from between 95-105 kDa in gel (fraction 4, figure 22). Identification performed by Mascot database searching post LC-MS/MS analysis of trypsin derived peptides. Colour indicates amino acids contained in peptides identified with expect scores of 0.05 or lower. Red sequences are attributable to the P450 monooxygenase module. Green sequences are from the oxidoreductase module. Purple amino acids belong to linker sequence.

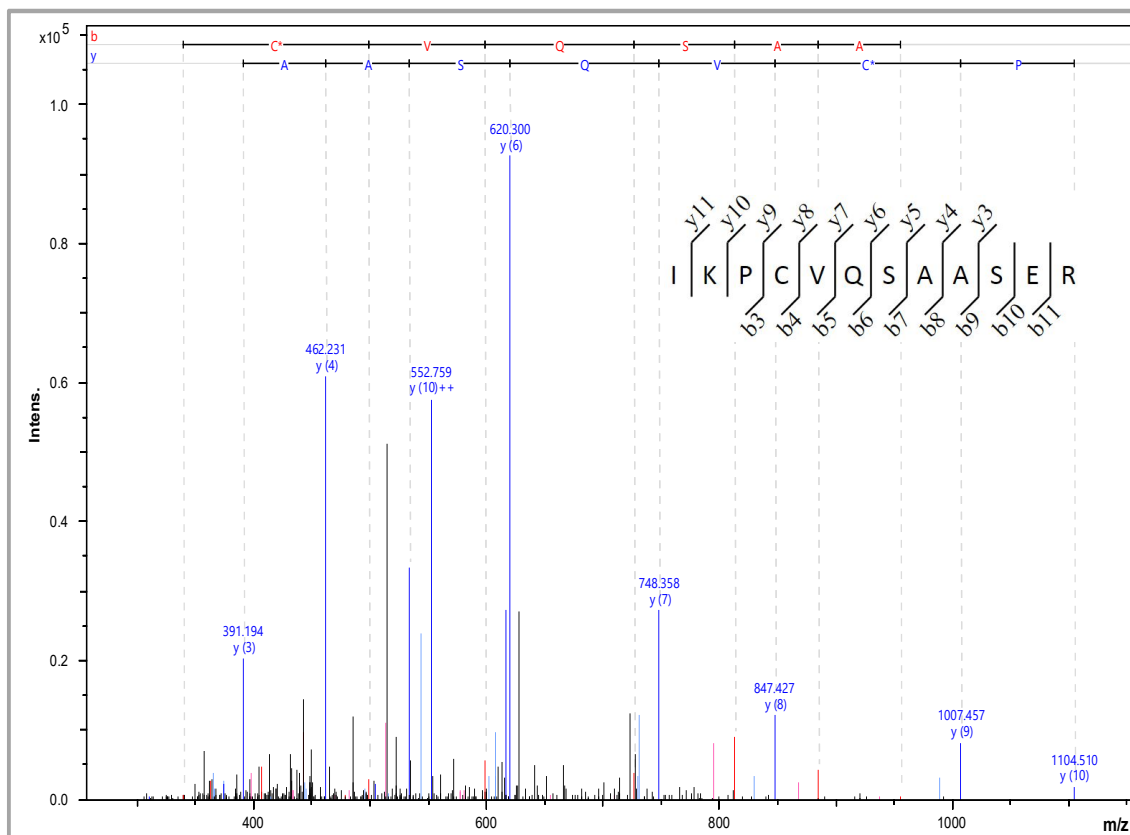
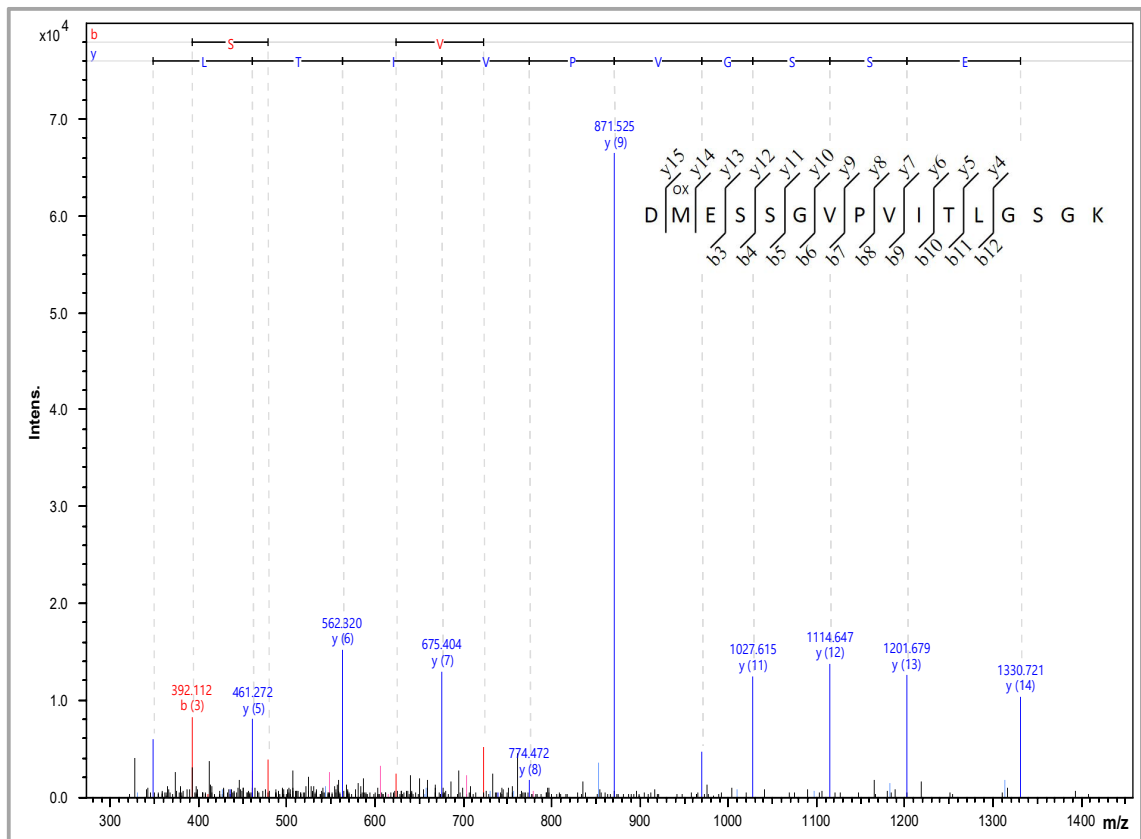


Figure 24. Annotated product ion spectrum for peptide identification IKPCVQSAASER (linker peptide pos 559-579). Mascot expect score =  $9.8e-07$ . Spectrum annotated with matches to theoretical b- (red) and y- ion series (blue).



**Figure 25. Annotated product ion spectrum for peptide identification DMESSGVPVITLGS GK (linker peptide pos 580-595). Mascot expect score =  $4e-09$ . Spectrum annotated with matches to theoretical b- (red) and y-ion series (blue).**

STORR peptides were not unique to fraction 4, with identification in 12 of the 15 fractions analysed. However, a clear spike in identification frequency was measured at the expected position (100 kDa – fraction 4), when measured in terms of sequence coverage, spectral counts or number of unique peptide sequences (table 11).

Fraction	Peptide IDs from sequences of:	Sequence coverage (%)	Spectral matches	No. peptide sequences
1	P450	1	1	1
2	P450 + OxR	17	12	11
3	P450 + OxR + Linker	49	41	35
4	P450 + OxR + Linker	76	149	55
5	P450 + OxR + Linker	46	38	32
6	P450 + OxR + Linker	17	11	11
7	P450 + OxR	14	11	10
8	P450 + OxR	10	8	7
9	P450 + OxR	17	11	11
10	N/A	0	0	0
11	OxR	5	6	5
12	OxR	1	2	2
13	N/A	0	0	0
14	N/A	0	0	0
15	OxR	0	1	1

**Table 11.** Comparison of sequence coverage, spectral counts and unique peptide sequences obtained in the 15 fractions excised from in-gel fractionated high-morphine cultivar protein extracts. Depth of red shading is proportional to the maximum value in the column.

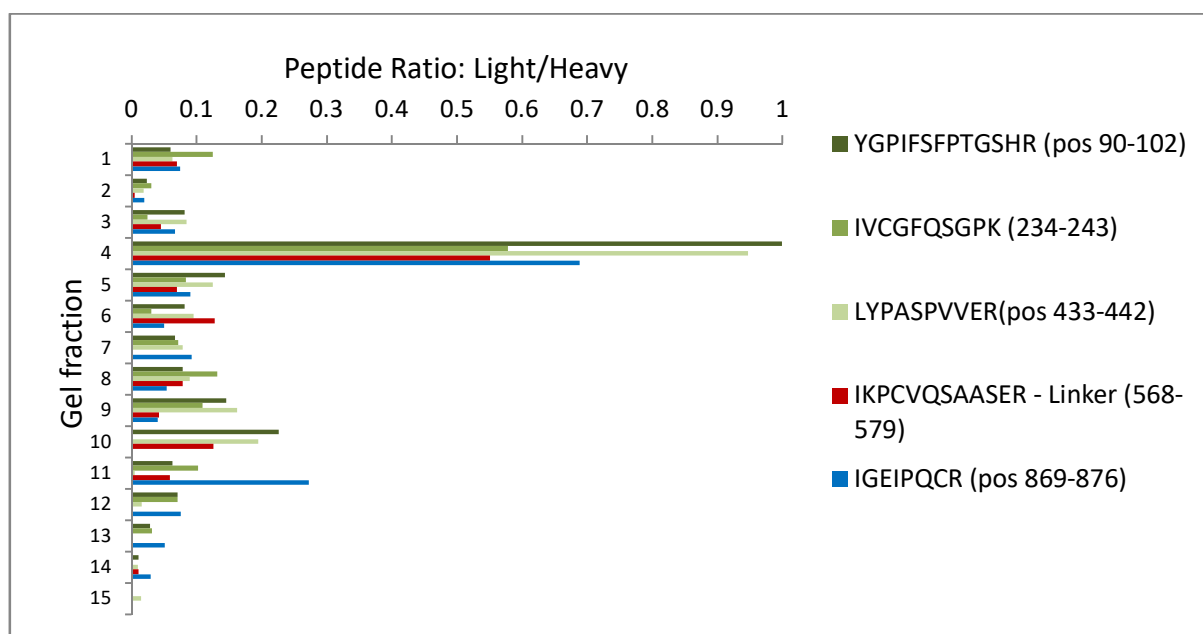
I explained to TW and MK that while the spectral counting approach to in-gel localisation (table 11) was consistent with a contiguous fusion there were imperfections in relative quantification between fractions. Spectral counting is reliant on duty cycle limitation providing a measure of quantification based on the number of times a peptide can be sampled during the acquisition. If the background complexity differs, as would be expected amongst 1D-PAGE fractions, the cycle time available to sample the targeted peptides will change, reducing quantitative accuracy. I further explained that to provide accurate relative inter-fraction quantification a common internal standard was required to normalise relative responses between fractions, the ideal standard being an isotopically heavy version of the STORR protein.

Subsequently, JC prepared a <sup>15</sup>N-labeled recombinant version of STORR protein. I ran the heavy protein into a 1D-PAGE gel, excised the band and digested as for the fractionated plant extract. I took aliquots of the remaining digest from fraction four and aliquoted a dilution series of the <sup>15</sup>N-labeled digest to determine the amount achieving an approximately equal ratio of light to heavy protein as measured by mass spectral response. I then spiked equal aliquots of the heavy digest into each 1D-PAGE fraction and repeated the original LC-MS/MS analysis.



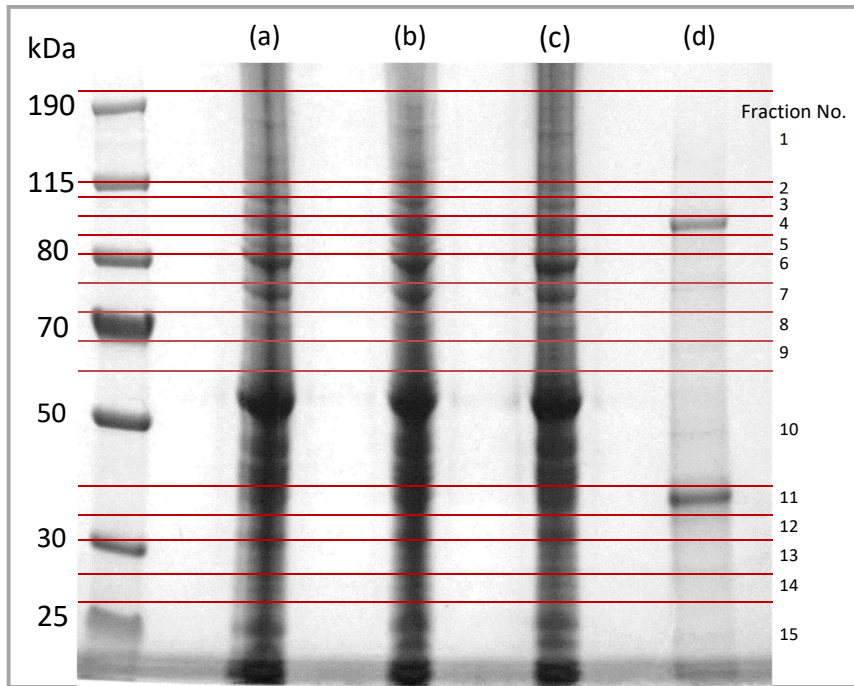
Five peptides were chosen for relative quantification of the fusion protein. These peptides spanned the sequence of the subunits and linker peptide. The quantified peptides were free of observed variable modifications such as oxidation or deamination in any of the light or heavy protein acquisition data sets. The peptides were also unique to the sequence of STORR. Relative peptide abundances between gel fractions were determined by comparing extracted ion chromatograms for MS peptide peak areas obtained from the endogenous and <sup>15</sup>N-labeled recombinant STORR digests (figure 26).

Results were consistent with the spectral counting analysis, showing a clear spike in abundance for all five quantified peptides in fraction 4, the expected mass of the fusion (figure 26).

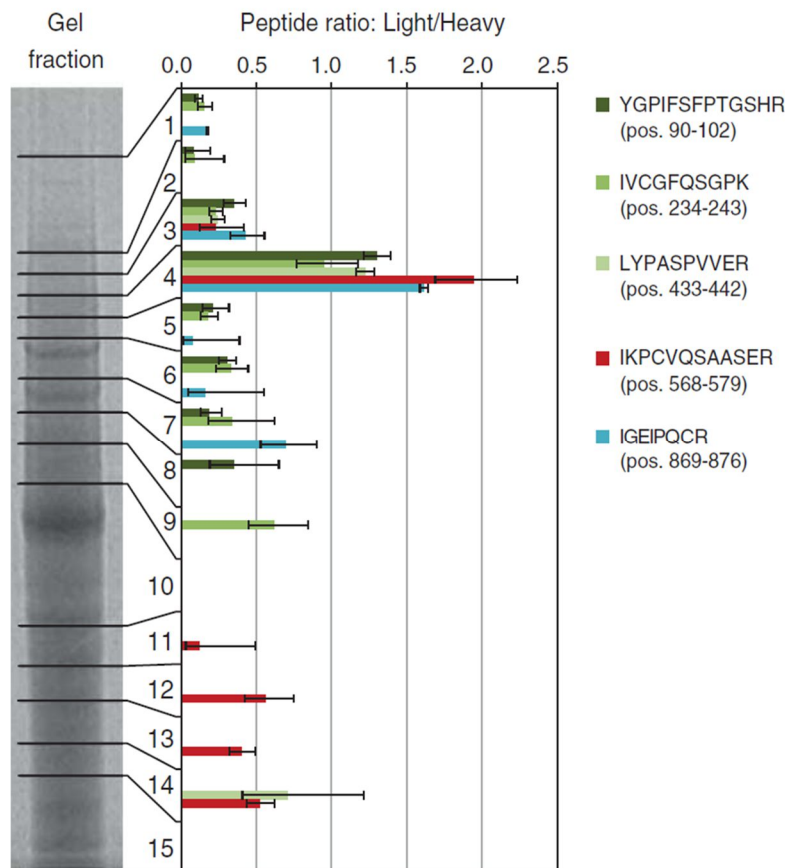


**Figure 26. Quantification of STORR peptides through 1D gel.** For relative quantification, a constant amount of a tryptic digest of <sup>15</sup>N-labeled recombinant STORR was spiked into the in-gel digest of each SDS-PAGE fraction before LC-MS/MS. Ratios of peak areas from extracted-ion chromatograms of light (endogenous) to heavy (labelled recombinant) versions of five peptides across the STORR protein sequence were compared. Ratios of normalized peak areas from extracted-ion chromatograms were converted to binary logarithms.

Upon peer review of the paper the editor requested the inter-fraction quantification be repeated to include biological replicates. TW produced three further high-morphine cultivar protein extracts, which I ran into a 1D-gel alongside an aliquot of the <sup>15</sup>N-labeled recombinant STORR protein (figure 27). Gel fractionation, digestion, spiking of heavy standard, LC-MS/MS acquisition and quantification data analysis were repeated as detailed in the paper. Normalised peak areas were converted to binary logarithms for calculation of means and standard errors of the mean. Figure 28 displays relative abundances of the five peptides monitored in each fraction.



**Figure 27.** 1D-PAGE of replicate high-morphine cultivar protein extracts (a-c), and <sup>15</sup>N-labeled recombinant STORR (d). Fractions from (a)-(c), as shown, were excised for tryptic digestion.



**Figure 28.** Quantification of STORR peptides through gel. For relative quantification, a constant amount of a tryptic digest of <sup>15</sup>N-labeled recombinant STORR was spiked into the in-gel digest of each SDS-PAGE fraction before LC-MS/MS. Ratios of peak areas from extracted-ion chromatograms of light (endogenous) to heavy (labelled recombinant) versions of five peptides from across the STORR protein sequence were compared. Ratios of normalized peak areas from extracted-ion chromatograms were converted to binary logarithms for the calculation of means and standard errors of the mean. Only measurements where the respective peptides were found in all three biological replicates are shown.

Highest relative abundance for all five peptides is in fraction four, around 90-105 kDa, which fits the predicted size of the STORR fusion protein (100.5 kDa) and where the <sup>15</sup>N-labeled recombinant version runs. It is likely that the resolution of the gel was insufficient to confine the protein to a single fraction with the amount of material loaded, although the presence of hydrolytic products cannot be discounted. Gel resolution may have been improved by loading less material. The gel quantification produces no evidence to support individual expression of the two proteins in un-fused forms, suggesting if the individual forms are present they are of low abundance.

Absolute quantification through the gel was not possible because the absolute amount of <sup>15</sup>N-labeled recombinant STORR was unknown. As shown in figure 27, the purified protein (d) contained a significant product around 35 kDa that restricted quantification of the recombinant product. Analysis by LC-MS/MS of the 35 kDa band in the recombinant expressed sample showed it to be the N-terminal portion of cytochrome P450 monooxygenase, containing the His-tag. Although further purification could be attempted were absolute quantification required, relative quantification through the gel was sufficient for size determination.

While my mass spectrometry analysis provides strong evidence for STORR being an expressed fusion, it is limited by the ambiguities of a bottom-up approach. Peptides spanning the length of the sequences are derived from the same gel fraction but there is no direct evidence that they result from a single proteoform, as following digestion peptides are no longer linked. The ideal analysis would be to purify the intact protein so that it could be observed as a continuous full length sequence in the mass spectrometer. Had an antibody been available this may have been possible using immune precipitation<sup>156</sup> or pull-down approaches.<sup>157</sup> In the absence of an antibody, intact protein LC-MS could be attempted,<sup>85,86</sup> but without further fractionation the complexity of the sample would make analysis difficult. The use of 2D-FT-ICR-MS/MS, where ICR pulse sequences are used to align product and precursor ions in co-fragmented mixtures, is an intriguing option with vast potential.<sup>158</sup> However, the technique is still in its infancy, with few practitioners and has only been published for relatively simple mixtures to date.

My contribution was vital to the publication of the paper as evidenced by the editor's requirement for the proteomics to be repeated with biological replicates to accept the manuscript. The work has been published in Science, has already been cited 19 times<sup>159-178</sup> and clearly has very high potential impact.

## Conclusion

The presented work demonstrates my use of mass spectrometry to: identify post translational modifications of histones (paper 1); employ a label free emPAI methodology to study the immunomodulation molecules used by two endemic parasites to modify host immune response (papers 2-4); quantify the changes in immune responses of dendritic cells (paper 5); identify a novel co-factor in uridine-5-oxylacetic acid biosynthesis (paper 6) and demonstrate the fusion protein structure of the remaining unidentified enzyme in the morphinan biosynthetic pathway (paper 7). While varied in their applications my proteomics analyses were vital to each of these studies. As a collection and in combination with this integrative chapter the work demonstrates both the power of mass spectrometry-based proteomics and my significant contribution to the field.

# Appendices

Appendices 1A-7C are the submitted papers and declarations from co-authors attesting to my contributions.

## Appendix 1A

Submitted paper 1 – Dynamics of plant histone modifications in response to DNA damage.

# Dynamics of plant histone modifications in response to DNA damage

Georgina E. DRURY\*, Adam A. DOWLE†, David A. ASHFORD†, Wanda M. WATERWORTH\*, Jerry THOMAS† and Christopher E. WEST<sup>1</sup>\*

\*Centre for Plant Sciences, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, U.K., and †Proteomics Technology Facility, Department of Biology, University of York, York YO10 5DD, U.K.

DNA damage detection and repair take place in the context of chromatin, and histone proteins play important roles in these events. Post-translational modifications of histone proteins are involved in repair and DNA damage signalling processes in response to genotoxic stresses. In particular, acetylation of histones H3 and H4 plays an important role in the mammalian and yeast DNA damage response and survival under genotoxic stress. However, the role of post-translational modifications to histones during the plant DNA damage response is currently poorly understood. Several different acetylated H3 and H4 N-terminal peptides following X-ray treatment were identified using MS analysis of purified histones, revealing previously unseen

patterns of histone acetylation in *Arabidopsis*. Immunoblot analysis revealed an increase in the relative abundance of the H3 acetylated N-terminus, and a global decrease in hyperacetylation of H4 in response to DNA damage induced by X-rays. Conversely, mutants in the key DNA damage signalling factor ATM (ATAXIA TELANGIECTASIA MUTATED) display increased histone acetylation upon irradiation, linking the DNA damage response with dynamic changes in histone modification in plants.

Key words: *Arabidopsis*, DNA double-strand break, DNA repair, epigenetics, recombination.

## INTRODUCTION

Repair of DNA DSBs (double-strand breaks), one of the most cytotoxic forms of DNA damage, takes place in the context of chromatin. Histones form the protein component of chromatin, and play critical roles in all aspects of DNA metabolism and genome maintenance. A total of eight histone proteins form the core nucleosome particle, around which 147 bp of DNA wraps in a left-handed superhelix to form the lowest-order structure of chromatin [1]. Each nucleosome is composed of two each of H2A, H2B, H3 and H4, 11–22 kDa proteins, in a tripartite assembly with a central (H3/H4)<sub>2</sub> tetramer flanked by two H2A/H2B dimers. Localized to the outer surface of the assembled nucleosome octamer, the N-terminal tails of H3 and H4 are rich in basic amino acids which confer a positive charge and high-affinity binding with negatively charged DNA [2].

Histones are subject to post-translational modification which residue-specific modification determines chromatin structure or function, including acetylation, phosphorylation, methylation, ubiquitination, SUMOylation and ADP-ribosylation. Post-translational histone modifications contribute to the varied and dynamic configuration of chromatin [3]. Post-translational modifications also orchestrate interaction of the histone tails with other proteins, as combinations of multiple modifications impart a histone ‘code’ that signals specifically to interacting protein partners. Chromatin configuration is intrinsically linked to gene expression and is tightly regulated during growth and development, displaying dynamic responses to changes in the environment [4].

Chromatin remodelling activities have been shown to be important in both the signalling and repair of DSBs in mammalian and yeast systems [5]. However, our understanding of chromatin modifications and their importance in plant DNA repair is limited to the phosphorylation of S<sup>139</sup> (single-letter code has

been used for amino acids) of the histone variant H2AX by the DSB signalling kinase ATM (ATAXIA TELANGIECTASIA MUTATED) [6]. This is a well-characterized and highly conserved response to induction of DNA DSBs, dependent on upstream MRN [MRE11 (meiotic recombination 11)–RAD50 (radiation sensitive 50)–NBS1 (Nijmegen breakage syndrome 1)] complex signalling in plants, yeast and animals [7,8]. In animals, H2AX phosphorylation has been shown to function in establishing stable repair foci at the site of DNA DSBs [9].

Chromatin remodelling during DNA repair in yeast and mammals involves hyperacetylation of H3 and H4 as early epigenetic changes to chromatin following DNA damage [5,10]. In addition to altering localized chromatin structure, these modifications function in DNA repair signalling pathways; histone H3 acetylation on K<sup>14</sup> (H3K14 acetylation) stimulates autophosphorylation of ATM [11]. In contrast, there are no data concerning histone acetylation in the plant response to DSBs, although acetylation has been reported after UV-treatment of plants, which predominantly induces formation of cyclobutane pyrimidine dimers and 6-4 photoproducts. UVB-tolerant maize lines showed high levels of covalent changes to histones compared with the UVB-sensitive lines. Hyperacetylation of H3 (diacetyl) and H4 (tetracetyl) N-termini in UVB-tolerant lines were enriched in the promoter and transcribed regions of genes associated with UVB response, implicating acetylation in the transcriptional response to UV-induced DNA damage [12].

Previous studies utilized MS analysis of plant histones in *Arabidopsis* [13,14], alfalfa [15] and soybean [16] to identify methylated and acetylated histone isoforms. In the present study we have used a LC (liquid chromatography)–MS/MS (tandem MS) approach to analyse post-translational modifications that are present on H3 and H4 following DNA-damage induction by IR (ionizing radiation), which results in a range of damage products including SSBs (single-strand breaks) and DSBs. We identify

Abbreviations used: ac, acetylated; ATM, ATAXIA TELANGIECTASIA MUTATED; CID, collision-induced dissociation; DDR, DNA damage response; DSB, double-strand break; DTT, dithiothreitol; HDAC, histone deacetylase; IPI, International Protein Index; IR, ionizing radiation; LC, liquid chromatography; me, methylated; MS/MS, tandem MS; TCA, trichloroacetic acid; TFA, trifluoroacetic acid.

<sup>1</sup> To whom correspondence should be addressed (email c.e.west@leeds.ac.uk).

several novel forms of acetylated H3 and H4 and demonstrate dynamic changes in the relative abundance of specific histone modifications upon induction of DNA damage in *Arabidopsis*.

## EXPERIMENTAL

### Plant growth

Wild-type Columbia-0 seeds were sterilized in 70% ethanol for 5 min, and resuspended in sterile 0.1% agar. Seeds were plated individually in Murashige and Skoog (Sigma) medium (20 g/l sucrose, 0.5 g/l Mes and 8 g of phytoagar; Duchefa). After a cold treatment of 2 days at 4°C, plates were transferred to a growth chamber (16 h of light and 8 h of dark cycle) and grown for 2 weeks before harvesting. IR treatment was delivered using a 320 kV X-ray irradiation system (NDT Equipment Services).

### Nuclei isolation

Seedlings were ground in liquid nitrogen to a fine powder and further homogenized in homogenization buffer [25 mM Pipes (pH 7.0), 10 mM NaCl, 5 mM EDTA (pH 8.0), 250 mM sucrose, 0.5% Triton X-100, 20 mM 2-mercaptoethanol, 0.2 mM PMSF and 1/50 ml protease inhibitor tablet (Roche catalogue number 04693132001)] for approximately 1 min in a Waring blender. The homogenate was filtered through 300 µm, 100 µm and 56 µm Sefar Nitex membrane layers and the filtrate was centrifuged at 4600 g for 15 min at 4°C and the supernatant removed. The pellet was then washed in homogenization buffer and centrifuged at 1000 g for 10 min at 4°C a further two times. The nuclear pellet was then resuspended in nuclei storage buffer [50 mM Hepes (pH 7.6), 110 mM KCl, 5 mM MgCl<sub>2</sub>, 50% (v/v) glycerol and 1 mM DTT (dithiothreitol)]. For protein samples used in immunoblotting, the nuclear pellet after the first centrifugation was suspended in nuclei electrophoresis buffer [5% SDS, 0.5 M Tris (pH 6.8) and 1 mM DTT].

### Acid extraction of histones

Nuclei were centrifuged at 1000 g for 10 min at 4°C and the supernatant removed. The remaining pellet was resuspended in prechilled extraction buffer (0.5 M HCl and 0.2 mM PMSF) and vortexed. Samples were placed on a rotator at 4°C for between 30 min and overnight. Samples were then centrifuged at 16000 g for 10 min at 4°C, and the supernatant removed and mixed 2:1 with prechilled 100% TCA (trichloroacetic acid)/water (final TCA concentration was 33%). Samples were mixed well and placed on ice for between 30 min and overnight. Samples were centrifuged at 16000 g for 10 min at 4°C and the supernatant poured off. The pellet was washed twice with 100% prechilled acetone and centrifuged at 16000 g for 5 min and then left to dry. Samples were stored at -80°C.

### Tricine-SDS/PAGE gel electrophoresis

Protein samples were separated using 10% acrylamide/bisacrylamide (19:1) Tricine gels containing ethylene glycol (30%, v/v) according to [17] using the Protean II xi Cell electrophoresis system (catalogue number 165-1834, Bio-Rad Laboratories). Protein samples were dissolved in Laemmli buffer (5% SDS, 5% 2-mercaptoethanol, 10% glycerol, 0.1 M Tris (pH 6.8) and 0.1% Bromophenol Blue) and denatured at 90°C for 10 min, and loaded on to the gel alongside Precision Xtra Plus protein marker (161-0377, Bio-Rad Laboratories) and calf thymus histones (catalogue number 10223565001, Roche). Proteins were electrophoresed at 100 V for 24 h using Tricine electrophoresis buffer (catalogue number B48, Fermentas). Proteins were stained

with Instant Coomassie (Expedeon ISB01L) for between 30 min and 1 h. The protein bands were excised and overlaid with 20% acetic acid and 5% methanol in 1.5 ml microfuge tubes. For immunoblotting, protein samples were separated on 16% acrylamide/bisacrylamide (19:1) minigels and electrophoresed for 2.5 h at 150 V.

### Western blotting

Protein concentrations in resuspended nuclear pellets were quantified using the Lowry protein assay (Bio-Rad Laboratories) with BSA as a standard. After electrophoresis, proteins were transferred on to a nitrocellulose membrane (catalogue number 10402495, Whatman) for 1 h at 100 V. The blots were probed with the following rabbit polyclonal primary antisera: anti-H3 (1:10000 dilution; catalogue number H0164, Sigma); anti-H3K18acK23ac (1:10000 dilution; catalogue number 17-615, Millipore); anti-H4K5acK8acK12acK16ac (1:5000 dilution catalogue number 06-598, Millipore); anti-H4K8ac (1:10000 dilution; catalogue number 06-760-MN, Millipore); anti-H4K12ac (1:5000 dilution; catalogue number 06-761-MN, Millipore); or anti-H4K16ac (1:10000 dilution; catalogue number 06-762-MN, Millipore). Immune complexes were detected by alkaline-phosphatase-conjugated anti-(rabbit IgG) (1:30000 dilution; catalogue number A2556, Sigma-Aldrich) and developed in BCIP (5-bromo-4-chloroindol-3-yl phosphate)/NBT (Nitro Blue Tetrazolium) solution (Invitrogen).

### Arg-C digestion

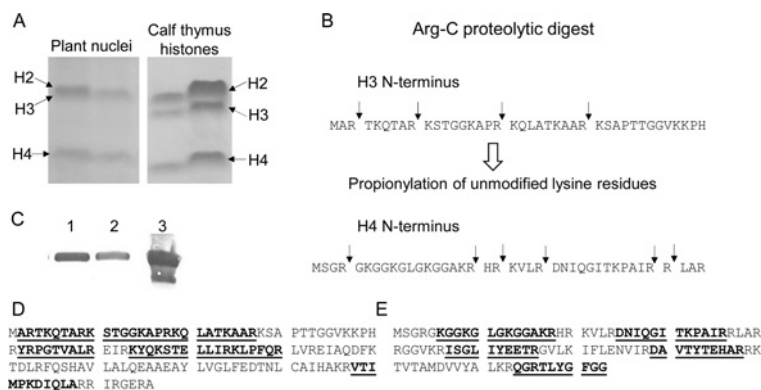
In-gel Arg-C digestion was performed after reduction with DTE (dithioerythritol) and S-carbamidomethylation with iodoacetamide. Gel pieces were washed twice with 50% (v/v) aqueous acetonitrile containing 25 mM ammonium bicarbonate, then once with acetonitrile and dried in a vacuum concentrator for 20 min. Gel pieces were rehydrated by adding 10 µl of 0.02 g/l Arg-C solution in 25 mM ammonium bicarbonate, and after 10 min sufficient 25 mM ammonium bicarbonate solution was added to cover the gel pieces. Digests were incubated overnight at 37°C.

### Propionylation

The digest supernatant was removed and adjusted to pH 10 by the addition of aqueous ammonium hydroxide, followed by the addition of 10 µl of propionic anhydride and incubation at 60°C for 1 h. Post-propionylation the samples were acidified with TFA (trifluoroacetic acid), dried down and reconstituted in 0.1% aqueous TFA.

### MS

HPLC was performed using a nanoAcquity UPLC system (Waters) equipped with a nanoAcquity Symmetry C18, 5 µm trap (180 µm×20 mm; Waters) and a nanoAcquity BEH130 1.7 µm C18 capillary column (75 m×250 mm; Waters). The trap wash solvent was 0.1% aqueous formic acid and the trapping flow rate was 10 µl/min. The trap was washed for 5 min after sample loading before switching flow to the capillary column. The separation used a gradient elution of two solvents: solvent A (0.1% formic acid) and solvent B (acetonitrile containing 0.1% formic acid). The flow rate for the capillary column was 300 nl/min, column temperature was 60°C and the gradient profile was as follows: initial conditions 5% solvent B (2 min), followed by a linear gradient to 35% solvent B over 20 min and then a wash with 95% solvent



**Figure 1** Isolation of histones and peptides for identification by MS

Histones were isolated from *Arabidopsis* seedlings and analysed by peptide digestion and MS/MS. **(A)** Coomassie Blue-stained Tricine gel. **(B)** Schematic overview of preparation of H3 and H4 N-terminal peptides for LC-MS/MS. **(C)** Immunoblot with H3 antisera. Lane 1 and 2, plant histones; lane 3, calf thymus. **(D)** H3 amino acid sequence with identified peptides in bold and underlined. **(E)** H4 amino acid sequence with identified peptides indicated as bold underlined text.

B for 2.5 min. The column was returned to initial conditions and re-equilibrated for 25 min before subsequent injections.

The nanoLC system was interfaced to a maXis UHR-QTOF (ultrahigh-resolution quadrupole-time-of-flight) mass spectrometer (Bruker Daltonics) with a nano-electrospray source fitted with a steel emitter needle [180  $\mu\text{m}$  outside diameter  $\times$  30  $\mu\text{m}$  inside diameter; Proxeon). Positive-ion MS and CID (collision-induced dissociation)-MS/MS spectra were acquired using AutoMSMS mode. Instrument control, data acquisition and processing were performed using Compass 1.3 SR3 software (micrOTOF control, Hystar and DataAnalysis, Bruker Daltonics). Instrument settings were: ion spray voltage, 1500 V; dry gas, 6 l/min; dry gas temperature, 160  $^{\circ}\text{C}$ ; and ion acquisition range  $m/z$ , 50–2200. AutoMSMS settings were for MS 0.5 s (acquisition of survey spectrum) and for MS/MS (CID with  $\text{N}_2$  as collision gas) ion acquisition range  $m/z$ , 350–1400; 0.1 s acquisition for precursor intensities above 100 000 counts; for signals of lower intensities down to 1000 counts acquisition time increased linearly to 1.5 s; the collision energy and isolation width settings were automatically calculated using the AutoMSMS fragmentation table; three precursor ions; absolute threshold 1000 counts; preferred charge states; and 24 singly charged ions excluded. Two MS/MS spectra were acquired for each precursor and previously selected target ions were excluded for 60 s.

### Database searching

MS/MS data were submitted to database searching using a locally running copy of the Mascot program (version 2.3, Matrix Science), through the Bruker ProteinScape interface (version 2.1). Spectra were searched against the IPI (International Protein Index) *Arabidopsis* database (37150 sequences and 14876683 residues). The search criteria specified were: enzyme, Arg-C; fixed modifications, carbamidomethyl (C); variable modifications, acetylation (K), oxidation (M), propionylation (K, N-term); peptide tolerance, 10 ppm; MS/MS tolerance, 0.1 Da. Peptides with an expect value of 0.05 or lower were considered significant.

## RESULTS

### Isolation of *Arabidopsis* histones and preparation for LC-MS/MS

Histones isolated from *Arabidopsis* seedling tissue and separated by electrophoresis were identified by immunoblotting with H3 antisera (Figures 1A and 1B). Discreet bands of H4 and H3

were excised from the gel and samples were reduced, alkylated and digested in-gel with Arg-C protease rather than trypsin which produces suboptimal histone peptides for MS due to their high lysine and arginine content. Endoproteinase Arg-C cleaves at arginine residues only, producing histone peptides of 3–14 residues in length (Figure 1C). Propionylation of H3 and H4 peptides was carried out to modify the  $\epsilon$ -amino groups of lysine residues and the N-termini to increase the hydrophobicity and retention of peptides on the C18 column prior to MS/MS. Following LC-MS/MS, the spectra were searched against the IPI *Arabidopsis* database, and the expected histone was the top-scoring match in each instance. We obtained sequence coverage of 45% and 54% of H3 and H4 respectively, with the N-termini of both H3 and H4 being well covered (Figures 1D and 1E).

### Identification of acetylated lysines in histones H3 and H4 following DNA damage

Acetylation of lysine residues has been shown to be both involved in the DDR (DNA damage response) in mammals and yeast [18,19] and in the transcriptional response to UV-irradiation in maize [12]. To investigate global acetylation patterns after DNA damage in *Arabidopsis*, we investigated the acetylation of H3 and H4 N-termini following a 160 Gy dose of X-rays. Good coverage of the N-terminal regions of both H3 and H4 was achieved by LC-MS/MS and Mascot searching to identify peptides (Figures 1D and 1E). Several criteria were used to confirm the assignment of sites of acetylation within peptides. Acetylation could readily be distinguished from trimethylation, which gives a mass increase that is 36 mDa higher than acetylation, using the high mass accuracy of the Bruker maXis. Low mDa differences between measured and calculated peptide masses were obtained (Table 1), which were consistent with acetylation and ruled out trimethylation. Mascot assignments of acetylation positions along with the spectra were submitted to the Prophossi spectrum annotation tool (<http://www.compbio.dundee.ac.uk/prophossi/bin/prophossi.cgi.pl>) [20], which labels spectra with product ions, and also matches all low-abundance and doubly charged product ions. Finally, manual inspection was performed to ensure the presence of crucial b- and y-type product ions in the MS/MS spectra (Figures 2 and 3).

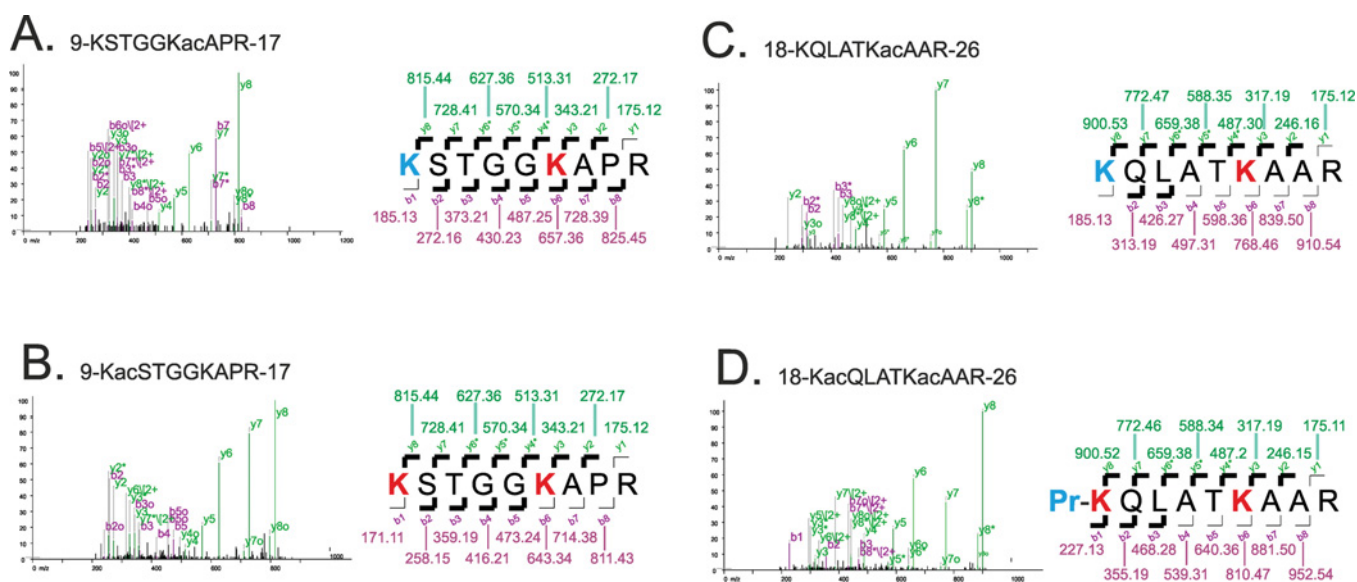
Acetylated peptide isoforms for two N-terminal H3 peptides, K<sup>9</sup>STGGKAPR<sup>17</sup> and K<sup>18</sup>QLATKAAR<sup>26</sup>, and one N-terminal H4 peptide, <sup>4</sup>GKGGKGLGKGGAKR<sup>17</sup>, were identified in DNA-damage-treated plants (Table 1). Both mono- and di-acetylated



**Table 1** Acetylated peptides from H3 and H4 N-termini identified by LC-CID-MS/MS

Acetylated lysines are indicated by Kac. Ion scores and expect values are from Mascot. Measured peptide masses are monoisotopic values calculated from masses of doubly charged precursor ions that can include propionyl groups (see Figure 3). Mass errors are differences between measured and theoretical masses.

Histone	Peptide	Acetylation sites	Ion score	Expect value	Measured mass (Da)	Mass error (mDa)
H3	K <sup>9</sup> STGGKacAPR <sup>17</sup>	K14	80	$2.0 \times 10^{-7}$	998.551	5
H3	K <sup>9</sup> acSTGGKacAPR <sup>17</sup>	K9 K14	51	$1.6 \times 10^{-4}$	984.535	6
H3	K <sup>18</sup> QLATKacAAR <sup>26</sup>	K23	58	$3.3 \times 10^{-5}$	1083.640	4
H3	K <sup>18</sup> acQLATKacAAR <sup>26</sup>	K18 K23	63	$9.8 \times 10^{-6}$	1125.651	8
H4	G <sup>4</sup> KGGKacGLGKGGAKR <sup>17</sup>	K8	55	$5.0 \times 10^{-5}$	1535.879	1
H4	G <sup>4</sup> KGGKGLGKGGAKacR <sup>17</sup>	K16	80	$1.5 \times 10^{-7}$	1423.832	6
H4	G <sup>4</sup> KacGGKGLGKGGAKacR <sup>17</sup>	K5 K16	60	$1.2 \times 10^{-5}$	1521.864	2
H4	G <sup>4</sup> KGGKacGLGKGGAKacR <sup>17</sup>	K8 K16	70	$1.3 \times 10^{-6}$	1465.843	7
H4	G <sup>4</sup> KGGKGLGKacGGAKacR <sup>17</sup>	K12 K16	63	$6.4 \times 10^{-6}$	1465.841	5
H4	G <sup>4</sup> KacGGKGLGKacGGAKacR <sup>17</sup>	K5 K12 K16	84	$5.2 \times 10^{-8}$	1451.820	<1
H4	G <sup>4</sup> KGGKacGLGKacGGAKacR <sup>17</sup>	K8 K12 K16	46	$3.0 \times 10^{-4}$	1507.849	2
H4	G <sup>4</sup> KacGGKacGLGKacGGAKacR <sup>17</sup>	K5 K8 K12 K16	67	$2.7 \times 10^{-6}$	1493.829	3

**Figure 2** Identification of mono- and di-acetylated peptides from histone H3

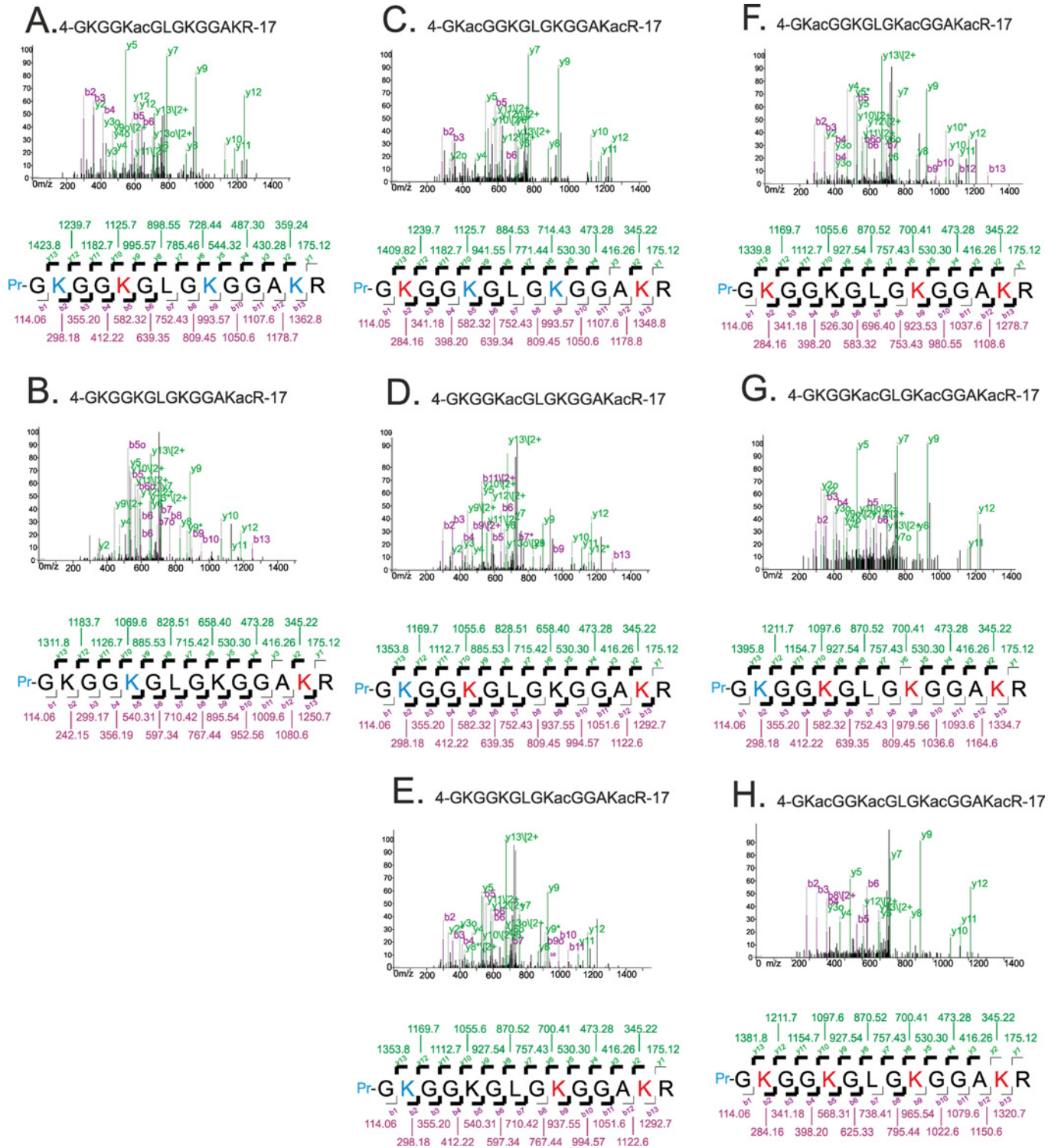
CID-MS/MS spectra were redrawn and annotated using ProPhoSI (<http://www.compbio.dundee.ac.uk/prophosi/bin/prophosi.cgi.pl>). Precursor ion  $m/z$  values and measured masses are indicated; all precursor ions were doubly charged. A bold black line indicates that a b- (purple) or y-ion (green) is observable in the spectrum. Acetylated lysine residues are in red; propionylated lysine residues are in blue; and N-terminal propionylation is labelled Pr.

forms of the two H3 peptides were present in addition to the non-acetylated forms. Their presence indicates that each of the lysines K<sup>9</sup>, K<sup>14</sup>, K<sup>18</sup> and K<sup>18</sup>K<sup>23</sup> were incompletely acetylated in histone H3. In this analysis, acetylation of K<sup>9</sup> is only observed in combination with K<sup>14</sup> acetylation, and likewise, acetylation of K<sup>18</sup> is only identified where K<sup>23</sup> is acetylated. In the case of the single N-terminal H4 peptide, G<sup>4</sup>KGGKGLGKGGAKR<sup>17</sup>, there were eight acetylated isoforms identified (two monoacetylated, three diacetylated, two triacetylated and one tetra-acetylated), which demonstrated acetylation of lysines K<sup>5</sup>, K<sup>8</sup>, K<sup>12</sup> and K<sup>16</sup> in histone H4 in different combinations and all four simultaneously. Of the eight different acetylated forms of the H4 N-terminal peptide isolated from irradiated *Arabidopsis*, CID-MS/MS evidence for only the tetra-acetylated isoform in higher plants has been published [12,21].

The MS/MS evidence for monoacetylated K<sup>9</sup>STGGKacAPR<sup>17</sup> (where ac indicates an acetylated residue) from histone H3 is very strong, with both b- and y-ion series observing that they lack only a single ion each (Figure 2A). The spectrum

of K<sup>9</sup>acSTGGKacAPR<sup>17</sup> also contains a y-ion series that lacks only one ion (the y<sub>1</sub>), plus a series of four b-ions (Figure 2B). The evidence for K<sup>18</sup>QLATKacAAR<sup>26</sup> and K<sup>18</sup>acQLATKacAAR<sup>26</sup> is also strong, with both y-ion series missing only the y<sub>1</sub> ion (Figures 2C and 2D).

For the two monoacetylated G<sup>4</sup>KGGKGLGKGGAKR<sup>17</sup> H4 peptides, convincing evidence for the assigned sites of acetylation was contained in the extensive b- and y-ion series observed (Figures 3A and 3B). Likewise, spectra of all three of the diacetylated peptides contain y-ion series that are complete except for the y<sub>1</sub> ion (Figures 3D and 3E), and two also contain extensive b-ion series (Figures 3D and 3E). Assignment of the sites of acetylation in G<sup>4</sup>KGGKacGLGKGGAKacR<sup>17</sup> is supported by the prominent y<sub>5</sub>, y<sub>7</sub> and y<sub>12</sub> ions at  $m/z$  530, 771 and 1409 respectively (Figure 3C). Two triacetylated histone H4 peptides, G<sup>4</sup>KGGKacGLGKGGAKacR<sup>17</sup> and G<sup>4</sup>KacGGKGLGKacGGAKacR<sup>17</sup>, were also identified by Mascot with expect values of  $5.2 \times 10^{-8}$  and  $3.0 \times 10^{-4}$  respectively (Table 1). The y-ion series for these two peptides are



**Figure 3** Identification of mono-, di-, tri- and tetra-acetylated peptides from histone H4

CID-MS/MS spectra were redrawn and annotated using ProPhoSI (<http://www.compbio.dundee.ac.uk/prophossi/bin/prophossi.cgi.pl>). Precursor ion  $m/z$  values and measured masses are indicated; all precursor ions were doubly charged. A bold black line indicates that a b- (purple) or y-ion (green) is observable in the spectrum. Acetylated lysine residues are in red; propionylated lysine residues are in blue; and N-terminal propionylation is labelled Pr.

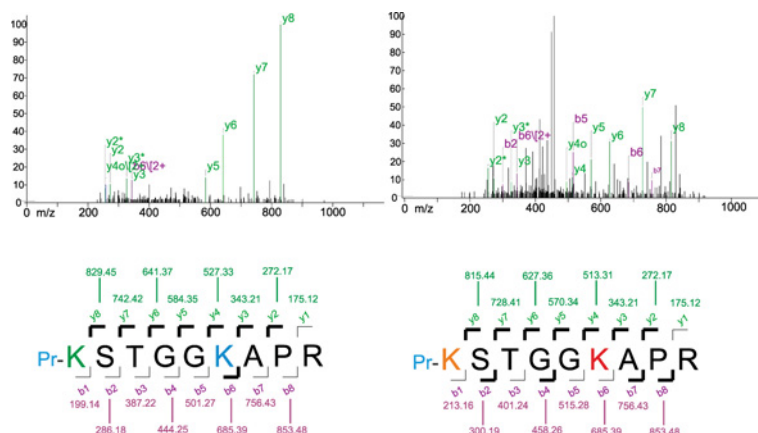
nearly complete, and the N-terminal b-ion series that are also present confirm the absence of acetylation of K<sup>8</sup> or K<sup>5</sup> in the two isoforms (Figures 3F and 3G respectively). The hyperacetylated form with acetyl modifications present at each of the four lysine residues (G<sup>4</sup>KacGGKacGLGKacGGAKacR<sup>17</sup>),

has previously been identified in UV-irradiated maize by Casati et al. [12]. Their published CID-MS/MS spectrum is very similar to that of the N-terminally propionylated version we obtained, which contains a nearly complete series of y-ions, plus a series of five b-ions from the N-terminus (Figure 3H).

**Table 2 Methylated peptide isoforms from H3 and H4 N-termini identified by LC-CID-MS/MS**

Acetylated lysine is indicated by Kac and dimethylated lysine by Kme<sub>2</sub>. Ion scores and expect values are from Mascot. Measured peptide masses are monoisotopic values calculated from masses of doubly charged precursor ions that include propionyl groups (see Figure 4). Mass errors are differences between measured and theoretical masses.

Histone	Peptide	Modification sites	Ion score	Expect value	Measured mass (Da)	Mass error (mDa)
H3	K <sup>9</sup> meSTGGKAPR <sup>17</sup>	K9	41	$2.8 \times 10^{-4}$	1026.587	5
H3	K <sup>9</sup> me <sub>2</sub> STGGKacAPR <sup>17</sup>	K9 K14	35	$7.8 \times 10^{-3}$	1026.581	1

**Figure 4 Identification of methylated and acetylated peptides from histone H3**

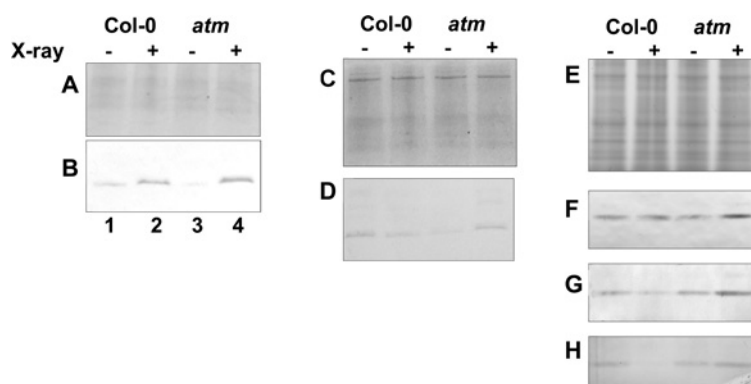
CID-MS/MS spectra were redrawn and annotated using ProPhoSI (<http://www.compbio.dundee.ac.uk/prophosi/bin/prophosi.cgi.pl>). Precursor ion *m/z* values and measured masses are indicated; all precursor ions were doubly charged. A bold black line indicates that a b- (purple) or y- (green) ion is observable in the spectrum. Acetylated lysine residues are in red; propionylated lysine residues are in blue; dimethylated lysine residues are in orange; and N-terminal propionylation is labelled Pr.

### Identification of methylated lysines in histone H3 following DNA damage

Methylation of lysine residues can be associated with gene silencing, for example, H3K9 methylation is necessary for silencing and the dimethylated form is associated with regions of heterochromatin, whereas trimethylated H3K27 and H3K9 are both found in euchromatin [3,14,22]. The highest levels of methylation were identified in the histone H3 isoform H3.1, including monomethylation of lysine residues K9 and K27, consistent with the role of H3.1 in packaging heterochromatin [13]. Analysis of methylated lysine residues in a peptide sequence was manually validated by confirming the presence of the diagnostic b- and y-type product ions in the MS/MS spectrum. The mass difference between b- and y-type product ions indicating trimethylation will result in addition of 42.047 to the mass of lysine, whereas acetylation will only add 42.011 Da. The accuracy of the mass measurements in the present study was sufficient to discriminate the 36 mDa mass difference between these two modifications (Table 2). We identified two unique methylated peptides of H3 from irradiated plants. One isoform was dimethylated at K9, which is enriched in heterochromatic regions [22], was concomitantly acetylated at K14 (K<sup>9</sup>me<sub>2</sub>STGGKacAPR<sup>17</sup>; where me indicates a methylated residue); the assignment is supported by a series of y-ions that lacks only the y<sub>1</sub> ion, plus intense b<sub>4</sub> and b<sub>7</sub> ions (Figure 4A). The other was monomethylated at K9 (K<sup>9</sup>meSTGGKAPR<sup>17</sup>); the assignment is supported by a series of y-ions that lacks only the y<sub>1</sub> ion (Figure 4B).

### Increase in acetylation of the H3 N-terminus following DNA damage

Previous reports indicate that the diacetylated forms for histone H3 K9acK14ac and K18acK23ac constitute some of the lowest abundance detectable modifications in plants [13]. Monoacetylated H3 peptides varied in relative abundance, K14ac being most abundant, then K18K23, whereas K9ac was not detected [13]. In the present study, four acetylated peptides on the H3 N-terminus were identified. The Arg-C digest after propionylation generated two peptides, K<sup>9</sup>STGGKAPR<sup>17</sup> and K<sup>18</sup>QLATKAAR<sup>26</sup>, which display acetylation at each of the four lysine residues (K9, K14, K18 and K23). However, our analysis does not reveal whether residues K14 and K18 are both acetylated within a single molecule of histone H3. The hyperacetylation of the H3 N-terminus was observed in response to DNA damage in mammals by the use of antisera specific to each lysine modification [23], but this analysis also did not reveal whether concurrent acetylation of multiple residues occurs in the same molecule. To further characterize hyperacetylation following X-ray treatment in *Arabidopsis*, a commercially available antibody to K14acK18ac was used to investigate whether K14 and K18, the residues spanning the two peptide fragments generated by the Arg-C digest, were both acetylated in irradiated tissue by immunoblotting. The antibody was used to detect differences in relative abundance of lysine modifications in protein samples from irradiated and unirradiated tissues (Figure 5A). Diacetylation of K14 and K18 was detected in all tissues, and was shown to increased abundance following IR in Col-0 (Figure 5B). Acetylation of K14 and K18 is therefore detectable



**Figure 5** Relative abundance of H3K14acK18ac and acetylated forms of H4 N-terminus increases following IR

(A) Coomassie Blue-stained Tricine-SDS/PAGE gel. (B) Western blot analysis using H3K14acK18ac. (C) Coomassie Blue-stained Tricine-SDS/PAGE gel. (D) Immunoblot for tetra-acetylated H4K5acK8acK12acK16ac. (E) Coomassie Blue-stained gel (F). Western blots with anti-H4K5ac (G), anti-H4K12ac (H), anti-H4K16ac antibodies. Lane 1, Col-0 unirradiated; lane 2, Col-0 160 Gy IR; lane 3, *atm-3* unirradiated; and lane 4, *atm-3* 160 Gy IR.

as an epigenetic mark present on histones in the absence of genotoxic stress that increases in response to DSB induction. A similar response was shown in the *atm*-null mutant background; however, a band of greater intensity is present following IR treatment, indicating that acetylation persists in the absence of a fully functioning DDR, and suggests that diacetylation of K14 and K18 occurs independently of ATM-mediated signalling.

#### Loss of tetra-acetylated H4 tail following DNA damage: K16 and K12 most likely targets for deacetylation

Hyperacetylation of the H4 tail has been shown to localize to transcription start sites of loci implicated in the DDR in maize [12], and be an epigenetic modification at DSB sites in studies with mammalian cells undergoing DDR [24]. Our MS analysis of *Arabidopsis* histones isolated from tissues undergoing repair of DSBs following X-ray treatment revealed the tetra-acetylated form was present, along with seven other isoforms. The tetra-acetylated peptide was found to decrease in abundance following X-ray treatment in Col-0, but increase in abundance in the *atm*-null mutant background, again illustrating that acetylation persists in the absence of a fully functioning DDR (Figures 5C and 5D). To identify deacetylation sites in Col-0 histones following IR, we used antibodies against each individual acetylated lysine residue; histones displayed consistent levels of acetylation on K5 (Figure 5F), whereas K12 and K16 show a decrease in relative abundance after irradiation and therefore represent candidate sites for deacetylation during the DDR (Figures 5G and 5H).

#### DISCUSSION

We used an LC-MS/MS approach to investigate acetyl and methyl modifications of *Arabidopsis* histones following DNA damage induction by X-ray treatment and investigated the DNA-damage specific modification of individual peptides through Western blot analysis.

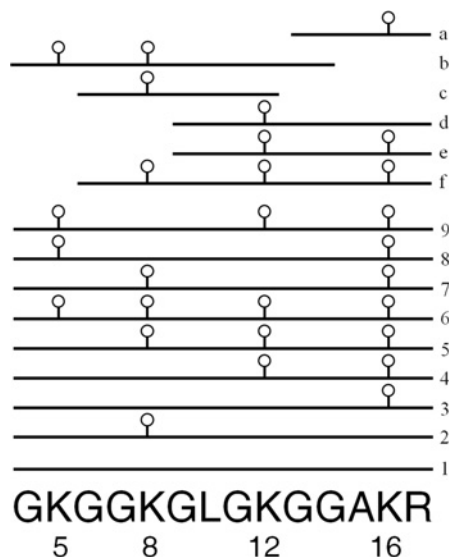
Our MS analysis provided good sequence coverage of both H3 and H4 (around 50% coverage) through the inclusion of Arg-C digestion and propionylation steps. The four H3 acetylated isoforms peptides described in the present study were also identified in a nucleosome-wide characterization of *Arabidopsis* histones, which additionally identified the monoacetylated peptide K<sup>18</sup>acQLATK<sup>23</sup> [14]. In addition, a study

of the relative abundance of peptide isoforms of histones isolated from *Arabidopsis* inflorescence tissue showed that the unmodified H3 peptide K<sup>18</sup>QLATKAAR<sup>26</sup> was the most abundant isoform, followed by the K18ac, K18acK23ac and then K23ac isoforms [13]. The MS analysis the present study was not quantitative, however we identified the two least abundant forms, K23ac and K18acK23ac, demonstrating the sensitivity of our analysis and confirming the presence of these isoforms following genotoxic stress. Studies of *Arabidopsis* using quantitative MS analysis in addition to an immunological detection report that K9ac accounts for a few per cent of the total histone population [13], and in human cells, H3K9ac levels are reduced in response to DNA damage [18]. This is consistent with the failure to detect this species in X-ray treated plant tissues, whereas the mutually exclusive methylated peptide was identified in the present study. The present study also detected the dimethylated and acetylated H3 peptide K9me<sub>2</sub>K14ac, previously shown to be low in abundance in both *Arabidopsis* [13] and alfalfa [15].

To investigate whether any hyperacetylated forms are associated with the plant response to genotoxic stress, immunoblotting was performed using antiserum specific against acetylated histone peptides. Acetylation of H3K14acK18ac increased in response to X-ray induction relative to untreated controls, indicating potential roles for chromatin modification in the plant DDR. Similar increases in H3K14ac were observed in yeast cells following UV treatment [25]. Although acetylation confers structural changes to chromatin, the relationship of these covalent modifications in the DDR is complex, as the combination of covalent histone marks also initiates ATP-dependent remodelling activities [24], with acetylation of the H3 tail dramatically increasing nucleosome repositioning by the RSC complex [26]. Although the present study demonstrates hyperacetylation of *Arabidopsis* H3 in response to DNA damage, this acetylation occurs in the absence of ATM indicating that this modification is independent of (or upstream of) the well-characterized ATM-kinase mediated DDR in plants.

Extensive acetylation of *Arabidopsis* histone H4 was detected in the present study, including nine forms of H4 G<sup>4</sup>KGGKGLGKGGAKR<sup>17</sup> and unmodified, mono-, di-, tri- and tetra-acetylated peptide isoforms represented (Figure 6, residues 1–9). Zhang et al. [14] identified acetylation sites by LC-MS/MS on six smaller tryptic peptides that are contained within the sequence GKGKGLGKGGAKR (Figure 6, peptides a–f). Their analysis identified modified peptides in common with the present





**Figure 6** Summary of peptides from histone H4 observed by LC-MS/MS

The nine GKGKGLGKGGAKR peptides observed in the present study, one unmodified and eight acetylated, are labelled 1–9. Smaller peptides included within the sequence GKGKGLGKGGAKR that were observed by Zhang et al. [14] (labelled a–f) are also shown for comparison. Sites of acetylation at lysine residues 5, 8, 12 and 16 are indicated by open circles.

study, which additionally demonstrates the acetylation of K5 without concomitant acetylation of K8 (Figure 6, peptides 8 and 9). This result contrasts with previous models of acetylation in plants [21], in which acetylation of K5, K8, K12 and K16 in *Brassica oleracea* was proposed to occur sequentially beginning with K16 and proceeding towards the N-terminal K5. In addition, peptides 2 and 7 (Figure 6) are also inconsistent with the proposed model, but suggest that acetylation of K16 is common.

The tetra-acetylated H4 tail has been shown to increase in abundance following UV exposure in UV-resistant maize [12], and is present specifically at transcription start sites of UV-tolerance-associated genes. In mammals, hyperacetylation of H4 occurs at endonuclease-induced DSBs; H4K8ac is enriched at break sites downstream of  $\gamma$ -H2AX, and two subunits of acetyltransferase NuA4 complex were also present [27]. Substitution of lysine residues of H4 at K5, K8, K12 and K16 by glutamine rendered cells more sensitive to DNA damage and impaired NHEJ (non-homologous end-joining) repair of DSBs demonstrating the importance of these residues in the mammalian DSB response [19]. Furthermore, tetra-acetylation is required for accumulation of BRCA1 (breast cancer 1) and Rad51 at DSB foci, and can extend for several kb either side of the DSB [27,28]. In contrast, the present study demonstrated that plants display a slight global decrease in relative abundance of tetra-acetylated H4 following 160 Gy IR. Interestingly this is not observed in *atm* mutants, which display a pronounced increase in the levels of tetra-acetylated H4 after exposure to X-rays. Mutant *atm* plants are deficient in the detection and signalling response to DSBs [6,29], and observed hyperacetylation of H4 in these lines after irradiation may be indicative of the accumulation of DSBs in the *atm* mutant background. Furthermore, the increased acetylation of H4 in *atm* mutants clearly indicates that hyperacetylation is upstream of ATM signalling. The observed deacetylation in irradiated plants may require functional ATM, as mammalian ATM interacts with the deacetylase HDAC (histone deacetylase) 1 *in vivo* with increased association following IR [30], and HDAC1 and HDAC2 have been shown to regulate acetylation and promote DSB repair [31].

## Summary

Using LC-MS/MS we have identified several different acetylated H3 and H4 N-terminal acetylation states following X-ray treatment, including four not previously described in *Arabidopsis*. Western blot analysis demonstrated hyperacetylation of histone H3 and a global decrease in H4 acetylation in response to X-rays. Acetylation levels were greater in an *atm*-null background, clearly showing that histone acetylation occurs in the absence of ATM, although deacetylation activity may be compromised. Future work investigating the hyper- and hypo-acetylation response of H3 and H4 will determine the significance of these epigenetic changes during genotoxic stress and repair.

## AUTHOR CONTRIBUTION

Georgina Drury, Wanda Waterworth and Chris West conceived the study; Georgina Drury, Wanda Waterworth and Chris West designed and performed the experiments, and evaluated the data; MS experiments and related data analysis were performed by Adam Dowle, David Ashford and Jerry Thomas. The paper was written by Georgina Drury, Wanda Waterworth and Chris West, with editorial assistance from Adam Dowle, David Ashford and Jerry Thomas.

## ACKNOWLEDGEMENTS

We thank Dr Iain Manfield (The Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, U.K.) for help and advice and Rajni Bhardwaj (University of Leeds, Leeds, U.K.) for assistance with medium and plant growth. We thank Dr David Martin (University of Dundee, Dundee, U.K.) who incorporated propionylation into Prophossi.

## FUNDING

This work was supported by the Biotechnology and Biological Science Research Council (BBSRC) [grant number BB/G001723/1]. MS work was performed using instruments of the University of York, Centre of Excellence in Mass Spectrometry.

## REFERENCES

- Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260
- Morales, V. and Richard-Foy, H. (2000) Role of histone N-terminal tails and their acetylation in nucleosome dynamics. *Mol. Cell. Biol.* **20**, 7230–7237
- Jenuwein, T. and Allis, C. D. (2001) Translating the histone code. *Science* **293**, 1074–1080
- Angel, A., Song, J., Dean, C. and Howard, M. (2011) A Polycomb-based switch underlying quantitative epigenetic memory. *Nature* **476**, 105–108
- Vempati, R. K., Jayani, R. S., Notani, D., Sengupta, A., Galande, S. and Halder, D. (2010) p300-Mediated acetylation of histone H3 lysine 56 functions in DNA damage response in mammals. *J. Biol. Chem.* **285**, 28553–28564
- Friesner, J. D., Liu, B., Culligan, K. and Britt, A. B. (2005) Ionizing radiation-dependent  $\gamma$ -H2AX focus formation requires ataxia telangiectasia mutated and ataxia telangiectasia mutated and Rad3-related. *Mol. Biol. Cell* **16**, 2566–2576
- Amiard, S., Charbonnel, C., Allain, E., Depeiges, A., White, C. I. and Gallego, M. E. (2010) Distinct roles of the ATR kinase and the Mre11–Rad50–Nbs1 complex in the maintenance of chromosomal stability in *Arabidopsis*. *Plant Cell* **22**, 3020–3033
- Falck, J., Coates, J. and Jackson, S. P. (2005) Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* **434**, 605–611
- Kinner, A., Wu, W., Staudt, C. and Iliakis, G. (2008)  $\gamma$ -H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Res.* **36**, 5678–5694
- Kurdistani, S. K. and Grunstein, M. (2003) Histone acetylation and deacetylation in yeast. *Nat. Rev. Mol. Cell. Biol.* **4**, 276–284
- Kim, Y. C., Gerlitz, G., Furusawa, T., Catez, F., Nussenzweig, A., Oh, K. S., Kraemer, K. H., Shiloh, Y. and Bustin, M. (2009) Activation of ATM depends on chromatin interactions occurring before induction of DNA damage. *Nat. Cell. Biol.* **11**, 92–96
- Casati, P., Campi, M., Chu, F., Suzuki, N., Maltby, D., Guan, S., Burlingame, A. L. and Walbot, V. (2008) Histone acetylation and chromatin remodeling are required for UV-B-dependent transcriptional activation of regulated genes in maize. *Plant Cell* **20**, 827–842

- 13 Johnson, L., Mollah, S., Garcia, B. A., Muratore, T. L., Shabanowitz, J., Hunt, D. F. and Jacobsen, S. E. (2004) Mass spectrometry analysis of *Arabidopsis* histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res.* **32**, 6511–6518
- 14 Zhang, K., Sridhar, V. V., Zhu, J., Kapoor, A. and Zhu, J. K. (2007) Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*. *PLoS ONE* **2**, e1210
- 15 Waterborg, J. H. (1990) Sequence analysis of acetylation and methylation in two histone H3 variants of alfalfa. *J. Biol. Chem.* **265**, 17157–17161
- 16 Wu, T., Yuan, T., Tsai, S. N., Wang, C., Sun, S. M., Lam, H. M. and Ngai, S. M. (2009) Mass spectrometry analysis of the variants of histone H3 and H4 of soybean and their post-translational modifications. *BMC Plant Biol.* **9**, 98
- 17 Bornemann, S., Rietschel, B., Baltruschat, S., Karas, M. and Meyer, B. (2010) A novel polyacrylamide gel system for proteomic use offering controllable pore expansion by crosslinker cleavage. *Electrophoresis* **31**, 585–592
- 18 Tjeertes, J. V., Miller, K. M. and Jackson, S. P. (2009) Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. *EMBO J.* **28**, 1878–1889
- 19 Bird, A. W., Yu, D. Y., Pray-Grant, M. G., Qiu, Q., Harmon, K. E., Megee, P. C., Grant, P. A., Smith, M. M. and Christman, M. F. (2002) Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. *Nature* **419**, 411–415
- 20 Neumann, H., Hancock, S. M., Buning, R., Routh, A., Chapman, L., Somers, J., Owen-Hughes, T., van Noort, J., Rhodes, D. and Chin, J. W. (2009) A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol. Cell.* **36**, 153–163
- 21 Earley, K. W., Shook, M. S., Brower-Toland, B., Hicks, L. and Pikaard, C. S. (2007) *In vitro* specificities of *Arabidopsis* co-activator histone acetyltransferases: implications for histone hyperacetylation in gene activation. *Plant J.* **52**, 615–626
- 22 Liu, C., Lu, F., Cui, X. and Cao, X. (2010) Histone methylation in higher plants. *Annu. Rev. Plant Biol.* **61**, 395–420
- 23 Lee, H. S., Park, J. H., Kim, S. J., Kwon, S. J. and Kwon, J. (2010) A cooperative activation loop among SWI/SNF,  $\gamma$ -H2AX and H3 acetylation for DNA double-strand break repair. *EMBO J.* **29**, 1434–1445
- 24 Xu, Y. and Price, B. D. (2011) Chromatin dynamics and the repair of DNA double strand breaks. *Cell Cycle* **10**, 261–267
- 25 Yu, S., Teng, Y., Waters, R. and Reed, S. H. (2011) How chromatin is remodelled during DNA repair of UV-induced DNA damage in *Saccharomyces cerevisiae*. *PLoS Genet.* **7**, e1002124
- 26 Chatterjee, N., Sinha, D., Lemma-Dechassa, M., Tan, S., Shogren-Knaak, M. A. and Bartholomew, B. (2011) Histone H3 tail acetylation modulates ATP-dependent remodeling through multiple mechanisms. *Nucleic Acids Res.* **39**, 8378–8391
- 27 Downs, J. A., Allard, S., Jobin-Robitaille, O., Javaheri, A., Auger, A., Bouchard, N., Kron, S. J., Jackson, S. P. and Cote, J. (2004) Binding of chromatin-modifying activities to phosphorylated histone H2A at DNA damage sites. *Mol. Cell* **16**, 979–990
- 28 Murr, R., Loizou, J. I., Yang, Y. G., Cuenin, C., Li, H., Wang, Z. Q. and Herceg, Z. (2006) Histone acetylation by Trp-1/Tip60 modulates loading of repair proteins and repair of DNA double-strand breaks. *Nat. Cell Biol.* **8**, 91–99
- 29 Culligan, K. M., Robertson, C. E., Foreman, J., Doerner, P. and Britt, A. B. (2006) ATR and ATM play both distinct and additive roles in response to ionizing radiation. *Plant J.* **48**, 947–961
- 30 Kim, G. D., Choi, Y. H., Dimtchev, A., Jeong, S. J., Dritschilo, A. and Jung, M. (1999) Sensing of ionizing radiation-induced DNA damage by ATM through interaction with histone deacetylase. *J. Biol. Chem.* **274**, 31127–31130
- 31 Miller, K. M., Tjeertes, J. V., Coates, J., Legube, G., Polo, S. E., Britton, S. and Jackson, S. P. (2010) Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining. *Nat. Struct. Mol. Biol.* **17**, 1144–1151

Received 9 December 2011/26 March 2012; accepted 11 May 2012

Published as BJ Immediate Publication 11 May 2012, doi:10.1042/BJ20111956

## Appendix 1B

Declaration from Georgina Drury attesting to my contribution to paper 1.



## Declaration of reviewee's contribution

### **Paper**

Drury, G.E., **DOWLE, A.A.**, Ashford, D.A., Waterworth, W.M., Thomas, J. & West, C.E. (2012) Dynamics of plant histone modifications in response to DNA damage. *Biochem. J.* **445(3)**, 393-401.

### **Candidate's contribution**

The submitting candidate performed proteomic analysis on isolated histones prepared by Georgina Drury. The reviewee was responsible for design and implementation of the proteomic experiments and analysis of all proteomic data, involving: histone propionylation, Arg-C digestion, LC-MS/MS analysis, database searching and manual validation of spectra generated to determine positions of acetylation and methylation. The work was provided as fee for service through the Bioscience Technology Facility, Department of Biology, University of York. Contributions to data analysis, production of figures and input into writing the paper were provided on a collaborative basis.

### **Declaration**

I attest that the summation above is a true reflection of the candidate's (Adam A. Dowle) contribution to the paper.

### **Candidate**

**Name:** Adam Dowle

**Signature:** 

**Date:** 6<sup>th</sup> June 2016

### **Co-author**

**Name:** Georgina Drury

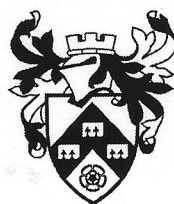
**Signature:** 

**Date:** 06-06-2016



## Appendix 1C

Declaration from Jerry Thomas attesting to my contribution to paper 1.



## Declaration of reviewee's contribution

### **Paper**

Drury, G.E., **DOWLE, A.A.**, Ashford, D.A., Waterworth, W.M., Thomas, J. & West, C.E. (2012) Dynamics of plant histone modifications in response to DNA damage. *Biochem. J.* **445(3)**, 393-401.

### **Candidate's contribution**

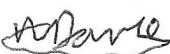
The submitting candidate performed proteomic analysis on isolated histones prepared by Georgina Drury. The reviewee was responsible for design and implementation of the proteomic experiments and analysis of all proteomic data, involving: histone propionylation, Arg-C digestion, LC-MS/MS analysis, database searching and manual validation of spectra generated to determine positions of acetylation and methylation. The work was provided as fee for service through the Bioscience Technology Facility, Department of Biology, University of York. Contributions to data analysis, production of figures and input into writing the paper were provided on a collaborative basis.

### **Declaration**

I attest that the summation above is a true reflection of the candidate's (Adam A. Dowle) contribution to the paper.

#### **Candidate**

**Name:** Adam Dowle

**Signature:** 

**Date:** 17<sup>th</sup> May 2016

#### **Co-author**

**Name:** Jerry Thomas

**Signature:** 

**Date:** 17-5-16

## Appendix 2A

Submitted paper 2 – The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory-secretory products.



## The secretome of the filarial parasite, *Brugia malayi*: Proteomic profile of adult excretory–secretory products

James P. Hewitson<sup>a</sup>, Yvonne M. Harcus<sup>a</sup>, Rachel S. Curwen<sup>b</sup>, Adam A. Dowle<sup>b</sup>,  
Agnes K. Atmadja<sup>c</sup>, Peter D. Ashton<sup>b</sup>, Alan Wilson<sup>b</sup>, Rick M. Maizels<sup>a,\*</sup>

<sup>a</sup> Institute of Immunology and Infection Research, University of Edinburgh, UK

<sup>b</sup> Department of Biology, University of York, UK

<sup>c</sup> Department of Parasitology, University of Indonesia, Jakarta, Indonesia

### ARTICLE INFO

#### Article history:

Received 9 January 2008

Received in revised form 8 February 2008

Accepted 11 February 2008

Available online 16 February 2008

#### Keywords:

Filaria  
Nematode  
In vitro culture  
Mass spectrometry  
Antigen identification

### ABSTRACT

The secretome of a parasite in its definitive host can be considered to be its genome *in trans*, to the extent that secreted products encoded by the parasite fulfill their function in the host milieu. The ‘extended phenotype’ of the filarial parasite, *Brugia malayi*, is of particular interest because of the evidence that infection results in potent down-modulation of the host immune response. We collected *B. malayi* ‘excretory–secretory’ (BES) proteins from adult parasites and using a combination of shotgun LC–MS/MS and 2D gel electrophoresis, identified 80 *B. malayi* and two host proteins in BES, of which 31 (38%) were detectable in whole worm extract (BmA). Products which were enriched in BES relative to BmA included phosphatidylethanolamine-binding protein (PEB), leucyl aminopeptidase (LAP, homologue of ES-62 from the related filaria *Acanthocheilonema viteae*), *N*-acetylglucosaminyltransferase (GlcNAcT) and galectin-1, in addition to the previously described major surface glycoprotein, glutathione peroxidase (gp29, GPX-1) and the cytokine homologue macrophage migration inhibitory factor (MIF-1). One of the most abundant released proteins was triose phosphate isomerase (TPI), yet many other glycolytic enzymes (such as aldolase and GAPDH) were found only in the somatic extract. Among the more prominent novel products identified in BES were a set of 11 small transthyretin-like proteins, and three glutamine-rich-repeat mucin-like proteins. Notably, no evidence was found of any secreted protein corresponding to the genome of the *Wolbachia* endosymbiont present in *B. malayi*. Western blotting with anti-phosphorylcholine (PC) monoclonal antibody identified that GlcNAcT, and not the ES-62 homologue, is the major PC-bearing protein in BES, while probing with human filariasis sera showed preferential reactivity to galectin-1 and to processed forms of myotactin. Overall, this analysis demonstrates selective release of a suite of newly identified proteins not previously suspected to be involved at the host–parasite interface, and provides important new perspectives on the biology of the filarial parasite.

© 2008 Elsevier B.V. All rights reserved.

### 1. Introduction

Helminth parasites continue to be a major global health problem with over 2 billion people infected across the world [1]. Their persistence is, in part, due to their success in escaping host immunity and surviving for years within the human body. Their long lifespan is considered to reflect the evolution of highly successful immune evasion strategies [2]. Manipulation of the host immune system is most likely to be mediated by molecules released from live parasites, and hence the definition of parasite secretomes will identify

candidate molecules involved in immune modulation [3]. Further, if the function of parasite secretions can be blocked by host antibodies, than secreted antigens may be critical in the formulation of future vaccines against parasitic helminth infections [4,5]. For these reasons, we have analysed the spectrum of proteins secreted by adult parasites of the human filarial nematode, *Brugia malayi*, aided by the recent release of the draft genome sequence of this pathogen [6].

In both human populations and experimental model systems, filarial infections with *B. malayi* and the closely related *Wuchereria bancrofti* are associated with immunological down-regulation. In infection, there is a dominance of regulatory cytokines, such as IL-10 and TGF- $\beta$ , which inhibit parasite-specific T cell proliferation and effector function [7,8], leading to a diverted Th2 phenotype characterised by high IL-4 and IgG4 isotype antibodies, but low levels of eosinophil-activating IL-5 and of reagenic IgE antibody [9].

\* Corresponding author at: Institute of Immunology and Infection Research, University of Edinburgh, Ashworth Laboratories, West Mains Road, Edinburgh EH9 3JT, UK. Fax: +44 131 650 5450.

E-mail address: r.maizels@ed.ac.uk (R.M. Maizels).

T cells from infected individuals are associated with the presence of a number of markers linked to regulatory T cells [10,11]. Defects in the antigen presenting cell (APC) population are also evident [12], as exemplified by reduced monocyte responsiveness to inflammatory stimuli [13]. Studies with mouse models confirm that, as in humans, both innate and adaptive arms of host immunity are targeted by filarial parasites [14–16].

The molecular basis of helminth-induced hypo-responsiveness has yet to be defined. One approach to investigate this is to characterise the molecules at the host–pathogen interface, given that these will have the opportunity to interact with, and potentially subvert, host immune cells. Early work used radiolabelling techniques to characterise the major surface and secretory products of adult *B. malayi* [17,18]. Subsequent *in vivo* studies have shown that *B. malayi* excretory–secretory (ES) products (hereafter referred to as BES) can inhibit parasite-specific proliferative responses [19], perhaps via the induction of a suppressive macrophage population [20].

More recently, as a result of large-scale parasite sequencing (genomic and EST) projects (<http://www.sanger.ac.uk/Projects/Helminths/>), the proteomic identification of helminth ES products has become possible. Among nematodes, numerous ES products from *Haemonchus contortus* [3], *Trichinella spiralis* [21], and *Teladorsagia circumcincta* [22] have been identified by reference to EST datasets. *Schistosoma mansoni* secretions from larvae [23,24], eggs [25] and adult guts [26], have also been determined by reference to transcriptomic data. However, proteomic studies on the secretions of filarial parasites revealed few matches to ESTs (Harcus, Curwen, Wilson and Maizels, unpublished). The release of the draft genome of *B. malayi* [6] has offered the opportunity for a more systematic characterisation of BES, and we now report on the use of two parallel proteomic approaches, gel-free (shotgun) LC–MS/MS and 2D gel electrophoresis to probe the adult *B. malayi* secretome.

## 2. Materials and methods

### 2.1. Parasites

*B. malayi*, obtained originally from TRS Laboratories, was maintained through *Aedes aegypti* mosquitoes and *Meriones unguiculatus* gerbils. Adult worms were recovered from the peritoneum approximately 4 months post-infection.

### 2.2. *In vitro* culture and collection of BES and BmA

Culture medium comprised serum-free RPMI-1640 supplemented with 25 mM HEPES, 100 U/ml penicillin, 100 µg/ml streptomycin (all Gibco) and 1% glucose (Sigma). Male and female parasites were washed extensively with culture medium, and then cultured at approximately 50 worms/50 ml medium at 37 °C in 5% CO<sub>2</sub>. Spent media were collected five times over an 8-day culture period. Media were filtered through a 0.22-µm syringe filter (Sartorius) to remove microfilariae, and then stored at –20 °C. Media from 2–3 batches of worms were concentrated and then diafiltered into 20 mM Tris–HCl pH 7.3 using an amicon ultrafiltration cell with regenerated cellulose 3000 MW cut-off (MWCO) filter, followed by further concentration using a centrifugal filter with a 5000 MWCO (all Millipore). Somatic extract of adult worms (BmA) was prepared by homogenisation of male and female worms on ice in 20 mM Tris pH 7.3, followed by centrifugation at 10,000 × g for 20 min at 4 °C. Protein concentrations were determined using Comassie Plus (Pierce) and samples were stored at –80 °C until use.

### 2.3. 1D gel electrophoresis

Independently prepared batches of BES or BmA (2 µg) were heated at 95 °C for 10 min in NuPAGE LDS sample buffer (Invitrogen) with 0.5 M 2-mercaptoethanol (Sigma), and then separated on NuPAGE 4–12% Bis–Tris gels using an X cell Surelock Minicell system with NuPAGE MES SDS running buffer (all Invitrogen) for 40 min at constant 200 V. Gels were silver stained using PlusOne (Amersham Biosciences, modified according to Ref. [27]) then scanned using a Linoscan (Heidelberg).

### 2.4. 2D gel electrophoresis

Parasite protein (10–20 µg) was diluted into a final volume of 125 µl with rehydration solution (7M urea, 2M thiourea (both BDH), 4% CHAPS (Sigma), 65 mM DTE (Fluka), 0.8% IPG buffer 3–10 (GE healthcare), trace bromophenol blue (Sigma)), and used to rehydrate 7 cm pH 3–10 IPG strips (Immobiline; GE healthcare) for 14 h at 20 °C. Isoelectric focusing was carried out (500 V for 30 min; 1000 V for 30 min; gradient to 8000 V for 5 h; total ~20 kV h) using an IPGphor (Pharmacia Biotech). Strips were reduced and alkylated as previously described [28] before electrophoresis in the second dimension using NuPAGE 4–12% Bis–Tris ZOOM gels and NuPAGE MES SDS running buffer (both Invitrogen) for 2 h 10 min at constant 100 V. Gels were silver stained and scanned as described above.

### 2.5. MALDI-ToF/ToF and LC–MS/MS

Protein spots were excised, destained in Farmer's reagent (20% sodium thiosulphate, 1% potassium ferricyanide), washed twice in 50% acetonitrile with 25 mM ammonium bicarbonate, and once with acetonitrile before vacuum drying. Samples were digested overnight at 37 °C with 10 µl of 0.01 µg/µl sequencing grade modified trypsin (Promega) in 25 mM ammonium bicarbonate, and then added to a MALDI target plate with an equal volume of α-cyanohydroxycinnamic acid (5 mg/ml; Sigma) in 50% acetonitrile containing 0.1% TFA. For LC–MS/MS, samples were diluted to 100 µl with 0.5 M triethylammonium bicarbonate. Disulphide bonds were reduced with 5 mM tris-(2-carboxyethyl)phosphine and thiols modified with 10 mM methyl methanethiosulphonate, then proteins were digested overnight at 37 °C with 1/20 (w/w) trypsin. Samples were acidified (pH 2.5) with TFA, passed through a strata C18-E solid phase extraction cartridge (55 µm, 70A, Phenomenex), and then dried and reconstituted in 10 µl 0.1% TFA. Peptides (3 µl aliquots) were injected onto a polystyrene-divinylbenzene polymeric monolithic column (200 µm i.d. × 5 cm; LC Packings, NL) linked to an Ultimate nano-HPLC system (Dionex), and separated using a linear gradient of 3–51% acetonitrile in 0.1% heptafluorobutyric acid over 20 min (3 µl/min flow rate). UV absorbance at 214 nm was monitored. Fractions were collected every 6 s onto a MALDI target plate, using a probot microfraction collector (Dionex), followed by post-column addition of 0.9 µl/min matrix and 6 mg/ml α-cyanohydroxycinnamic acid in 60% acetonitrile.

Positive-ion MALDI mass spectra were obtained using an Applied Biosystems 4700 Proteomics Analyzer (Applied Biosystems, Foster City, CA, USA) in reflectron mode. MS spectra were acquired over a mass range of *m/z* 800–4000 and monoisotopic masses were obtained from centroids of raw, unsmoothed data. CID-MS/MS was performed on the 20 strongest peaks with a signal to noise greater than 50 for both the gel spots and each LC–MS/MS fraction. For the latter, a fraction-to-fraction precursor exclusion of 200 ppm was used. For CID-MS/MS, a source 1 collision energy of 1 kV was used, with air as the collision gas. The precursor mass window was set to a relative resolution of 50, and the metastable suppressor was enabled. The default calibration was used for

MS/MS spectra, which were baseline-subtracted (peak width 50) and smoothed (Savitsky-Golay with three points across a peak and polynomial order 4); peak detection used a minimum S/N of 5, local noise window of 50 *m/z*, and minimum peak width of 2.9 bins. Filters of S/N 20 and 30 were used for generating peak lists from MS and MS/MS spectra, respectively.

## 2.6. Database searching and bioinformatics

Mass spectral data from protein spots were submitted to database searching using a locally running copy of the Mascot program (Matrix Science Ltd., version 2.1). Batch-acquired MS/MS data was submitted to a MS/MS ion search through the Applied Biosystems GPS Explorer software interface (version 3.6) to Mascot. Mass spectral data from LC-MS/MS were submitted to database searching using TS2Mascot (Matrix Science, version 1.0.0) and Mascot 2.1. Search parameters allowed a maximum of one missed cleavage, the modification of cysteine, the possible oxidation of methionine, peptide tolerance of 100 ppm and MS/MS tolerance of 0.1 Da. Spectra were searched against both a recent version of the NCBI non-redundant protein database and a *Brugia* coding sequence database composed of both genomic (<http://www.tigr.org/tdb/e2k1/bma1/intro.shtml>) and EST ([http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gireport.pl?gudb=b\\_malayi](http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gireport.pl?gudb=b_malayi)) sequences. The significance threshold was set at  $p < 0.05$ , and identification required that each protein contained at least one peptide with an expect

value  $< 0.05$ . By using a decoy database, the LC-MS/MS false discovery rate for single peptide matches was estimated to be 6.1% for BES and 4.7% for BmA. Single peptide matches are marked (\*) in Tables 1 and 2. For each protein match identified by MASCOT, the programme calculated the corresponding exponentially modified protein abundance index (emPAI) as the transformed ratio of the number of experimentally observed peptides to the total number of peptides that can theoretically be detected within the operating mass range and retention range of the instrument. To establish whether the proteins present in BES contained predicted signal sequences, sequences were analysed using SignalP [29] and SecretomeP [30]. ClusterW sequence alignment and phylogenetic tree construction (bootstrap, midpoint, 1000 replications) was performed using MacVector version 9.5.2.

## 2.7. Western blotting

BES was separated by 2D gel electrophoresis as described above, transferred to nitrocellulose membrane (Biorad) using an Xcell II blot module (90 min, 30 V; Invitrogen), and blocked overnight at 4 °C (Starting Block T20; Pierce). To detect phosphorylcholine (PC)-bearing proteins, membranes were probed with 1/1000 Bp-1 [31] overnight at 4 °C, washed extensively in TBS with 0.05% tween (TBST), and incubated with 1/2000 rabbit anti-mouse Ig HRP (1 h, room temperature; DakoCytomation). To assess the binding of human IgG to BES, sera from

**Table 1**  
Proteins present in BES identified by LC MS/MS

Name	Pub.locus	TIGR locus	Gene index	Score	SS?	emPAI	Spot
Cytosol stress response/chaperones							
1 14-3-3-like protein 2	Bm1.10970	13662.m00125		74	No	0.43	
2 Chaperonin-10 kDa	Bm1.56470	15397.m00021		34*	No	0.35	
3 Heat shock protein 70 kDa	Bm1.43675	14977.m04983		113*	SS	0.33	
4 Small heat shock protein P27	XXX	XXX	TC7940	33*	NCSS	0.18	
Anti-oxidants							
5 Glutathione peroxidase, major surface antigen gp29	Bm1.40465	14972.m07803		150	SS	1.31	
6 Oxidoreductase, aldo/keto reductase	Bm1.28070	14922.m00060		52	No	0.12	
7 Superoxide dismutase	XXX	XXX	TC8075	191	No	1.07	
8 Superoxide dismutase precursor (Extracellular)	XXX	XXX	AA509030	64	SS	0.30	
9 Thioredoxin peroxidase 2	XXX	XXX	TC7730	45	NCSS	0.72	
10 Translationally controlled tumor protein	Bm1.31480	14955.m00256		39*	NCSS	0.21	
Cytosol energy metabolism							
11 Enolase	Bm1.24115	14703.m00079		346	NCSS	0.83	13
12 Glycosyl hydrolases family 31 protein	Bm1.40580	14972.m07829		69	SS	0.11	
13 Inorganic pyrophosphatase	Bm1.16955	14271.m00285		51	SS	0.14	
14 MF-1 antigen (endochitinase)	Bm1.17035	14274.m00229	TC8062	115	SS	0.19	
15 6-Phosphofructokinase	Bm1.01930	12616.m00133		39	No	0.05	
16 Triose phosphate isomerase	Bm1.29130	14940.m00172		796	No	11.11	5, 6, 7, 8
Structural/cytoskeletal							
17 Actin	Bm1.16810	14258.m00140		60	No	0.13	
18 Actin 1	Bm1.34925	14965.m00431		59	No	0.28	
19 Calmodulin	Bm1.50415	14992.m10856		34*	NCSS	0.35	
20 Calsequestrin family protein	Bm1.40185	14972.m07743		41	SS	0.24	
21 Lethal protein 805, isoform d	Bm1.12945	13929.m00009		166	SS	0.62	17
22 Major sperm protein	Bm1.55755	15304.m00111		125	NCSS	0.21	
23 Major sperm protein 2	Bm1.13600	14015.m00090		281	NCSS	4.40	
24 Major sperm protein 3	Bm1.16920	14269.m00019		44	No	0.31	
25 Muscle positioning protein 4	XXX	XXX	AI783143	88*	NCSS	0.20	
26 Muscle positioning protein 4	Bm1.05930	13207.m00046		45	NCSS	0.17	
27 Myotactin form B	Bm1.53510	15059.m00091		194	No	0.08	21, 22, 23
28 Secretory protein (LS110p)	XXX	XXX	TC9625	173	SS	2.22	
29 Tropomyosin family protein	Bm1.02060	12630.m00063		69	No	0.34	
Protein digestion and folding							
30 Calreticulin precursor	Bm1.23560	14677.m00169		71	SS	0.36	
31 Cyclophilin-2 cyp-2	Bm1.55850	15309.m00029		139	No	0.53	4
32 Cyclophilin-5 cyp-5	Bm1.24035	14702.m00390		38	SS	0.26	1
33 FKBP-12	Bm1.49010	14990.m07926		55	NCSS	1.41	
34 $\gamma$ -Glutamyltranspeptidase family protein	Bm1.09950	13531.m00015	TC8816	431	SS	4.91	14, 15, 16

Table 1 (Continued)

Name	Pub.locus	TIGR locus	Gene index	Score	SS?	emPAI	Spot
35	Leucyl aminopeptidase	Bm1_56305	15373.m00009	568	SS	2.18	18
36	Protein disulphide isomerase	Bm1_39250	14972.m07552	36	SS	0.09	
37	Serine carboxypeptidase	Bm1_43130	14977.m04868	71*	SS	0.10	
38	Ubiquitin	Bm1_25395	14761.m00205	95	No	0.35	1
39	Ubiquitin-like protein SMT3	Bm1_45210	14979.m04551	59	NCSS	1.32	
Lectins and glycosyltransferases							
40	N-Acetylglucosaminyltransferase	Bm1_07275	13311.m00333	TC7929	370	NCSS	3.70
41	C-Type lectin domain containing protein	Bm1_40520	14972.m07815	45*	NCSS	0.08	
42	Galectin Bm-GAL-1	Bm1_24940	14731.m01012	556	No	6.13	9, 10
43	Galectin Bm-GAL-2	Bm1_46750	14981.m02389	101	No	0.18	
Protease inhibitors							
44	Cystatin CPI-2	XXX	XXX	TC7871	62	SS	0.22
Lipid binding							
45	Nematode polyprotein allergen (gp15/400), NPA-1	Bm1_50995	14992.m10973	274	No	0.45	
46	Phosphatidylethanolamine-binding protein 1	Bm1_41005	14973.m02599	279	NCSS	8.36	
47	Phosphatidylethanolamine-binding protein 2	Bm1_31500	14956.m00481	44	NCSS	0.71	
Host cytokine homologues							
48	Macrophage migration inhibitory factor 1	Bm1_28435	14930.m00337	176	NCSS	2.38	2
Nuclear							
49	DNA repair protein Rad4 containing protein	Bm1_03115	12787.m00392	39	NCSS	0.06	
50	Snf5 homologue	Bm1_46120	14980.m02744	71	NCSS	0.19	
51	High mobility group protein	Bm1_25620	14768.m00190	46	No	0.33	
Transthyretin-like family proteins							
52	Transthyretin-like family protein	Bm1_26590	14830.m00078	194	SS	1.58	
53	Transthyretin-like family protein	XXX	XXX	TC7985	182	SS	1.11
54	Transthyretin-like family protein	XXX	XXX	TC8258	161	SS	3.24
55	Transthyretin-like family protein	XXX	XXX	TC8116	131	SS	1.35
56	Transthyretin-like family protein	XXX	XXX	TC7986	86	SS	2.29
57	Transthyretin-like family protein	XXX	XXX	AA592049	59*	TRUN	0.32
58	Transthyretin-like family protein	Bm1_20065	14486.m00069	AI105565	50	SS	0.33
59	Transthyretin-like family protein	Bm1_15250	14164.m00122	53	SS	0.86	
60	Transthyretin-like family protein	Bm1_04380	12984.m00011	48*	SS	0.25	
61	Transthyretin-like family protein	Bm1_26585	14830.m00077	41	SS	0.87	
62	Transthyretin-like family protein	Bm1_06445	13250.m00031	TC8095	35	SS	0.30
Proteins of undetermined function							
63	Conserved cysteine-glycine protein 1	Bm1_38150	14972.m07327	124	SS	0.52	
64	Conserved hypothetical protein	Bm1_19065	14396.m00009	35	No	0.06	
65	DJ-1 family protein	Bm1_07685	13325.m00230	49	No	0.68	
66	Fasciclin domain containing protein	Bm1_17270	14284.m00379	50	No	0.08	
67	G15-6A protein-related	Bm1_49930	14990.m08112	35	SS	0.22	
68	Hypothetical protein	Bm1_57465	15533.m00023	TC7900	107	No	0.09
69	Hypothetical protein	Bm1_19875	14469.m00102	65	TRUN	0.42	
70	Hypothetical protein	Bm1_46475	14980.m02820	36	SS	0.20	
71	Hypothetical (mucin-like) protein	Bm1_11505	13734.m00156	207	No	0.24	
72	Hypothetical (mucin-like) protein	Bm1_01245	12551.m00090	188	No	0.31	
73	Hypothetical (mucin-like) protein	Bm1_09845	13513.m00023	145	No	0.31	
74	Immunogenic protein 3	Bm1_07780	13333.m00082	TC8120	133	SS	1.09
75	Immunoglobulin I-set domain containing protein	Bm1_45475	14979.m04613	73	No	0.02	
76	Immunoglobulin I-set domain containing protein	Bm1_35295	14968.m01452	63	NCSS	0.26	
77	Major allergen	XXX	XXX	TC8813	272	SS	2.13
78	PDZ-domain protein scribble	Bm1_33310	14961.m05155	35	No	0.04	
79	Recombinant antigen R1	Bm1_11105	13673.m00035	35	SS	0.23	
80	SXP-1 protein	Bm1_42870	14975.m04515	125	SS	1.10	
Name	Pub.locus	Locus	Gene index	Score	SS?	emPAI	Spot
Host							
81	Fibronectin	n/a	P11276	352	n/a	0.10	
82	Serum albumin	n/a	P02769	256	n/a	0.25	19

Pub.locus is the stable gene ID, where the draft genome sequence includes the gene locus; XXX denotes gaps in the draft genome. TIGR locus is the original annotation. Gene index indicates an EST, and is given either when the genomic sequence appears truncated compared to an EST or is not present. Score refers to the mascot score. SS? shows the presence of a classical N-terminal signal sequence ("SS" from signalP), an internal non-classical sequence ("NCSS" from secretomeP), or the lack of a signal sequence ("no" negative for both signalP and secretomeP). TRUN indicates that the sequence is truncated at the N-terminus and so the presence of signal sequences could not be determined. Exponentially modified protein abundance index (emPAI) gives an approximate quantification of the protein. Spot refers to the spot number in Fig. 2C.

\* Indicates single peptide hits.

previously characterised microfilaraemic or pathology individuals [32] were pooled (five representative individuals per group). Normal human sera were obtained from non-exposed UK residents. Membranes were probed with 1/500 sera dilutions in TBST over-

night at 4 °C, washed in TBST, and then with 1/2000 HRP conjugated rabbit anti-human IgG (1 h room temperature; DakoCytomation P0214). Following further washing in TBST, all blots were developed using ChemiGlow West, according to the manufacturers instruc-



**Table 2**  
Proteins present in BmA identified by LC–MS/MS

	Name	Pub.locus	TIGR locus	Gene index	Score	emPAI	Ratio
<b>Cytosol stress response/chaperones</b>							
1	14-3-3-like protein 2	Bm1_10970	13662.m00125		82	1.46	0.3
2	Chaperonin protein HSP60	Bm1_56580	15413.m00008		52 <sup>†</sup>	0.16	xxx
3	Heat shock protein 70	Bm1_43675	14977.m04983	TC7686	613	1.56	0.2
4	Heat shock protein 70C	Bm1_17800	14318.m00070		98	0.16	xxx
5	HSP-like 86.9 kDa C30C11.4	Bm1_23190	14656.m00228		96	0.19	xxx
6	Heat shock protein 90	Bm1_51495	14992.m11078		269	0.65	xxx
7	Small heat shock protein	Bm1_19805	14459.m00246	TC8028	52	0.47	xxx
8	Small heat shock 19.4 kDa protein ZC395.10	Bm1_30230	14950.m01862		52 <sup>†</sup>	0.27	xxx
9	Small heat shock protein OV25-1	Bm1_14535	14083.m00056		95	0.68	xxx
10	Small heat shock protein P27	XXX	XXX	TC7940	242	0.98	0.2
<b>Anti-oxidants</b>							
11	Glutathione peroxidase, major surface antigen gp29	Bm1_40465	14972.m07803		47 <sup>†</sup>	0.19	6.9
12	Oxidoreductase, aldo/keto reductase	Bm1_28070	14922.m00060		94	0.27	0.4
13	Thioredoxin	Bm1_46700	14981.m02379	TC7847	80 <sup>†</sup>	0.22	xxx
14	Thioredoxin peroxidase 2	XXX	XXX	TC7730	63	0.47	1.5
<b>Cytosol energy metabolism</b>							
15	Adenylate kinase isoenzyme 1	Bm1_13790	14037.m00196		148	0.61	xxx
16	Carbohydrate phosphorylase	Bm1_16060	14231.m00144		119	0.25	xxx
17	Enolase	Bm1_24115	14703.m00079	TC7838	541	0.65	1.3
18	Fructose-bisphosphate aldolase 1	Bm1_15350	14176.m00093		306	0.65	xxx
19	Glyceraldehyde 3-phosphate dehydrogenase	Bm1_41940	14975.m04318	TC7820	206	0.49	xxx
20	Inorganic pyrophosphatase	Bm1_16955	14271.m00285		51 <sup>†</sup>	0.15	0.9
21	Phosphoenolpyruvate carboxykinase (GTP)	Bm1_25195	14749.m00214		86	0.23	xxx
22	Phosphoglycerate kinase	Bm1_01925	12616.m00132		117	0.43	xxx
23	Triosephosphate isomerase	Bm1_29130	14940.m00172		390	0.65	17.1
<b>Structural/cytoskeletal</b>							
24	Actin 1	Bm1_34925	14965.m00431		482	0.65	0.4
25	Actin	Bm1_16810	14258.m00140		432	0.65	0.2
26	Calponin protein 3	XXX	XXX	TC8546	295	0.65	xxx
27	Major sperm protein	Bm1_55755	15304.m00111	TC7970	96 <sup>†</sup>	0.22	1.0
28	Major sperm protein 2	Bm1_13600	14015.m00090		112	0.90	4.9
29	Myosin essential light chain	XXX	XXX	TC7908	104	0.40	xxx
30	Myosin heavy chain	Bm1_40715	14972.m07860	TC7898	52	0.05	xxx
31	Myosin regulatory light chain 1	Bm1_40180	14972.m07742	TC8201	117	0.26	xxx
32	Profilin family protein	Bm1_21620	14590.m00346		57	0.43	xxx
33	Secretory protein (LS110p)	XXX	XXX	TC9625	52 <sup>†</sup>	0.27	8.2
34	Tropomyosin family protein	Bm1_02060	12630.m00063		88	0.35	1.0
35	Troponin-like EF hand family protein	Bm1_48810	14990.m07885	TC7791	42	0.27	xxx
36	Tubulin alpha chain	Bm1_38680	14972.m07435	TC7853	67	0.20	xxx
37	Tubulin alpha chain	Bm1_12120	13818.m00233		43	0.11	xxx
38	Tubulin beta chain	Bm1_25780	14773.m00919	TC7923	86	0.09	xxx
<b>Protein digestion and folding</b>							
39	Cyclophilin-2 cyp-2	Bm1_55850	15309.m00029	TC7957	108	0.53	1.0
40	Cyclophilin-5, cyp-5	Bm1_24035	14702.m00390		83 <sup>†</sup>	0.26	1.0
41	FKBP-12	Bm1_49010	14990.m07926	TC8539	50 <sup>†</sup>	0.32	4.4
42	γ-Glutamyltranspeptidase family protein	Bm1_09950	13531.m00015	TC8816	83	0.36	13.6
43	Leucyl aminopeptidase	Bm1_56305	15373.m00009		56	0.12	18.2
44	Proteasome subunit alpha type 1	Bm1_17805	14318.m00071		59	0.20	xxx
45	Protein disulphide isomerase	Bm1_39250	14972.m07552	TC8291	49	0.09	1.0
46	Transglutaminase	Bm1_28935	14937.m00487		96	0.10	xxx
47	Ubiquitin	Bm1_09050	13432.m00246		61	0.27	xxx
48	Ubiquitin-like protein SMT3	Bm1_45210	14979.m04551		49 <sup>†</sup>	0.53	2.5
<b>Lectins and glycosyltransferases</b>							
49	N-Acetylglucosaminyltransferase	Bm1_07275	13311.m00333	TC7929	42	0.33	11.2
<b>Protease inhibitors</b>							
50	Cystatin CPI-2	XXX	XXX	TC7871	77	0.23	1.0
<b>Lipid binding</b>							
51	Nematode polyprotein allergen (gp15/400), NPA-1	Bm1_50995	14992.m10973		332	0.65	0.7
52	Phosphatidylethanolamine-binding protein	Bm1_41005	14973.m02599	TC8147	107	0.41	20.4
<b>Nuclear</b>							
53	Nucleosome assembly protein 1	Bm1_27810	14919.m00204	TC8159	50	0.15	xxx
<b>Proteins of undetermined function</b>							
54	AT19640p-related	Bm1_34700	14963.m01811		43 <sup>†</sup>	0.11	0.9
55	Conserved cysteine-glycine protein 1	Bm1_38150	14972.m07327		108	0.15	3.5
56	Disorganized muscle protein 1	Bm1_40320	14972.m07771		68	0.16	xxx
57	Dosage compensation protein dpy-30	XXX	XXX	TC9176	64	0.19	xxx
58	G15-6A protein-related	Bm1_49930	14990.m08112	TC8020	65	0.50	0.4
59	Hypothetical protein	Bm1_06545	13257.m00082		39	0.04	xxx



Table 2 (Continued)

	Name	Pub_locus	TIGR locus	Gene index	Score	emPAI	Ratio
60	Immunoglobulin I-set domain containing protein	Bm1.35295	14968.m01452	TC7669	145	0.51	0.5
61	Thrombospondin type 1 domain containing protein	Bm1.13945	14041.m00077		48	0.05	xxx
62	TPR domain containing protein	Bm1.46100	14980.m02740		49	0.18	xxx
63	Tumor domain containing protein	Bm1.52240	14992.m11233		36	0.04	1.0

Pub.locus is the stable gene ID, where the draft genome sequence includes the gene locus; XXX denotes gaps in the draft genome. TIGR locus is the original annotation. Gene index indicates an EST, and is given either when the genomic sequence appears truncated compared to an EST or is not present. Ratio refers to the ratio of emPAI values for proteins detected in both BES and BmA (BES: BmA). Identities are from *Brugia malayi* coding sequence database; no host proteins were detected in BmA when peptides were searched against the NCBI database.

\* Indicates single peptide hits.

tions (Alpha Innotech) and imaged using a FluorChem SP (Alpha Innotech).

### 2.8. RT-PCR of *N*-acetylglucosaminyltransferase

Total RNA was extracted from adult mixed sex worms using TRIzol (Invitrogen), and reverse transcribed with MMLV reverse transcriptase (Stratagene) using standard protocols. GlcNAcT was amplified using a forward primer specific for the 5'-end of the gene coding sequence Bm1.07275 (ATGCGTTACTGCCTCTCATT) and a reverse primer specific for the 3'-end of EST TC7929 (TAGCACTCAAACATTGATATATT). An N-terminal extension to the protein was predicted from an open-reading frame in the designated non-coding sequence of Bm1.07275, and amplified using the above reverse primer and the forward primer ATGAAACCT-GAAATGTTCTCG. PCR conditions were as follows: 40 cycles of 95 °C 30 s, 60 °C 30 s, and 72 °C 2 min. Reaction products were separated on 1% agarose gels and visualised using ethidium bromide.

## 3. Results and discussion

### 3.1. Comparison of BES with whole worm extract (BmA)

The secretions of *in vitro* cultured adult *B. malayi* were collected under serum-free conditions using previously optimised protocols [18]. Preliminary 1D SDS-PAGE analysis (Fig. 1) of several different batches of concentrated BES revealed a general consistency, both in protein concentration and composition, although minor differences were evident. Protein yields from different batches averaged  $51 \pm 4 \mu\text{g/L}$  culture media, which is equivalent to  $\sim 30 \text{ ng}$  per worm per day. To address the question of whether adult *B. malayi* worms selectively secrete a specific subset of proteins, we compared BES to a soluble worm homogenate (BmA). 1D gel electrophoresis revealed that BES and BmA had distinct banding patterns (Fig. 1). 2D gel electrophoresis confirmed the differing protein compositions of the two parasite preparations (Fig. 2A and B), although visual inspection reveals a number of common proteins. Importantly, 2D gel electrophoresis shows that BES (comprising approximately 20 strongly staining spots and 70+ additional molecular species) is markedly less complex than BmA (in excess of 200 spots). This is consistent with the adult parasites differentially secreting a defined set of proteins, and argues against the ES simply reflecting non-specific leakage of proteins from the worms during the *in vitro* culture period.

To identify the proteins present in BES, selected spots were picked from 2D gels and subjected to MALDI-ToF/ToF analysis (Fig. 2C shows spots; Fig. 2D gives identities). Because 2D gel electrophoresis does not favour the identification of less abundant protein species or those with extremes of *pI*, we performed shotgun LC-MS/MS as a complementary approach, which allows the simultaneous identification of multiple proteins in a complex mixture. LC-MS/MS was also carried out on BmA to confirm differences in protein composition compared to BES, as suggested by 1D and 2D

gel electrophoresis (Figs. 1, 2A and B). By using LC-MS/MS, we were able to identify 80 *B. malayi* and 2 host proteins in BES (Table 1), and a further 63 *B. malayi* proteins in BmA (Table 2). Of the 82 positively identified BES proteins, 38% (31/82) were also detected in BmA. Thirty spots were subjected to individual LC-MS/MS analysis and produced adequate spectra for database searching (Fig. 2C); of these 23 were positively identified (Fig. 2D). All identified proteins spots were also present in the LC-MS/MS analysis (Table 1). One spot (19) corresponded to host serum albumin.

We also compared the relative abundance of proteins in BES and BmA by their emPAI values (exponentially modified Protein Abundance Index [33]). While this measure may not define the absolute abundance of different proteins, it is a useful means of estimating abundance of the same protein in different mixtures. Such analysis has been previously used to estimate changes in bacterial protein expression during different growth conditions [34]. Hence, we were able to infer that the concentration of a number of proteins differed substantially between BES and BmA, as graphically represented in Fig. 3. Specifically, phosphatidylethanolamine-binding protein (PEB Bm1.41005, 20-fold enrichment in BES), leucyl aminopeptidase (LAP, 18-fold),  $\gamma$ -glutamyltranspeptidase (14-fold)

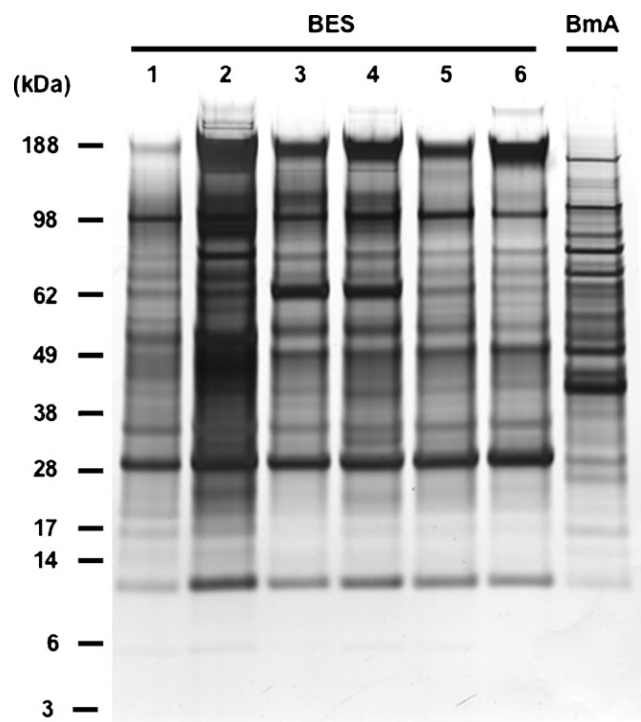
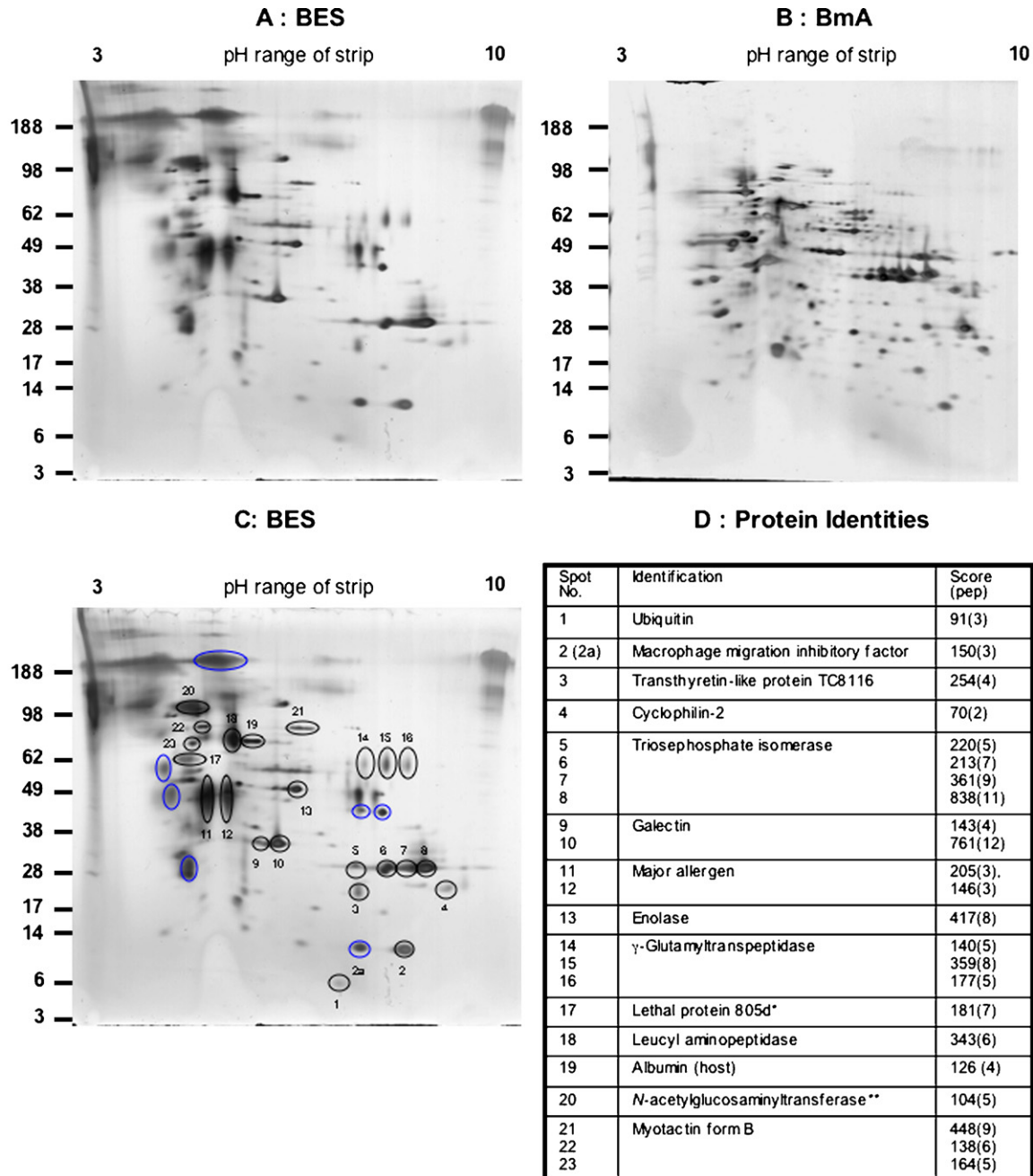


Fig. 1. One-dimensional SDS-PAGE of BES and BmA. Six independent batches of BES were compared to BmA by 1D SDS-PAGE; each lane was loaded with 2  $\mu\text{g}$  of protein. Molecular weight markers are indicated on the left.



**Fig. 2.** Two-dimensional SDS-PAGE of BES and BmA. (A) BES (20  $\mu$ g) was separated by isoelectric focusing on immobilized pH 3–10 gradients and by second dimension SDS-PAGE electrophoresis. Molecular weight markers are indicated on the left. Gel is representative of six different batches of BES. (B) BmA (20  $\mu$ g) separated as in (A). (C) 2D SDS-PAGE as shown in (A), with identified proteins circled in black and numbered in accordance with Table 1, whereas non-identified spots are circled in blue. One spot (2a) could not be unequivocally identified although it was a weak match to MIF-1 (*cf.* spot 2). (D) Table of 23 spots identified from gel C. Score refers to the mascot score and pep refers to the number of matching peptides identified. \*Unique peptides from spot matched two different genomic sequences (13507.m000111 and 13929.m00009) that overlap significantly (data not shown) and so are likely to be the same protein. Score is derived from putative full-length consensus. \*\*Score derived from putative full-length GlcNAcT as detailed in Supplementary Fig. 1.

and N-acetylglucosaminyltransferase (11-fold) appear to be preferentially secreted proteins, given that they are substantially over-represented in BES when compared to BmA. Furthermore, several proteins that are highly represented in BES, as judged by both 2D gel electrophoresis and emPAI (Fig. 2C and Table 1), such as the galectin Bm1.24940 (spots 9 and 10) and macrophage migration inhibitory factor-1 (MIF-1) (spot 2) were at undetectable levels in BmA using LC-MS/MS.

As might be expected, the majority of full-length parasite proteins identified in BES (66%; 52/78) have some form of predicted

signal sequence, either classical N-terminal (33/78) or non-classical internal (19/78). However, the remaining 33% of proteins lacked detectable signal sequences, including triose phosphate isomerase (TPI), which is the most abundant protein in BES, as judged by 2D gel spot intensity (Fig. 2C, spots 5–8), consistent with its high emPAI value (Table 1). It has been suggested that this is the result of signal peptide prediction programs being optimised using mammalian, rather than worm, proteins [25]. Alternatively, such proteins may be released into BES as a result of holocrine secretion, whereby the entire contents of a cell are released into the surroundings, as

has been previously described in *S. mansoni* [24]. Although holocrine secretion may be occurring to a certain extent, it cannot fully explain the abundance of TPI in BES (17-fold enrichment compared to BmA), given that the only other glycolytic enzyme detected in BES, enolase, is not particularly preferentially secreted (1.28 emPAI ratio BES:BmA). Additionally, abundant glycolytic enzymes, such as GAPDH and aldolase, present in BmA at similar levels to TPI, are undetectable in BES. TPI has been previously described in the secretions of schistosome larvae [23,24] and eggs [25], as well as adult *H. contortus* [3], indicating that it may play an important, as yet undiscovered, role at the helminth–host interface. These results also show it is inappropriate to use the lack of a detectable signal sequence as a reason to exclude a parasite protein from potentially fulfilling a role in the extracellular environment.

3.2. Abundant proteins in BES

The majority of BES comprises a relatively small number of abundant proteins (Fig. 2A). Prominent amongst these is the galectin Bm1\_24940 (GAL-1), identified by both LC–MS/MS and 2D gel electrophoresis (Table 1 and Fig. 2C, spots 9 and 10). A second less abundant galectin (GAL-2; Bm1\_46750) was also detected using LC–MS/MS. Galectins are an extensive family of sugar-binding proteins with an affinity for *N*-acetylglucosamines, through a conserved carbohydrate-recognition domain (CRD) [35]. Galectins have previously been identified in other helminths, with *B. malayi* GAL-1 being most similar to an *O. volvulus* galectin (AAA20541 [36]), and GAL-2 to a *H. contortus* homologue (O44126 [37]) (Fig. 4A). Both *B. malayi* galectins are tandem-repeat types, with duplicated CRDs; in each case, the C-terminal CRD contains a full sugar-binding consensus sequence [38], while N-terminal domains are divergent (Fig. 4B). No function has yet been ascribed to any helminth galectin,

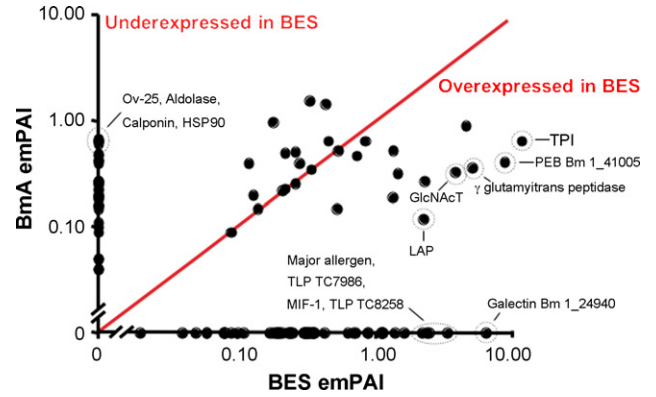


Fig. 3. Relative abundance of proteins in BES and BmA. Bivariate plot of emPAI values of proteins from BES and BmA. Proteins over-represented in either BES or BmA are highlighted.

but it is interesting to note that a number of mammalian galectins have the ability to inhibit both Th1 and Th2-mediated inflammation [39,40]. Moreover, the suppressive phenotype of regulatory T cells is dependent on their preferential expression of galectins 1 and 10 [41,42].

The binding of mammalian galectins to their target glycoproteins is determined by the availability of their polylactosamine ligands. In turn, this is regulated by the activity of Golgi-resident glycosyltransferases, such as core 2 β-1,6-*N*-acetylglucosaminyltransferase (core 2 GnT) and β-1,6-*N*-acetylglucosaminyltransferase V (GnT V), which promote the addition of polylactosamine to *O*- and *N*-linked glycans, respectively [43]. These enzymes allow galectins to bind and cross-link

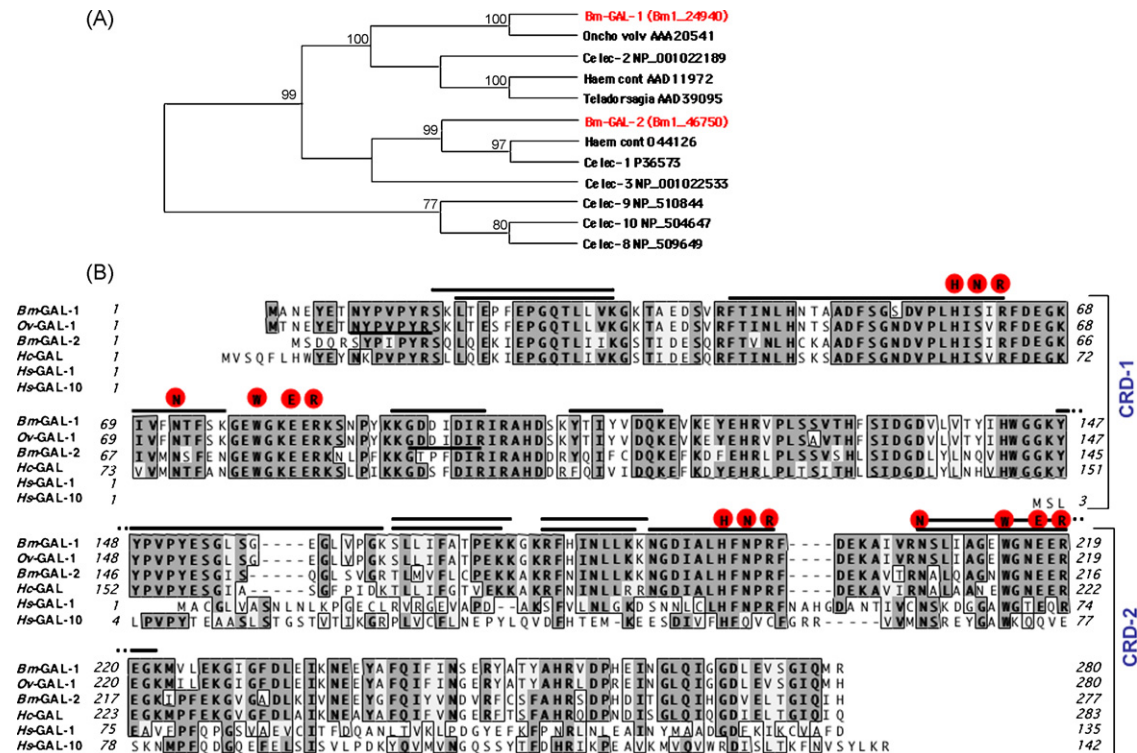
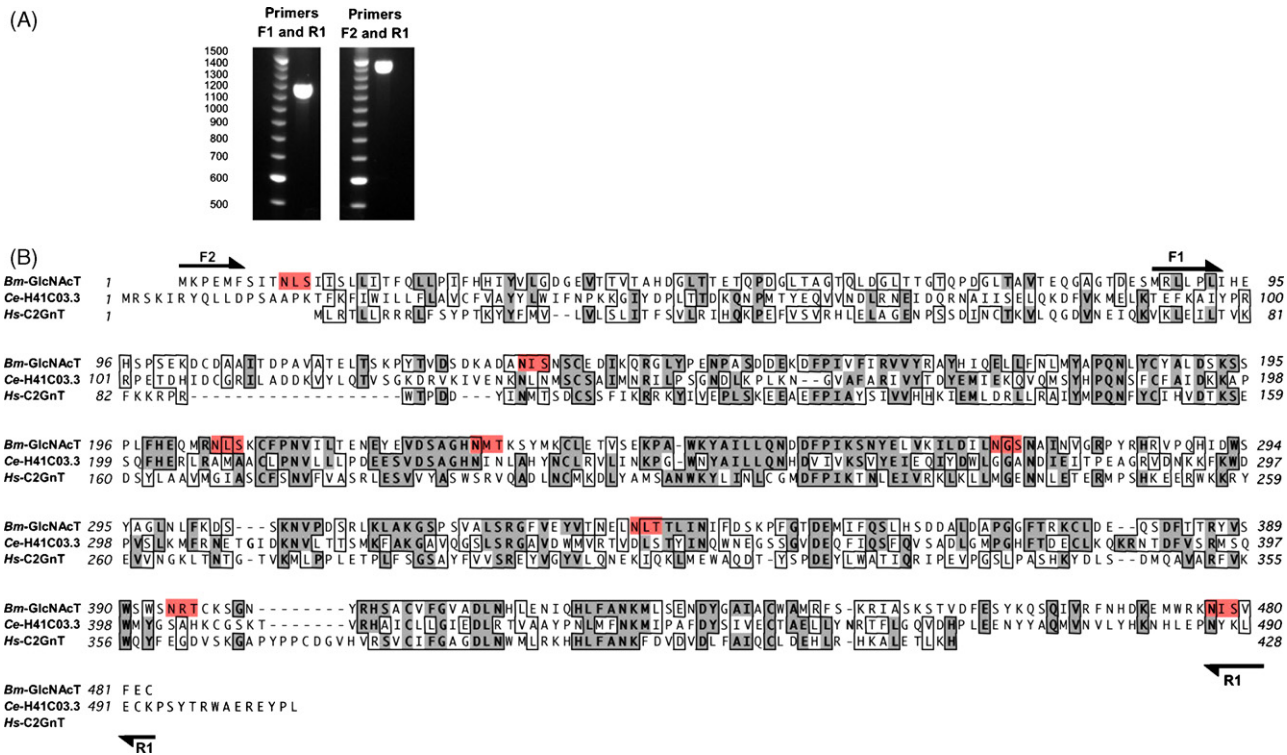


Fig. 4. Comparison of secreted *B. malayi* galectins with parasitic nematode and human galectins. (A) Phylogenetic tree of parasitic nematode galectins. Closely related galectins from *C. elegans* are included for comparison. Values indicate % occurrence of relevant nodes. (B) Sequence alignment of BES galectins with their closest identified homolog. Human (Hs) galectin 1 and 10 are included for comparison. Dark grey boxes indicate identity and light grey boxes indicate similarity. Red circles indicate conserved residues known to interact with carbohydrate ligands [38]. Thick black lines indicate tryptic peptide fragments for the *B. malayi* galectins. Each CRD (carbohydrate recognition domain) is indicated.

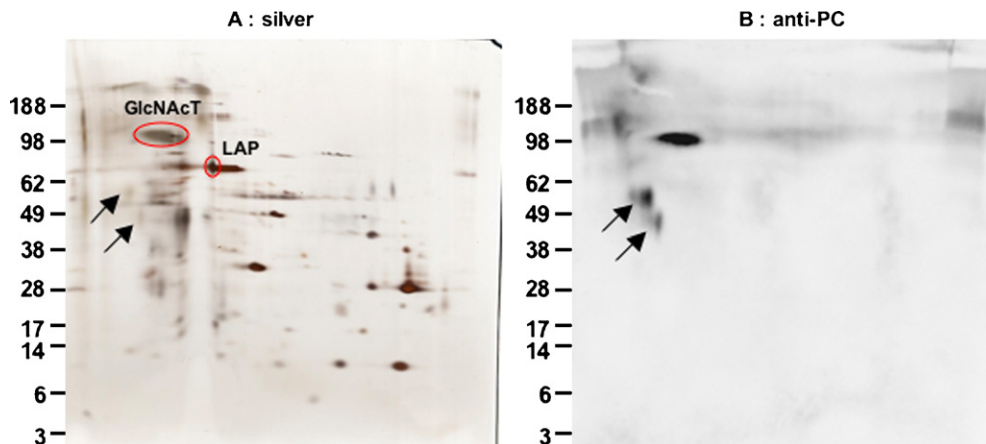




**Fig. 5.** Detection of putative full-length *N*-acetylglucosaminyltransferase. (A) Expression of the predicted *GlcNAcT* mRNA was confirmed by RT-PCR of adult *B. malayi* cDNA. Left panel: amplification using primers specific for the draft genome predicted N-terminus of Bm1.07275 (F1) and the predicted C-terminal coding sequence of EST consensus TC7929 (R1). Right panel: amplification of putative full-length *GlcNAcT* cDNA with primer F2, for upstream N-terminus based on homology between genomic sequence and *C. elegans* *GlcNAcT*, and R1 as in left panel. PCR product size is indicated by DNA ladder on the left. (B) Sequence alignment of *B. malayi* *GlcNAcT* with *C. elegans* homologue (H41C03.3) and human core 2 GnT1. Dark grey boxes indicate identity and light grey boxes indicate similarity. Red boxes indicate potential *N*-glycosylation sites of the *B. malayi* protein.

cell surface proteins, creating galectin–glycoprotein lattices, which can contribute to immune down-regulation by increasing the signalling threshold required for T cell activation [44,45]. In this regard, another abundant protein in BES was identified as *N*-acetylglucosaminyltransferase (*GlcNAcT*; Fig. 2C, spot 20), with different peptides assigned to either draft genome sequence (Bm1.07275) or to an overlapping EST consensus (TC7929) (Supplementary Fig. 1). We first verified by RT-PCR that these sequences corresponded to a single, longer gene (Fig. 5A). Inspection of upstream genome sequence indicated a further, N-terminal,

exon that was again confirmed by RT-PCR (Fig. 5A), allowing us to infer the complete sequence of *B. malayi* *GlcNAcT* (Supplementary Fig. 1). The discrepancy between the calculated (54 kDa) and observed size (~98 kDa) of this protein may be explained by substantial post-translational modifications—the protein contains eight *N*-glycosylation consensus motifs (N-X-S/T), and none of these sequences were identified as tryptic peptides by MALDI-ToF/ToF, consistent with their modification (Fig. 5B and Supplementary Fig. 1). The *B. malayi* *GlcNAcT* protein has significant homology to human core 2 GnT, as well as to a hypothetical protein from *C. ele-*



**Fig. 6.** Attachment of PC-residues to *N*-acetylglucosaminyltransferase. (A) 2D SDS-PAGE gel of BES (10 µg). Circles indicate *GlcNAcT* and LAP. (B) Duplicate of (A), Western blotted and probed with anti-PC Bp-1 monoclonal antibody. The PC-bearing spot was identified as *GlcNAcT* (Fig. 2C). Representative of two independent batches of BES. We were unable to identify the minor PC positive proteins (arrows).

gans (Fig. 5B). As yet, it is not known whether *B. malayi* GlcNAcT is capable of inducing protein glycosylation (and galectin binding) in the extracellular milieu, given that glycosyltransferases are usually functional in the Golgi. It is also possible that GlcNAcT may have an alternative extracellular function, as had been demonstrated for a secreted form of human GnT V, which can promote angiogenesis [46].

The *B. malayi* homologue of *Acanthocheilonema viteae* ES-62, LAP, was readily detectable in BES (Fig. 2C, spot 18). The anti-inflammatory properties of ES-62 are well known [47], and are dependent on its modification with phosphorylcholine (PC) side chains [48]. Furthermore, the attachment of PC to *B. malayi* proteins allows them to modulate immune cells, demonstrated by their inhibition of mitogen-dependent T cell proliferation [49]. To probe LAP for PC side chains, we used the anti-PC monoclonal antibody Bp-1 [31] on 2D Western blots of BES (Fig. 6A and B). Surprisingly, we could find no evidence for the attachment of PC to LAP. Instead, we identified the most prominent PC-bearing species in BES as the GlcNAcT described above. This is consistent with previous detection of a ~90 kDa PC-conjugated protein in BES [50]. We were unable to identify the less intense PC-bearing proteins of 45–55 kDa by MALDI-ToF/ToF. It has previously been noted that *B. pahangi* secretes 3-fold less LAP than *A. viteae* does of ES-62 [51], and our data indicate that in *B. malayi*, the majority of PC is coupled to a different carrier protein than LAP. Further functional comparisons between PC-conjugated proteins from the two species will doubtless yield extremely interesting results, and experiments designed to assess the ability of BES to modulate inflammatory diseases may also be warranted.

Also prominent in BES are two proteins,  $\gamma$ -glutamyltranspeptidase (Fig. 2C, three isoforms, spots 14–16) and a protein designated as “major allergen” (two isoforms, spots 11 and 12), that are known to be targets for host IgE, either in experimental animal models or human infection [52,53]. The function of “major allergen” is unknown, although it has weak homology to two proteins from *C. elegans* (ZC412.3 and T10G3.3).

### 3.3. Potential immune evasion proteins

*B. malayi* release a number of proteins whose function appears to directly interfere with host effector mechanisms. Prominent among these are several anti-oxidant proteins, which can detoxify the potentially damaging reactive oxygen and nitrogen intermediates produced by immune cells. Both the cytoplasmic and extracellular forms of superoxide dismutase were detected (Table 1; [54]). The major cuticular protein of adult *B. malayi*, gp29 (glutathione peroxidase; [55,56]), which may protect parasite lipids from reactive oxygen attack [57], was also released. Additionally, thioredoxin peroxidase 2 [58] was identified using LC-MS/MS. As well as the above enzymes, two scavenger proteins are present in BES which can act as anti-oxidants in a non-enzymatic manner: translationally controlled tumour protein [59], and host albumin [60]. It is unclear how host albumin comes to be associated with adult *B. malayi*, although previous studies have detected it on the surface of the microfilarial stage of other filarial parasites [61,62].

Another mechanism by which filarial parasites may dampen the immune response is through the release of protease inhibitors [63]. *B. malayi* encodes three cysteine-protease inhibitors, cystatins (CPI-1, -2 and -3 [64]), and CPI-2 was identified in BES through LC-MS/MS (Table 1). CPI-2 homologues are prominent in the secretions of a range of filarial nematodes (reviewed in Ref. [63]). Bm-CPI-2 can inhibit host proteases involved in antigen processing and presentation, including the key enzyme asparaginyl endopeptidase [65]. Two PEB proteins identified in BES by LC-MS/MS (Table 1; Bm1.41005 and Bm1.31500) may represent

additional protease inhibitors. Mouse PEB, to which the *B. malayi* proteins share significant sequence similarity (Supplementary Fig. 2), has been shown to be a novel type of serine protease inhibitor [66]. However, it is not yet known whether the *B. malayi* proteins, or homologues from other helminths (such as *O. volvulus* Ov-16 [67] and *Toxocara canis* TES-26 [68]) share this inhibitory function.

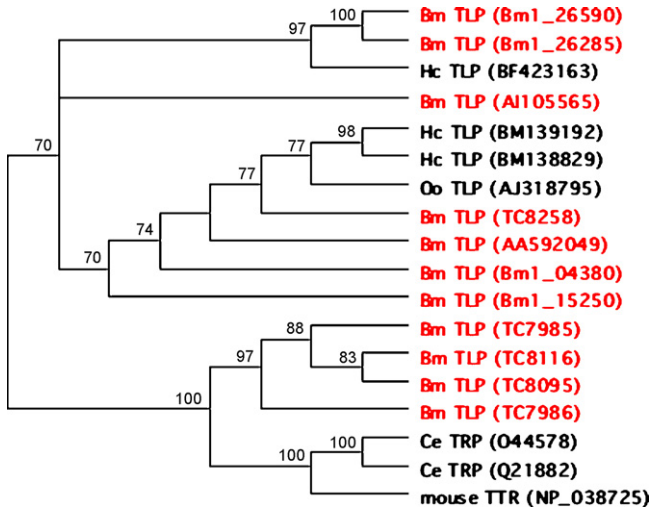
*B. malayi* secrete three members of the immunophilin family (Table 1), specifically two cyclophilins (CYP-2 (see also Fig. 2C, spot 4) and CYP-5) and FKBP (FK506-Binding Protein)-12. Immunophilins are characterised by peptidyl-propyl *cis-trans* isomerase activity, and have roles in protein folding and as molecular chaperones [69]. It is unclear what, if any, role these proteins have in parasite immune evasion, although secreted cyclophilins have been identified in the ES of *H. contortus* [3], and schistosome larvae and eggs [67]. Of note, a cyclophilin (CYP-18) from the protozoan parasite *Toxoplasma gondii* is directly involved in host-parasite crosstalk, as it can stimulate protective Th1 responses through its binding to the chemokine receptor CCR5 [70].

A wide variety of pathogens, from viruses to helminths, are able to subvert the host immune response through the production of host cytokine mimics [71]. *B. malayi* produces two homologues of macrophage migration inhibitory factor (Bm-MIF-1 and -2) [72,73]. In our analysis, MIF-1 is readily detectable in BES through both LC-MS/MS and 2D gel electrophoresis (Table 1 A and Fig. 2C, spot 2). It is known that MIF-2 is expressed at a lower level [73], and so its release may be below our lower limit of detection. Similarly, we have not detected the expression of homologues of host TGF- $\beta$  (Bm-TGH1 and TGH-2), although secretion of the latter has been reported [74]. TGF- $\beta$  family members are potent at extremely low concentrations, and again may be present below our lower detection threshold. There was also no evidence of secretion of two recently assigned cytokine homologues, similar to IL-16 and IL-17, which have recently been identified at the genomic level [6].

### 3.4. Newly identified secretory proteins and unidentified proteins

The benefit of applying a proteomic-based approach to the study of BES is it can identify previously undiscovered secretory proteins. In this respect, we have found that adult *B. malayi* secrete 11 members of the transthyretin-like protein (TLP) family (Table 1). The *B. malayi* draft genome reveals 15 TLP genes [6], of which 6 are detected in BES. The remaining five proteins (TC8116 – also identified as spot 3, Fig. 2C – as well as TC7985, TC8258, TC7986, AA592049) are tentative consensus or ESTs from the *B. malayi* gene index project, and have no match in the current draft version of the genome. TLP are distantly related to both the classical transthyretin (TTR) and transthyretin-related protein (TRP) families (Fig. 7) [75]. TTR is responsible for the transport of thyroid hormones, as well as vitamin A and its derivative retinoic acid [76]. TLP are also secreted by the cattle helminths *Ostertagia ostertagi* (AJ31875; [77]) and *H. contortus* (BM139192, BM138829, BF423163 [3]), although their function is currently unknown.

In addition to newly discovered secreted proteins with close matches from other parasites, 7 ES products matched predicted proteins in the draft genome with no known homologue. Three closely related hypothetical proteins (Bm1.11505; Bm1.01245; Bm1.09845) were identified, which are unusual in being composed of multiple homologous repeats of ~70 amino acids, including an acidic tract (QEEEEEE), as well as being rich in serines and threonines (Supplementary Fig. 3). Together these characteristics support the assignment of mucin-like proteins in Table 1. Four other hypothetical proteins were identified, of which two had weak homology to entries from either *C. elegans* (Bm1.46475 to KO1A11.1) or *C. briggsae* (Bm1.19065 to CBG09470).



**Fig. 7.** Transthyretin-like proteins. Phylogenetic tree of TLP, TRP and TTR from *B. malayi* (Bm), *H. contortus* (Hc), *O. ostertagi* (Oo), *C. elegans* (Ce) and mouse. Values indicate % occurrence of relevant nodes.

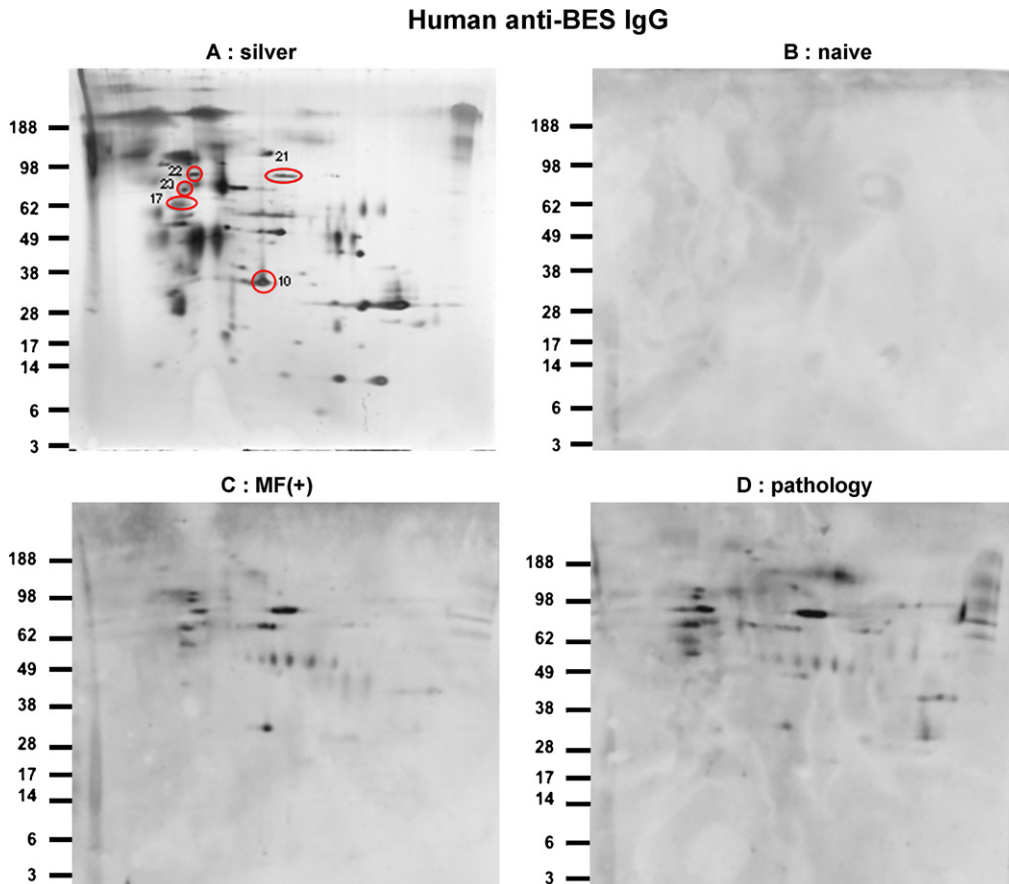
As is normal with LC-MS/MS analysis, many peptide masses could not be unequivocally assigned to an individual gene product. However, we were surprised that 7 out of 30 spots from 2D analysis also could not be matched, despite generating good spectra and exhaustive searching of genomic and EST datasets (Fig. 2C, blue

circles). Because the draft genome is estimated to be 90% complete, it may be that these proteins corresponded to genes that have yet to be identified at the genomic level. However, as sufficient sample quantities become available, *de novo* peptide sequence deduction will be carried out, from which experimental searching of genomic DNA with degenerate primers may allow us to isolate the corresponding gene sequences.

*B. malayi* contains a rickettsial endosymbiont, *Wolbachia* (wBm), which is essential for the survival and reproduction of the parasite [78]. It has also been postulated that *Wolbachia*-derived products may impact upon the host immune system. We therefore, specifically searched for matches to the 1.1 Mb genome of wBm [79], but found none. Hence, in our system at least, *Wolbachia* proteins are not strongly represented in the secretions of viable adult parasites. It is, of course, very likely that dying parasites would release *Wolbachia* proteins, and the relative influence of parasite and endosymbiont products on host immunity remains to be determined.

3.5. Human sera recognition

To define which, if any, of the identified BES proteins were targets for immune recognition in human filarial infections, Western blots were carried out using two pools of characterised sera from Indonesian filariasis patients. We have previously reported on the reactivity of individual members of these pools to extracts of whole worm [32], which will measure immune responses to the totality of somatic antigens, many of which are likely to be presented only after parasite demise. Immunoreactivity to BES, however, is likely to



**Fig. 8.** Recognition patterns of anti-BES IgG in *B. malayi* infected patients. (A) One of 4 replicate 2D gels of BES, silver stained. (B) Replicate, Western blotted and probed with sera from non-exposed “naïve” individuals. (C) Replicate, Western blotted and probed with sera from microfilaraemic patients. (D) Replicate, Western blotted and probed with sera from elephantiasis pathology patients. Immunogenic proteins are circled in (A) and are numbered as in Fig. 2C with identities in Table 1.



reflect the immune response directed against live parasites. A particularly important question is whether clinical outcome can be correlated to the recognition of a particular antigen/group of antigens.

To begin to answer this question, we probed 2D blots of BES with pooled sera from filariasis cases and control individuals. The silver-stained replicate of BES proteins in this analysis is shown in Fig. 8A, and no significant reactivity was observed with sera from uninfected UK residents (naïve; Fig. 8B). However, asymptomatic microfilaraemic (MF+; Fig. 8C) or elephantiasis (pathology; Fig. 8D) serum pools both reacted with a number of common proteins. Overall, it is clear that sera from elephantiasis patients bind a broader spectrum of proteins than sera from microfilaraemic individuals. In general, protein abundance and immunogenicity correlated poorly. For instance, the most abundant protein in BES, TPI, was not a target for human sera (Fig. 8C and D). Instead, the most immunogenic protein in both groups was identified as myotactin form B, which is present as spots 21–23 (Figs. 2 and 8).

It is notable that myotactin B is represented by three distinct spots of 70–90 kDa on 2D gels. However, the predicted *B. malayi* myotactin B gene (Bm1.53510) encodes a much larger protein of 285 kDa. Analysis of the tryptic peptides generated by MALDI-ToF/ToF shows that the spots represent different regions of the full-length myotactin B protein; spot 21 peptides corresponding to amino acids 303–677, spot 22 to 983–1539, and spot 23 to 1020–1347 (Supplementary Fig. 4). This profile is likely to reflect proteolytic processing of the full-length protein, either directly from the parasite or following its secretion, although alternative splicing cannot be excluded. Another immunogenic protein (spot 17) was identified as lethal protein 805d. Both this protein and myotactin B are related to *C. elegans* myotactin (also known as lethal protein 805), a protein that links the contractile apparatus of the worm to its hypodermis [80]. The significance of this protein's secretion by *B. malayi* is unclear, although its immunogenicity in infected patients is consistent with a role as a decoy for the host immune response. Galectin Bm1.24940 was also an antibody target in both groups of filariasis patients, although perhaps surprisingly, it was recognised to a greater extent by (the generally less immunoreactive) MF+ sera (Fig. 8C). This work paves the way for future studies to analyse at the individual level the response of different human antibody isotypes, particularly IgG4 and IgE, to specific identified BES antigens.

#### 4. Conclusions

The secretome of the long-lived mammalian stage of a successful parasite will encompass a spectrum of proteins required for every facet of the parasitic life-style, from metabolism, reproduction, and modification of the physiological environment, as well as immune evasion. For some of the adult BES proteins we have identified (such as the cystatin, CPI-2), a specific role in immune modulation has already been established [63,65]. Our analysis greatly extends the number of candidate immune evasion products, including a number of unexpected players, such as the galectins, *N*-acetylglucosaminyltransferase, the transthyretin-like proteins, and the phosphatidylethanolamine-binding proteins, each of which fully warrant further investigation.

Many of the additional secreted proteins identified here, however, have no obvious role in the host–parasite immune interaction. It is likely that some act in other arenas, for example in aiding parasite co-location and reproduction, but if so then they may serve equally well as the target of an antibody-inducing vaccine aimed at preventing transmission. Inevitably, some of the products designated as 'secretory', may in fact simply be cellular proteins which are sufficiently abundant and soluble to be released by live para-

sites over time. Nevertheless, if such release occurs *in vivo*, then the immunological consequences may be significant, not only as host innate immune receptors may have evolved to recognise these as pathogen-associated molecular patterns, but also (as in the case of filarial tropomyosin) because they may also be effective targets of vaccine-induced immunity [81,82]. It will be interesting to test TPI as a potential vaccine target in the same manner.

The analysis reported here has been limited in that quantities of BES proteins were, in some instances, less than desired; and also because both genomic and EST datasets offer incomplete coverage. Currently the draft *B. malayi* genome covers around 90% of coding regions (13 of 80, or 16%, of identities we report match ESTs but not genomic data). As the sensitivity of mass spectrometry increases, and the scope of filarial database information widens, additional BES proteins will be identified. The identification of individual ES proteins will now facilitate discovery of complementary host receptors, and host cell types, with which parasite products may interact. In terms of the wider molecular biology of parasite secretions, future work should also expand on our knowledge of post-translational modifications, in particular glycosylations and other side chains such as PC [48] and DMAE [83]. Attention should be paid to the source of ES, perhaps using identified proteins as markers (for example, where monoclonal antibodies are available), looking particularly to distinguish dedicated secretory organs such as the amphid glands, from the uterine contents of gravid females which may be voided together with newly released MF. Finally, and perhaps most importantly, the study of filarial ES proteins should encompass the immature stages, the blood-dwelling MF and the infective L3 stage, to complete our understanding of the *B. malayi* secretome.

#### Acknowledgements

JPH, YMH and RMM acknowledge Programme Grant support from the Wellcome Trust. PDA, RSC and AW acknowledge funding support from BBSRC and the Wellcome Trust.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molbiopara.2008.02.007.

#### References

- [1] Hotez PJ, Molyneux DH, Fenwick A, et al. Control of neglected tropical diseases. *N Engl J Med* 2007;357:1018–27.
- [2] Maizels RM, Balic A, Gomez-Escobar N, Nair M, Taylor M, Allen JE. Helminth parasites: masters of regulation. *Immunol Rev* 2004;201:89–116.
- [3] Yatsuda AP, Krijgsveld J, Cornelissen AWCA, Heck AJ, De Vries E. Comprehensive analysis of the secreted proteins of the parasite *Haemonchus contortus* reveals extensive sequence variation and differential immune recognition. *J Biol Chem* 2003;278:16941–51.
- [4] Maizels RM, Holland M, Falcone FH, Zang XX, Yazdanbakhsh M. Vaccination against helminth parasites: the ultimate challenge for immunologists? *Immunol Rev* 1999;171:125–48.
- [5] Bethony J, Loukas A, Smout M, et al. Antibodies against a secreted protein from hookworm larvae reduce the intensity of hookworm infection in humans and vaccinated laboratory animals. *Faseb J* 2005;19:1743–5.
- [6] Ghedin E, Wang S, Spiro D, et al. Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 2007;317:1756–60.
- [7] King CL, Mahanty S, Kumaraswami V, et al. Cytokine control of parasite-specific anergy in human lymphatic filariasis. Preferential induction of a regulatory T helper type 2 lymphocyte subset. *J Clin Invest* 1993;92:1667–73.
- [8] Mahanty S, Ravichandran M, Raman U, Jayaraman K, Kumaraswami V, Nutman TB. Regulation of parasite antigen-driven immune responses by interleukin-10 (IL-10) and IL-12 in lymphatic filariasis. *Infect Immun* 1997;65:1742–7.
- [9] Maizels RM, Sartono E, Kurniawan A, Selkirk ME, Partono F, Yazdanbakhsh M. T cell activation and the balance of antibody isotypes in human filariasis. *Parasitol Today* 1995;11:50–6.
- [10] Steel C, Nutman TB. CTLA-4 in filarial infections: implications for a role in diminished T cell reactivity. *J Immunol* 2003;170:1930–8.

- [11] Babu S, Blauvelt CP, Kumaraswami V, Nutman TB. Regulatory networks induced by live parasites impair both Th1 and Th2 pathways in patent lymphatic filariasis: implications for parasite persistence. *J Immunol* 2006;176:3248–56.
- [12] Semnani RT, Nutman TB. Toward an understanding of the interaction between filarial parasites and host antigen-presenting cells. *Immunol Rev* 2004;201:127–38.
- [13] Sasisekhar B, Aparna M, Augustin DJ, et al. Diminished monocyte function in microfilaremic patients with lymphatic filariasis and its relationship to altered lymphoproliferative responses. *Infect Immun* 2005;73:3385–93.
- [14] MacDonald AS, Maizels RM, Lawrence RA, Dransfield I, Allen JE. Requirement for in vivo production of IL-4, but not IL-10, in the induction of proliferative suppression by filarial parasites. *J Immunol* 1998;160:4124–32.
- [15] Hoerauf A, Satoguina J, Saeftel M, Specht S. Immunomodulation by filarial nematodes. *Parasite Immunol* 2005;27:417–29.
- [16] Taylor M, Le Goff L, Harris A, Malone E, Allen JE, Maizels RM. Removal of regulatory T cell activity reverses hyporesponsiveness and leads to filarial parasite clearance in vivo. *J Immunol* 2005;174:4924–33.
- [17] Maizels RM, Denham DA, Sutanto I. Secreted and circulating antigens of the filarial parasite *Brugia pahangi*: analysis of in vitro released components and detection of parasite products in vivo. *Mol Biochem Parasitol* 1985;17:277–8.
- [18] Kwan-Lim G-E, Gregory WF, Selkirk ME, Partono F, Maizels RM. Secreted antigens of filarial nematodes: survey and characterisation of in vitro excretory/secretory (E/S) products of adult *Brugia malayi* filarial parasites. *Parasite Immunol* 1989;11:629–54.
- [19] Miller S, Schreuer D, Hammerberg B. Inhibition of antigen-driven proliferative responses and enhancement of antibody production during infection with *Brugia pahangi*. *J Immunol* 1991;147:1007–13.
- [20] Allen JE, MacDonald AS. Profound suppression of cellular proliferation mediated by the secretions of nematodes. *Parasite Immunol* 1998;20:241–7.
- [21] Robinson MW, Greig R, Beattie KA, Lamont DJ, Connolly B. Comparative analysis of the excretory–secretory proteome of the muscle larva of *Trichinella pseudospiralis* and *Trichinella spiralis*. *Int J Parasitol* 2007;37:139–48.
- [22] Craig H, Wastling JM, Knox DP. A preliminary proteomic survey of the in vitro excretory/secretory products of fourth-stage larval and adult *Teladorsagia circumcincta*. *Parasitology* 2006;132:535–43.
- [23] Knudsen GM, Medzihradsky KF, Lim KC, Hansell E, McKerrow JH. Proteomic analysis of *Schistosoma mansoni* cercarial secretions. *Mol Cell Proteomics* 2005;4:1862–75.
- [24] Curwen RS, Ashton PD, Sundaralingam S, Wilson RA. Identification of novel proteases and immunomodulators in the secretions of schistosome cercariae that facilitate host entry. *Mol Cell Proteomics* 2006;5:835–44.
- [25] Cass CL, Johnson JR, Califf LL, et al. Proteomic analysis of *Schistosoma mansoni* egg secretions. *Mol Biochem Parasitol* 2007;155:84–93.
- [26] Delcroix M, Medzihradsky K, Caffrey CR, Fetter RD, McKerrow JH. Proteomic analysis of adult *S. mansoni* gut contents. *Mol Biochem Parasitol* 2007;154:95–7.
- [27] Yan JX, Wait R, Berkelman T, et al. A modified silver staining protocol for visualization of proteins compatible with matrix-assisted laser desorption/ionization and electrospray ionization–mass spectrometry. *Electrophoresis* 2000;21:3666–72.
- [28] Curwen RS, Ashton PD, Johnston DA, Wilson RA. The *Schistosoma mansoni* soluble proteome: a comparison across four life-cycle stages. *Mol Biochem Parasitol* 2004;138:57–66.
- [29] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004;340:783–95.
- [30] Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel* 2004;17:349–56.
- [31] Sutanto I, Maizels RM, Denham DA. Surface antigens of a filarial nematode: Analysis of adult *Brugia pahangi* surface components and their use in monoclonal antibody production. *Mol Biochem Parasitol* 1985;15:203–14.
- [32] Kurniawan A, Sartono E, Partono F, Yazdanbakhsh M, Maizels RM. Specificity of predominant IgG4 antibodies to adult and microfilarial stages of *Brugia malayi*. *Parasite Immunol* 1998;20:155–62.
- [33] Ishihama Y, Oda Y, Tabata T, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 2005;4:1265–72.
- [34] Graham RL, Sharma MK, Ternan NG, Weatherly DB, Tarleton RL, McMullan G. A semi-quantitative GeLC–MS analysis of temporal proteome expression in the emerging nosocomial pathogen *Ochrobactrum anthropi*. *Genome Biol* 2007;8:R110.
- [35] Cooper DN, Barondes SH. God must love galectins; he made so many of them. *Glycobiology* 1999;9:979–84.
- [36] Klion AD, Donelson JE. OvGalBP, a filarial antigen with homology to vertebrate galactoside-binding proteins. *Mol Biochem Parasitol* 1994;65:305–15.
- [37] Newlands GF, Skuce PJ, Knox DP, Smith SK, Smith WD. Cloning and characterization of a beta-galactoside-binding protein (galectin) from the gut of the gastrointestinal nematode parasite. *Parasitology* 1999;119:483–90.
- [38] Cooper DN. Galectinomics: finding themes in complexity. *Biochim Biophys Acta* 2002;1572:209–31.
- [39] Toscano MA, Commodaro AG, Ibarregui JM, et al. Galectin-1 suppresses autoimmune retinal disease by promoting concomitant Th2- and T regulatory-mediated anti-inflammatory responses. *J Immunol* 2006;176:6323–32.
- [40] Katoh S, Ishii N, Nobumoto A, et al. Galectin-9 inhibits CD44-hyaluronan interaction and suppresses a murine model of allergic asthma. *Am J Resp Crit Care Med* 2007;176:27–35.
- [41] Kubach J, Lutter P, Bopp T, et al. Human CD4+CD25+ regulatory T cells: proteome analysis identifies galectin-10 as a novel marker essential for their energy and suppressive function. *Blood* 2007;110:1550–8.
- [42] Garin MI, Chu CC, Golshayan D, Cernuda-Morollon E, Wait R, Lechler RI. Galectin-1: a key effector of regulation mediated by CD4+CD25+ T cells. *Blood* 2007;109:2058–65.
- [43] Daniels MA, Hogquist KA, Jameson SC. Sweet 'n' sour: the impact of differential glycosylation on T cell responses. *Nat Immunol* 2002;3:903–10.
- [44] Demetriou M, Granovsky M, Quaggin S, Dennis JW. Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. *Nature* 2001;409:733–9.
- [45] Rabinovich GA, Toscano MA, Jackson SS, Vasta GR. Functions of cell surface galectin-glycoprotein lattices. *Curr Opin Struct Biol* 2007;17:513–20.
- [46] Saito T, Miyoshi E, Sasai K, et al. A secreted type of beta 1,6-N-acetylglucosaminyltransferase V (GnT-V) induces tumor angiogenesis without mediation of glycosylation: a novel function of GnT-V distinct from the original glycosyltransferase activity. *J Biol Chem* 2002;277:17002–8.
- [47] Melendez AJ, Harnett MM, Pushparaj PN, et al. Inhibition of FcεRI-mediated mast cell responses by ES-62, a product of parasitic filarial nematodes. *Nat Med* 2007;13:1375–81.
- [48] Harnett W, McInnes IB, Harnett MM. ES-62, a filarial nematode-derived immunomodulator with anti-inflammatory potential. *Immunol Lett* 2004;94:27–33.
- [49] Lal RB, Kumaraswami V, Steel C, Nutman TB. Phosphocholine-containing antigens of *Brugia malayi* nonspecifically suppress lymphocyte function. *Am J Trop Med Hyg* 1990;42:56–64.
- [50] Maizels RM, Burke J, Denham DA. Phosphorylcholine-bearing antigens in filarial nematode parasites: analysis of somatic extracts and in vitro secretions of *Brugia malayi* and *B. pahangi* and infection sera. *Parasite Immunol* 1987;9:49–66.
- [51] Stepek G, Houston KM, Goodridge HS, Devaney E, Harnett W. Stage-specific and species-specific differences in the production of the mRNA and protein for the filarial nematode secreted product, ES-62. *Parasitology* 2004;128:91–8.
- [52] Lobos E, Zahn R, Weiss N, Nutman TB. A major allergen of lymphatic filarial nematodes is a parasite homolog of the γ-glutamyl transpeptidase. *Mol Med* 1997;2:712–24.
- [53] Lobos E, Nutman TB, Hothersall JS, Moncada S. Elevated immunoglobulin E against recombinant *Brugia malayi* gamma-glutamyl transpeptidase in patients with bancroftian filariasis: association with tropical pulmonary eosinophilia or putative immunity. *Infect Immun* 2003;71:747–53.
- [54] Tang L, Ou X, Henkle-Dührsen J, Selkirk ME. Extracellular and cytoplasmic CuZn superoxide dismutases from *Brugia* lymphatic filarial nematode parasites. *Infect Immun* 1994;62:961–7.
- [55] Maizels RM, Gregory WF, Kwan-Lim G-E, Selkirk ME. Filarial surface antigens: the major 29,000 MW glycoprotein and a novel 17,000–200,000 MW complex from adult *Brugia malayi* parasites. *Mol Biochem Parasitol* 1989;32:213–27.
- [56] Cookson E, Blaxter ML, Selkirk ME. Identification of the major soluble cuticular protein of lymphatic filarial nematode parasites (gp29) as a secretory homolog of glutathione peroxidase. *Proc Natl Acad Sci USA* 1992;89:5837–41.
- [57] Tang L, Smith VP, Gounaris K, Selkirk ME. *Brugia pahangi*: the cuticular glutathione peroxidase (gp29) protects heterologous membranes from lipid peroxidation. *Exp Parasitol* 1996;82:329–32.
- [58] Ghosh I, Eisinger SW, Raghavan N, Scott AL. Thioredoxin peroxidases from *Brugia malayi*. *Mol Biochem Parasitol* 1998;91:207–20.
- [59] Gnanasekar M, Ramaswamy K. Translationally controlled tumor protein of *Brugia malayi* functions as an antioxidant protein. *Parasitol Res* 2007;101:1533–40.
- [60] Selkirk ME, Smith VP, Thomas GR, Gounaris K. Resistance of filarial nematode parasites to oxidative stress. *Int J Parasitol* 1998;28:1315–32.
- [61] Maizels RM, Philipp M, Dasgupta A, Partono F. Human serum albumin is a major component on the surface of microfilariae of *Wuchereria bancrofti*. *Parasite Immunol* 1984;6:185–90.
- [62] Philipp M, Worms MJ, McLaren DJ, Ogilvie BM, Parkhouse RME, Taylor PM. Surface proteins of a filarial nematode: a major soluble antigen and a host component on the cuticle of *Litomosoides carinii*. *Parasite Immunol* 1984;6:63–82.
- [63] Hartmann S, Lucius R. Modulation of host immune responses by nematode cystatins. *Int J Parasitol* 2003;33:1291–302.
- [64] Gregory WF, Maizels RM. Cystatins from filarial parasites: evolution, adaptation and function in the host–parasite relationship. *Int J Biochem Cell Biol*; in press.
- [65] Manoury B, Gregory WF, Maizels RM, Watts C. *Bm-CPI-2*, a cystatin homolog secreted by the filarial parasite *Brugia malayi*, inhibits class II MHC-restricted antigen processing. *Curr Biol* 2001;11:447–51.
- [66] Hengst U, Albrecht H, Hess D, Monard D. The phosphatidylethanolamine-binding protein is the prototype of a novel family of serine protease inhibitors. *J Biol Chem* 2001;276:535–40.
- [67] Lobos E, Weiss N, Karam M, Taylor HR, Ottesen EA, Nutman TB. An immunogenic *Ochocerca volvulus* antigen: a specific and early marker of infection. *Science* 1991;251:1603–5.
- [68] Gems DH, Ferguson CJ, Robertson BD, Page AP, Blaxter ML, Maizels RM. An abundant, trans-spliced mRNA from *Toxocara canis* infective larvae encodes a 26 kDa protein with homology to phosphatidylethanolamine binding proteins. *J Biol Chem* 1995;270:18517–22.



- [69] Bell A, Monaghan P, Page AP. Peptidyl-prolyl *cis*-*trans* isomerases (immunophilins) and their roles in parasite biochemistry, host-parasite interaction and antiparasitic drug action. *Int J Parasitol* 2006;36:261–76.
- [70] Golding H, Aliberti J, King LR, et al. Inhibition of HIV-1 infection by a CCR5-binding cyclophilin from *Toxoplasma gondii*. *Blood* 2003;102:3280–6.
- [71] Alcami A. Viral mimicry of cytokines, chemokines and their receptors. *Nat Rev Immunol* 2003;3:36–50.
- [72] Pastrana DV, Raghavan N, FitzGerald P, et al. Filarial nematode parasites secrete a homologue of the human cytokine macrophage migration inhibitory factor. *Infect Immun* 1998;66:5955–63.
- [73] Zang XX, Taylor P, Meyer D, et al. Homologues of human macrophage migration inhibitory factor from a parasitic nematode: gene cloning, protein activity and crystal structure. *J Biol Chem* 2002;277:44261–7.
- [74] Gomez-Escobar N, Gregory WF, Maizels RM. Identification of *Bm-tgh-2*, a filarial nematode homolog of *C.elegans daf-7* and human TGF- $\beta$ , expressed in microfilarial and adult stages of *Brugia malayi*. *Infect Immun* 2000;68:6402–10.
- [75] Eneqvist T, Lundberg E, Nilsson L, Abagyan R, Sauer-Eriksson AE. The transthyretin-related protein family. *Eur J Biochem* 2003;270:518–32.
- [76] Zanotti G, D'Acunzio MR, Malpeli G, Folli C, Berni R. Crystal structure of the transthyretin-retinoic-acid complex. *Eur J Biochem* 1995;234:563–9.
- [77] Vercauteren I, Geldhof P, Peelaers I, Claerebout E, Berx G, Vercruysse J. Identification of excretory-secretory products of larval and adult *Ostertagia ostertagi* by immunoscreening of cDNA libraries. *Mol Biochem Parasitol* 2003;126:201–8.
- [78] Hoerauf A, Volkmann L, Hamelmann C, et al. Endosymbiotic bacteria in worms as targets for a novel chemotherapy in filariasis. *Lancet* 2000;355:1242–3.
- [79] Foster J, Ganatra M, Kamal I, et al. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* 2005;3:e121.
- [80] Hresko MC, Schriefer LA, Shrimankar P, Waterston RH. Myotactin, a novel hypodermal protein involved in muscle-cell adhesion in *Caenorhabditis elegans*. *J Cell Biol* 1999;146:659–72.
- [81] Hartmann S, Adam R, Marti T, Kirsten C, Seidinger S, Lucius R. A 41-kDa antigen of the rodent filaria *Acanthocheilonema viteae* with homologies to tropomyosin induces host protective immune responses. *Parasitol Res* 1997;83:390–3.
- [82] Jenkins RE, Taylor MJ, Gilvary NJ, Bianco AE. Tropomyosin implicated in host protective responses to microfilariae in onchocerciasis. *Proc Natl Acad Sci USA* 1998;95:7550–5.
- [83] Hintz M, Schares G, Taubert A, et al. Juvenile female *Litomosoides sigmodontis* produce an excretory/secretory antigen (Juv-p120) highly modified with dimethylaminoethanol. *Parasitology* 1998;117:265–71.

## Appendix 2B

Declaration from James Hewitson attesting my contribution to paper 2.



UNIVERSITY  
*of York*

## Declaration of reviewee's contribution

### Paper

Hewitson, J.P., Harcus, Y.M., Curwen, R.S., **DOWLE, A.A.**, Atmadja, A.K., Ashton, P.D., Wilson, A., Maizels, R.M. (2008) The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory-secretory products *Mol Biochem Parasitol.*, **160(1)**, 8-21

### Candidate's contribution

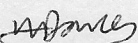
The submitting candidate performed the mass spectrometry component of the paper. Gel spots, bands and protein extracts were provided by James Hewitson. The reviewee's contribution to the work included: protein digestion, MALDI-MS/MS, LC-MALDI-MS/MS, database searching and associated data analysis. The aim of the proteomics analysis was to compare both within and between fractions using emPAI. The work was provided as fee for service through the Bioscience Technology Facility, Department of Biology, University of York. Contributions to data analysis and input into writing the paper were provided on a collaborative basis.

### Declaration

I attest that the summation above is a true reflection of the candidate's (Adam A. Dowle) contribution to the paper.

### Candidate

Name: Adam Dowle

Signature: 

Date: 1<sup>st</sup> June 2016

### Co-author

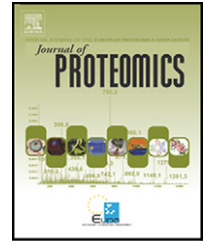
Name: James Hewitson

Signature: 

Date: 01.06.16

## Appendix 3A

Submitted paper 3 – Proteomic analysis of secretory products from the model gastrointestinal nematode *Heligmosomoides polygyrus* reveals dominance of venom allergen-like (VAL) proteins.

available at [www.sciencedirect.com](http://www.sciencedirect.com)[www.elsevier.com/locate/jprot](http://www.elsevier.com/locate/jprot)

# Proteomic analysis of secretory products from the model gastrointestinal nematode *Heligmosomoides polygyrus* reveals dominance of Venom Allergen-Like (VAL) proteins

James P. Hewitson<sup>a</sup>, Yvonne Harcus<sup>a</sup>, Janice Murray<sup>a</sup>, Maaïke van Agtmaal<sup>a</sup>, Kara J. Filbey<sup>a</sup>, John R. Grainger<sup>a,1</sup>, Stephen Bridgett<sup>b</sup>, Mark L. Blaxter<sup>b</sup>, Peter D. Ashton<sup>d</sup>, David A. Ashford<sup>d</sup>, Rachel S. Curwen<sup>c</sup>, R. Alan Wilson<sup>c</sup>, Adam A. Dowle<sup>d</sup>, Rick M. Maizels<sup>a,\*</sup>

<sup>a</sup>Institute of Immunology and Infection Research, University of Edinburgh, Edinburgh EH9 3JT, UK

<sup>b</sup>Gene Pool, Ashworth Laboratories, University of Edinburgh, Edinburgh EH9 3JT, UK

<sup>c</sup>Department of Biology, University of York, York YO10 5DD, UK

<sup>d</sup>Technology Facility, Proteomics Laboratory, Department of Biology, University of York, York YO10 5DD, UK

## ARTICLE INFO

Available online 29 June 2011

Keywords:

Parasite

Immunomodulation

Protease

Secretion

Signal peptide

## ABSTRACT

The intestinal helminth parasite, *Heligmosomoides polygyrus bakeri* offers a tractable experimental model for human hookworm infections such as *Ancylostoma duodenale* and veterinary parasites such as *Haemonchus contortus*. Parasite excretory–secretory (ES) products represent the major focus for immunological and biochemical analyses, and contain immunomodulatory molecules responsible for nematode immune evasion. In a proteomic analysis of adult *H. polygyrus* secretions (termed HES) matched to an extensive transcriptomic dataset, we identified 374 HES proteins by LC–MS/MS, which were distinct from those in somatic extract HEx, comprising 446 identified proteins, confirming selective export of ES proteins. The predominant secreted protein families were proteases (astacins and other metalloproteases, aspartic, cysteine and serine-type proteases), lysozymes, apyrases and acetylcholinesterases. The most abundant products were members of the highly divergent venom allergen-like (VAL) family, related to *Ancylostoma* secreted protein (ASP); 25 homologues were identified, with VAL-1 and -2 also shown to be associated with the parasite surface. The dominance of VAL proteins is similar to profiles reported for *Ancylostoma* and *Haemonchus* ES products. Overall, this study shows that the secretions of *H. polygyrus* closely parallel those of clinically important GI nematodes, confirming the value of this parasite as a model of helminth infection.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

Infection with intestinal nematode parasites such as hookworm, whipworm and *Ascaris* remains an enormous global health problem, with over 25% of the world's population

infected [1]. Moreover, similar pathogens account for major morbidity and economic loss among livestock in temperate climates [2]. The high prevalence and longevity of these parasites in immunocompetent hosts reflects a sophisticated array of mechanisms to modulate, disrupt and divert the host

\* Corresponding author at: Institute of Immunology and Infection Research, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. Tel.: +44 131 650 5511; fax: +44 131 650 5450.

E-mail address: [rick.maizels@ed.ac.uk](mailto:rick.maizels@ed.ac.uk) (R.M. Maizels).

<sup>1</sup> Current address: Mucosal Immunology Unit, Laboratory of Parasitic Diseases, NIAID, National Institutes for Health, Bethesda, MD 20892, USA.

immune response [3,4]. However, the identification of molecular mediators of parasite immunomodulation is still at an early stage [5–7]. For these reasons, the recent expansion in genomic [8–11], transcriptomic [12–21] and proteomic [22–30] analyses of parasitic nematodes provides an exciting platform for new discoveries.

A major theme in helminth research is the analysis of products released by live parasites which are likely to fulfil the many biological imperatives faced by a pathogen, including invasion of the host, creation of a suitable niche, and evasion of host immunity. These molecules, termed excretory–secretory (ES) products, have been the particular target of proteomic studies aimed at characterising the “secretome” of the major human [25–27] and veterinary [22,23,28–30] parasites. In addition, many prominent individual ES proteins have been identified, most notably members of a large multi-gene Venom Allergen-Like (VAL) family [28,31,32], first characterised in ES of the canine nematode *Ancylostoma caninum* and named *Ancylostoma* Secreted Protein (ASP) [33]. Members of this gene family include effective vaccine molecules in experimental models [34], indicating also the potential for ES proteins as new immunoprophylactics against helminth infections in man and animals.

Because the major human intestinal helminth species do not normally infect laboratory animals, model systems with natural rodent nematode parasites are invaluable in gaining insights into the factors regulating infection and immunity. The murine intestinal nematode parasite, *Heligmosomoides polygyrus*, provides a widely studied system [35], and much is now known of the immunology of infection and the immune components which combine to protect the host [36–40]. Parasite-infected mice feature multiple levels of immunosuppression, including amelioration of allergy [41,42], autoimmune diabetes [43,44] and colitis [45–48]. At least part of the immunosuppression can be accounted for by expanded regulatory T cell activity [42,49–51] and suppressive B cell populations [52] in infected mice, which are also able to transfer immunosuppression to uninfected recipients [42,52].

Significantly, the immunomodulatory effects of live *H. polygyrus* infection can be reproduced with the soluble products (HES) collected from adult parasites cultivated *in vitro*. HES converts naive murine T cells into suppressive regulatory T cells [53], interferes with the ability of dendritic cells to stimulate effector T cells and suppresses antibody responses to unrelated antigens [54], and can prevent the development of airway allergy in mice (O’Gorman, McSorley et al., manuscript in preparation). Hence the nature of the HES products is of intense interest for potential novel immunomodulators that might be exploited in therapy of allergy and autoimmunity. More broadly, intestinal nematodes co-habit a complex ecosystem with commensal microbes, and bacterial–parasite interactions are also likely to be important in the establishment of a long term nematode infection [55].

Despite the sophistication of the cellular immune analyses of *H. polygyrus* infection, few molecular products from this parasite have yet been described [24,56–60]. Indeed, genomic and transcriptomic datasets are only now being developed for this organism (Harcus et al., manuscripts in preparation). Taking a proteomic approach, we have identified the majority of proteins secreted by adult *H. polygyrus*, and show that there

is a predominance of VAL/ASP-like products, which demonstrates that the overall composition and functional profile of HES closely parallels those of *Ancylostoma* and *Haemonchus* parasites. These results pave the way to use the mouse model for more precise determination of the role of many individual proteins in the biological processes of infection, intestinal establishment, and manipulation of the host immune response.

## 2. Materials and methods

### 2.1. Parasites and HES

The original stock of *H. polygyrus bakeri* used in these studies was kindly supplied to us by Professor JM Behnke, University of Nottingham, UK. The life cycle of *H. polygyrus* was maintained in CBAx57BL/6F1 mice infected with 500 infective larvae by gavage, and adult worms were recovered 14 days later. Adult worms were washed extensively before incubation in serum-free RPMI1640 medium supplemented with 1% glucose, 100 U/ml penicillin, 100 µg/ml streptomycin, 2 mM L-glutamine, and 100 µg/ml gentamicin (Gibco). Culture supernatants were recovered at 3–4 day intervals and replaced each time with fresh medium over a 3 week period. Worms remained viable throughout this time frame. Pooled supernatants were diafiltrated into PBS over a 3000 MWCO Amicon membrane, and the resultant HES (*H. polygyrus* excretory/secretory products) material stored at –80 °C [59]. The profile of proteins released each week did not differ significantly (Supplementary Fig. A1). Soluble somatic extracts of adult worms (*H. polygyrus* extract; HEx) were prepared by homogenisation in a ground-glass hand-held homogeniser (VWR-Jencons, UK) in ice-cold PBS, followed by centrifugation at 13,000 g for 30 min, from which the supernatant was collected and stored at –80 °C until use.

### 2.2. 2-D gel electrophoresis and spot identification

HES and HEx (25 µg per gel) were separated and silver stained as previously described [25], then scanned with a Linoscan 1450 (Heidelberg). Protein spots of interest were prepared for mass spectrometry analysis as before [25], and positive-ion MALDI mass spectra were obtained using a Bruker ultraflex III in reflectron mode, equipped with a Nd:YAG smart beam laser. MS spectra were acquired over a mass range of m/z 800–4000, and monoisotopic masses were obtained using a SNAP averaging algorithm. The ten strongest peaks of interest, with a S/N greater than 30, were selected for MS/MS fragmentation in LIFT mode. Bruker flexAnalysis software (version 3.3) was used to perform the spectral processing and peak list generation for both the MS and MS/MS spectra.

### 2.3. LC–MS/MS

Tryptic HES peptides were prepared essentially as before [25] and then loaded onto a nanoAcquity UPLC system equipped with a nanoAcquity Symmetry C<sub>18</sub>, 5 µm trap (180 µm × 20 mm) and a nanoAcquity BEH130 1.7 µm C<sub>18</sub> capillary column (75 µm × 250 mm; all Waters). The trap was washed for 5 min with 0.1% (v/v) formic acid at 10 µL/min. Subsequently, flow



was switched to the capillary column, and peptides were separated by gradient elution (Solvent A=0.1% (v/v) formic acid; Solvent B=acetonitrile with 0.1% (v/v) formic acid; Initial gradient conditions 5% solvent B (2 min), then a linear gradient to 35% solvent B over 120 min, followed by a linear gradient to 50% solvent B over 5 min, and finally wash with 95% solvent B for 10 min. Flow rate was 300 nL/min and column temperature was 60 °C). The nanoLC system was interfaced with a maXis UHR-TOF mass spectrometer (Bruker Daltonics) with a nano-electrospray source fitted with a steel emitter needle (180 µm O.D. × 30 µm I.D.; Proxeon). Instrument control, data acquisition and processing were performed using Compass 1.3 SR3 software (microTOF control, Hystar and DataAnalysis; Bruker Daltonics). Positive ESI-MS & MS/MS spectra were acquired using AutoMSMS mode. Instrument settings were: ion spray voltage: 1500 V, dry gas: 6 L/min, dry gas temperature 160 °C, ion acquisition range: *m/z* 50–2200. AutoMSMS settings were: MS: 0.5 s (acquisition of survey spectrum), MS/MS (CID with N<sub>2</sub> as collision gas): ion acquisition range: *m/z* 300–1500, 5 precursor ions, absolute threshold 1000 counts, acquisition time: 0.1 s for precursor intensities ≥100,000 counts increasing linearly to 1 s for precursor intensities of 1000 counts, collision energy and isolation width settings were calculated automatically using the AutoMSMS fragmentation table, preferred charge states: 2–4, singly charged ions excluded, one fragmentation spectrum was acquired for each precursor and former target ions were excluded for 30 s.

#### 2.4. Database searching and bioinformatics

Tandem mass spectral data were submitted to database searching using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.1), through the Bruker ProteinScape interface (version 2.1). Search parameters required trypsin specificity, the carbamidomethylation of cysteine, allowed a maximum of one missed cleavage, and the possible oxidation of methionine. Spectra were searched against an in-house database composed of >460,000 cDNA sequences from both normalised and non-normalised libraries made from adult worm mRNA (Harcus et al., manuscript in preparation; <http://www.nematodes.org/nembase4/overview.shtml>). The database was supplemented with existing NCBI depositions for *H. polygyrus*. Sequencing was performed on a Roche 454 instrument yielding reads of ~200 nt, and assembled into isotigs each representing a distinct transcript using Newbler 2.5. For gel spot identifications a peptide tolerance of 250 ppm and MS/MS tolerance of 0.5 Da were employed. For LC-MS/MS Mascot searches, MudPit scoring was used with a peptide tolerance of 10 ppm and MS/MS tolerance of 0.1 Da. The significance threshold was set at *p*<0.05, and hits were manually inspected for the presence of open reading frames. All LC-MS/MS data were filtered to only accept peptides with expect values <0.05, and single peptide hits were further filtered requiring expect values <0.01. All protein matches were required to contain at least one unique peptide sequence not matched in any higher ranked proteins. The LC-MS/MS data were also searched against a Mascot generated decoy database, containing a random set of sequences with the same average amino acid composition and sequence length as the target database. Comparison of the number of sequences identified in the target and decoy databases estimated a false discovery rate of 2.01% HES and 2.94% for HEX

for peptide matches above identity threshold. Spectra were also searched against Swiss-Prot to identify potential murine or bacterial proteins present in HES and HEX. Here the false discovery rate was higher (33.1% for HES and 6.92% for HEX), likely reflecting the paucity of *H. polygyrus* sequences in the database, and their low level of sequence identity with other species. The exponentially Modified Protein Abundance Index (emPAI) for each identification was calculated according to the ratio of observed and observable peptides for each protein.

Identified protein sequences were subject to analysis by SignalP3.0 to ascertain presence of predicted signal peptide [61]. In some instances, protein sequences were judged to be truncated, by the absence of a start methionine and/or by homology to known protein sequences from other organisms. Proteins and conserved domains were identified by BLAST, and gene ontology (GO) categories determined with Interproscan version 31.0 (<http://www.ebi.ac.uk/Tools/pfa/iprscan/>). Protein sequences lacking conserved domains, but with significant similarity (BLAST score >40) to nematode proteins were labelled conserved nematode proteins (CSN=Conserved Secreted, No signal peptide; CSP=Conserved Secreted with signal-Peptide; CXN=Conserved eXtract, No signal peptide; CXP=Conserved eXtract with signal-Peptide). Novel sequences (BLAST score <40) were labelled NSN, NSP, NXN and NSP as described above. ClustalW sequence alignments were performed using MacVector version 11.1.1. To identify potential N- and O-glycosylation sites, NetNGlyc 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>) and NetOGlyc 3.1 (<http://www.cbs.dtu.dk/services/NetOGlyc/>) were used respectively. Accession numbers given for nucleotide and protein sequences are those deposited with ENA or NCBI.

#### 2.5. Antibody generation and Western blotting

HES (1 µg) was separated by 2-D gel electrophoresis as described above, blotted as before [25], and then blocked in 5% skimmed milk powder (Marvel)-TBS with 0.05% Tween 20 (TBST) for 2 h at room temperature. Polyclonal antibodies to Hp-VAL-1, -2 and -4 were generated to 6-His-tagged recombinant proteins expressed in pET21 (Novagen)-transformed *E. coli*, solubilised in 8 M urea and purified by metal chelating chromatography under the same chaotropic conditions; rats were immunised with 100 µg of recombinant protein co-precipitated with alum, boosted on days 28 and 35 with 100 µg protein in alum, and serum collected on day 42. To assess binding to HES, membranes were probed with 1/1000 sera dilutions in block solution overnight at 4 °C, washed extensively in TBST, and then with 1/2000 rabbit anti-rat Ig (1 h room temperature; DakoCytomation). Following further washing in TBST, blots were developed using ChemiGlow West, according to the manufacturer's instructions (Alpha Innotech) and imaged using a FluorChem SP (Alpha Innotech).

#### 2.6. Surface radio-iodination, immunoprecipitation and surface staining

Adult *H. polygyrus* were surface radio-labelled as described in earlier publications [62] but using Pierce Iodination Reagent (Iodogen) as the catalyst for generating nascent iodine [63]. Eppendorf tubes (1.5 ml) were coated with 200 µl of a 1 mg/ml solution of Iodination reagent (Pierce) in chloroform, dried, washed with PBS, before transfer of approximately 500 adult

worms and 500  $\mu\text{Ci}$   $^{125}\text{I}$ odine (Perkin Elmer) on ice. The sample was incubated with frequent agitation for 10 min, quenched by the addition of a saturated solution of L-tyrosine (Sigma), and surface radio-labelled parasite material produced as for HEx as described above, except that parasites were homogenised in PBS containing 1.5% nOG detergent and 1% protease inhibitor cocktail (Sigma P8340). Surface labelled parasite proteins were then separated by 2-D gel electrophoresis as above, and then dried and autoradiographed as before [62]. Immunoprecipitates were performed following pre-clearing of radiolabelled parasite extract with Protein G agarose beads (Millipore, 16–266) in the presence of MOPC 31 C IgG1 isotype control (for mouse anti-VAL monoclonal antibodies) or naïve rat serum (for rat anti-VAL polyclonal serum) for 30 min at room temperature. Unbound parasite material was then incubated with 2  $\mu\text{g}$  of mAb to VAL-1 (clone 3–36), VAL-2 (clone 4-S4), VAL-4 (clone 2–11) or MOPC 31 control IgG1 in non-denaturing IP buffer (20 mM Tris pH 8, 150 mM NaCl, 1 mM EDTA, 10% glycerol, 1% Triton-X100) for 2 h, then with Protein G agarose beads overnight, at 4 °C with rotation.

The production and specificity of anti-VAL mAb from the spleens of infected mice is to be described elsewhere [64]. Alternatively, 5  $\mu\text{l}$  polyclonal anti-VAL-1, 2, 4 rat sera or naïve rat sera was used. Beads were washed 5  $\times$  5 min in IP buffer, and bound proteins eluted by boiling in NuPAGE LDS sample buffer (Invitrogen)/0.5 M 2-mercaptoethanol, before separation on 1-D SDS-PAGE and autoradiograph as before [64]. For sections, adult *H. polygyrus* worms were snap-frozen on dry ice in Cryo-M-Bed mountant (Bright Instruments), cryostat sections (5  $\mu\text{m}$ ; Leica) cut onto Polysine™ slides (VWR), dried and then fixed in 100% acetone for 10 min. Sections were washed twice with PBS for 10 min, and then incubated with the mouse mAb described above (50  $\mu\text{g}/\text{mL}$  in 1% FCS/PBS) for 2 h at room temperature, washed twice in PBS as before, and then incubated with secondary anti-mouse Ig TRITC (1/100 in PBS) for 1 h at room temperature. Sections were washed extensively and then mounted in anti-fade Vectashield mountant (Vector Labs), before imaging with an Olympus fluorescent microscope.

### 3. Results

#### 3.1. Mass spectrometric identification of HES proteins from 2D PAGE gels

*H. polygyrus* adult excretory–secretory (HES) products were collected from parasites cultured in serum-free medium and concentrated over a 3000 MW cut-off membrane. To determine the identity of HES proteins, we used both 2-dimensional gel electrophoresis (2DGE) and ‘shotgun’ proteomics approaches. HES contained over 100 discernable polypeptides when analysed by silver staining of 2DGE (Fig. 1A), in a pattern clearly distinct from that of *H. polygyrus* somatic soluble protein extract (HEx, Fig. 1B). Analysis of 53 of the spots visible by 2DGE of HES (Fig. 1C) provided identities when matched against an in-house transcriptomic database of adult *H. polygyrus* mRNA (composed of 466,844 Roche 454 sequence reads, Harcus et al., manuscript in preparation). Additional spots were examined but did not provide sufficient material for MS identification (data not shown).

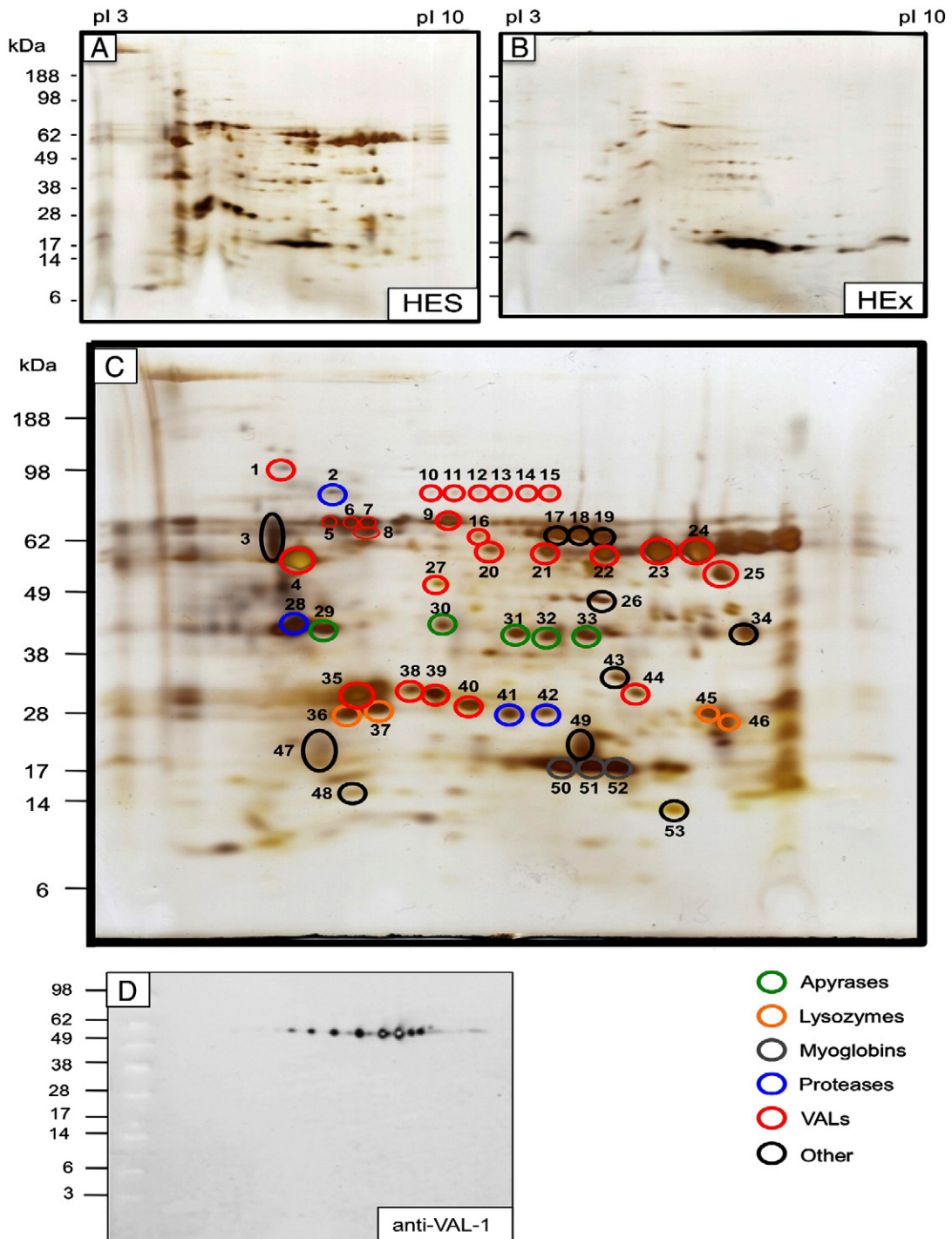
The most abundant products, as judged by intensity of silver staining, were found to be members of the VAL/ASP gene family (e.g. spots 4–9, 20–25, 35, 38–40), as well as apyrases, lysozymes, myoglobins and proteases. Table 1 summarises the full list of parasite proteins identified in this manner, which also includes a galectin, vitellogenin, chitinase, enolase and two novel gene sequences. We also observed that some proteins were present in multiple spots (e.g. variants of VAL-1 and VAL-2 were present in spots 20–24 and 5–9, respectively, and apyrase-2 in spots 31–33). Consistent with this, a polyclonal rat serum generated against recombinant VAL-1.1 recognised a chain of at least 8 spots by Western blot (Fig. 1D). Although several secreted HES proteins show micro-variation at the amino acid level (see below), this does not fully account for the observed differences in pI, as the same variant may be present in several different spots (e.g. VAL-1.2 is present in spots 21–24), and conversely more than one sequence variant was often seen in the same spot (see Table 1 and Supplementary Fig. A2), as the amino acid polymorphisms identified do not lead to large changes in predicted pI (data not shown).

#### 3.2. Mass spectrometric identification by LC–MS/MS

For a more exhaustive analysis of HES components, we employed LC–MS/MS on a total of 40  $\mu\text{g}$  of HES, resulting in the identification of a total of 374 secreted HES proteins; 100 of the most abundant (as ranked by Mascot score) are presented in Table 2, with the full listing given in Supplementary Table A1. Among the HES proteins identified were a selection of proteases, particularly metalloproteases (zinc metalloproteases and a large number of astacins), cysteine proteases (cathepsin B, legumain and necpain), aspartyl proteases (necepsin), and various serine proteases (cathepsin A, dipeptidyl peptidase four, serine carboxypeptidases and trypsin family proteins). Several other classes of enzymes were abundantly secreted, including relatively high levels of acetylcholinesterases, apyrases, chitinases and lysozymes. Protease inhibitors (cystatins, Kunitz inhibitors and serpins), transthyretin-related (TTR) proteins, chondroitin proteoglycans, and lectins (both C-type lectins and galectins) were also detected, as were a large number of proteins of unknown function with homologues in other nematodes (conserved nematode proteins) and novel proteins as yet unidentified in other helminths. In addition, as discussed further below, no fewer than 25 distinct VAL proteins were identified. Most of the abundant proteins had been localised by 2DGE to specific spots, as indicated where appropriate in Table 2. Searching of LC–MS/MS data against Swiss-Prot revealed that HES contained two different Ig kappa light chains (accession numbers P01654 and P01837) and a single IgG1 heavy chain (P01869). No significant matches were found to any bacterial protein sequences. We confirmed the presence of trace amounts of host immunoglobulin by ELISA, which showed that 1  $\mu\text{g}$  of HES contains  $0.135 \pm 0.010$  ng murine IgG1 (data not shown). It is likely that this antibody is bound to the adult worm *in vivo*, and subsequently dissociates from the parasite during the *in vitro* culture period. No murine proteins were detected in HEx.

In parallel, the soluble somatic protein extract (HEx) of the worm was analysed, yielding 446 identities thereby greatly extending analysis beyond previously available information





**Fig. 1 – 2-Dimensional analysis of *H. polygyrus* secreted proteins (HES) and soluble somatic extract (HEx). A. HES silver stain. B. HEx silver stain. Note that the major spots of 15–18 kDa have previously been identified as myoglobins [24]. C. HES annotated with spots analysed by MS/MS as presented in Table 1. D. Anti-VAL-1 rat polyclonal antibody on Western blot.**

[24]. Table 3 presents 100 of the most abundant (ranked by Mascot score) and a full listing is given in Supplementary Table A2. Complete mass spectrometric data for HES and HEx are presented in Supplementary Tables A3 and A4 respectively. In contrast to HES, many of the HEx products were ribosomal proteins and protein synthesis factors, as well as cytoskeletal components (actin, tropomyosin and tubulin) and also cytosolic

enzymes involved in glycolysis, lipid binding and redox reactions. GO annotation of HES and HEx (Fig. 2 and Supplementary Table A5) indicated that translation (GO:0006412) was the most common “biological process” term for HEx, and structural component of ribosome (GO:0003735) was the most common “molecular function”, indicative of the large number of ribosomal proteins present. In contrast, proteolysis

**Table 1 – Proteins identified in HES from spots.**

Spot	Identity	Domain	Hpb adult database	Score	Peptides	Accession No
1	VAL-3.1	Double SCP	isotig05790	64	2	JF914909
2	MEP-1 Zinc metalloprotease	Peptidase M13	isotig02967	144	5	(truncated)
3	CHI-1 Chitinase	Chitinase GH19	isotig04840	268	3	HE573246
4	VAL-3.1	Double SCP	isotig05790	462	6	JF914909
5	VAL-2.1/2.2/2.3 (a)	Double SCP	isotig03106	43	1	JF914906
6	VAL-2.1/2.2/2.3 (a)	Double SCP	isotig03106	266	7	JF914906
7	VAL-2.1/2.2/2.3 (a)	Double SCP	isotig03106	206	6	JF914906
8	VAL-2.1/2.2/2.3 (a)	Double SCP	isotig03106	32	1	JF914906
9	VAL-2.1/2.2/2.3 (a)	Double SCP	isotig03106	429	7	JF914906
10	VAL-5	Double SCP	isotig04839	186	8	JF914911
11	VAL-5	Double SCP	isotig04839	146	10	JF914911
12	VAL-5	Double SCP	isotig04839	132	8	JF914911
13	VAL-5	Double SCP	isotig04839	145	5	JF914911
14	VAL-5	Double SCP	isotig04839	169	7	JF914911
15	VAL-5	Double SCP	isotig04839	121	4	JF914911
16	VAL-9	Double SCP	isotig05765	495	10	JF914917
17	ACE-1 Acetylcholinesterase	Esterase lipase	isotig05694	21	1	JF439067
18	ACE-1 Acetylcholinesterase	Esterase lipase	isotig05694	32	2	JF439067
19	ACE-1 Acetylcholinesterase	Esterase lipase	isotig05694	53	2	JF439067
20	VAL-1.1	Double SCP	isotig06320	16	1	JF914902
21A	VAL-1.2	Double SCP	isotig03069	43	1	JF914903
21B	VAL-1.4	Double SCP	isotig01653	38	1	JF914905
22A	VAL-1.2	Double SCP	isotig03069	129	4	JF914903
22B	VAL-1.1	Double SCP	isotig06320	103	4	JF914902
22 C	VAL-1.4	Double SCP	isotig01653	70	2	JF914905
23A	VAL-1.2	Double SCP	isotig03069	419	5	JF914903
23B	VAL-1.1	Double SCP	isotig06320	402	5	JF914902
24A	VAL-1.1	Double SCP	isotig06320	613	6	JF914902
24B	VAL-1.2	Double SCP	isotig03069	385	5	JF914903
25	VAL-6	Double SCP	isotig03505	401	4	JF914912
26	Enolase (b)	Enolase	isotig06965	103	3	(truncated)
27	VAL-8.1	Double SCP	isotig02308	257	6	JF914916
28	MEP-3 Zinc metalloprotease	Peptidase M13	isotig05155	177	3	(truncated)
29	APY-1.1 Apyrase	Apyrase	isotig02250	68	3	JF721961
30	APY-3 Apyrase	Apyrase	isotig03589	188	5	JF721966
31	APY-2 Apyrase	Apyrase	isotig07051	213	6	JF721965
32	APY-2 Apyrase	Apyrase	isotig07051	323	3	JF721965
33	APY-2 Apyrase	Apyrase	isotig07051	185	5	JF721965
34	PHP-1, PHA-domain protein	PHA02954	isotig03547	180	4	HE573241
35	VAL-4	Single SCP	isotig10387	127	3	JF914910
36A	LYS-1 Lysozyme	Muramidase GH25	isotig08802	411	6	HE573247
36B	LYS-3 Lysozyme	Muramidase GH25	isotig08606	126	3	HE573249
37A	LYS-1 Lysozyme	Muramidase GH25	isotig08802	483	6	HE573247
37B	LYS-3 Lysozyme	Muramidase GH25	isotig08606	170	3	HE573249
38A	VAL-7.3	Single SCP	isotig02284	94	3	JF914915
38B	VAL-7.2	Single SCP	isotig01525	61	2	JF914914
39	VAL-7.2	Single SCP	isotig01525	48	1	JF914914
40	VAL-7.1	Single SCP	isotig01524	296	6	JF914913
41	NAS-1.1/1.2 Nematode Astacin (c)	ZnMc astacin-like	isotig04178	280	4	(truncated)
42	NAS-3.3 Nematode Astacin (d)	ZnMc astacin-like	isotig01791	141	2	(truncated)
43	Galectin	Double GLECT	isotig09082	219	4	(truncated)
44	VAL-10	Single SCP	isotig04979	27	1	JF914918
45	LYS-2 Lysozyme	Muramidase GH25	isotig05074	220	4	HE573248
46	LYS-2 Lysozyme	Muramidase GH25	isotig05074	179	3	HE573248
47	Vitellogenin	VWD	contig00471	64	2	(truncated)
48	NSP-4 Novel Secreted Protein	None	isotig11873	60	1	HE573242
49	NSP-16 Novel Secreted Protein	None	isotig05257	39	2	HE573243
50	Myoglobin	Globin	isotig04274	64	1	HE573244
51	Myoglobin	Globin	isotig04274	125	3	HE573244
52	Myoglobin	Globin	isotig04274	94	3	HE573244
53	TTR-1 Transthyretin-related	DUF290	isotig04612	452	6	HE573245

(GO:0006508) and metalloendopeptidase activity (GO:0004222) were the most common biological process and molecular function terms for HES. It is important to note that a greater number of proteins in HES compared to HEx failed to match GO terms (molecular function 54.8% HES Vs 23.5% HEx; biological process 68.7% HES Vs 41.3% HEx), consistent with the specialised nature of the parasite secretions unique to its intestinal niche, and in contrast to the somatic extract which is generally composed of proteins common to most eukaryotic organisms.

### 3.3. Comparison of proteins in HES and HEx reveals preferentially secreted proteins

Only 104 of 374 (27.8%) HES proteins were detectable in HEx (Table 4). The most abundant of these “somatic” proteins present in HES were the myoglobins and vitellogenins, both of which are extremely highly expressed by adult worms, the former representing the dominant species on 2D profiles (Fig. 1B). Hence, while the secreted components in HES in general represent a selective subset of the whole worm proteome, abundant somatic constituents such as myoglobin are also found in the *in vitro* parasite culture medium. In this regard it is noteworthy that both myoglobin and vitellogenin contain N-terminal signal sequences. It is also possible that as core egg proteins, vitellogenins diffuse from the eggs released by adult females during *in vitro* culture, or are present in intrauterine contents that accompany egg release. Comparison of exponentially modified Protein Abundance Index (emPAI) values for the 104 proteins common to HES and HEx indicated that there were large differences in the level of certain proteins between the two parasite preparations (*e.g.* VAL-1 and VAL-2 variants were highly abundant in HES and detected at only trace amounts in HEx; Table 4). This is not unexpected given that parasite secretions originate from the worm itself.

### 3.4. HES and HEx components differ significantly in proportion of predicted signal peptide sequences and in extent of novel gene products

Of the 374 HES proteins identified, 291 (77.8%) contained a predicted N-terminal signal peptide (Fig. 3). In addition, 100 (26.7%) did not correspond to any annotated gene in the NCBI database, and of these 70 (18.7%) were novel proteins with no database match. The remaining 30 sequences matched predicted or hypothetical proteins of unknown function from *C. elegans* or other nematodes, and were classified as conserved nematode proteins. When these 100 secreted proteins of unknown function were examined for the presence of a potential signal peptide, approximately 85% of each (60/70 novel, 25/30 conserved) were signal peptide-positive, indicating that an important set of novel secreted proteins are

present in HES. The novel proteins in particular were mostly <150 amino acids, with an average predicted molecular weight of 16.5 kDa (range 5.4–55.5 kDa). In contrast to HES, only 25.1% (112/446) of proteins identified in HEx encoded signal peptides, and HEx also contained noticeably fewer proteins of unknown function, with only 15 conserved nematode proteins (of which 4 were also detected in HES) and 9 novel proteins (4 detected in HES).

### 3.5. VAL proteins are associated with the parasite surface

In parallel with HES analysis, we also investigated the adult parasite surface proteins accessible to radio-iodination of live worms. When surface iodination was employed, solubilised proteins were analysed by 2D gel electrophoresis and autoradiography, and showed mobilities similar to VAL-1 and -2 (compare Fig. 4A with Fig. 1C) which could be specifically immunoprecipitated with monoclonal and polyclonal antibodies specific for VAL-1 and VAL-2 (Fig. 4B), indicating that both VAL-1 and VAL-2, highly enriched in HES, are present on the surface of the adult worm. In contrast, we did not detect surface expression of the similarly abundantly secreted VAL-4.

The distribution of VAL-1 and -2 was then investigated using monoclonal antibodies specific for each protein [64] to stain frozen sections of adult worms (Fig. 4C–F). Both antibodies gave a highly restricted punctate pattern of labelling the body wall, in a series of structures at contralateral sites that may represent longitudinal neuronal fibres or secretory tissues. Interestingly, in *A. caninum*, different VAL (ASP) products showed distinct localization patterns, with anti-Ac-ASP-4 staining the cuticle of adult worms, while antibodies to ASP-3 and -6 bind to glandular structures, and anti-ASP-5 to the gut, yet all four proteins are also found in adult worm ES [65].

### 3.6. Sequence analysis of the *H. polygyrus* VAL gene family

The most striking characteristic of HES is the predominance of members of the VAL gene family, both in terms of the abundance of certain members, particularly VAL-1, 2, 3, 4 and 7, as well as the large number (25) of different VAL proteins detected. VAL proteins show a conserved overall structure built around the SCP modular domain of ~200 amino acids (sperm-coating protein; pfam accession PF00188), typically containing 5 disulphide bonds. Members of this gene family generally contain either a single SCP domain, or two tandem domains, which are not necessarily closely related to each other in sequence. Fig. 5 shows a schematic of the 25 secreted VAL proteins of adult *H. polygyrus*, including 21 full length proteins, 8 of which (VAL-4, 7, 10, 15, 19, 21, 22 and 25) are single domain and 13 double domain (VAL-1, 2, 3, 5, 6, 8, 9, 12, 13, 14, 16, 17 and 20). Phylogenetically, the single- and double-

#### Notes to Table 1

Accession numbers from NCBI GenBank or European Nucleotide Archive (ENA) are given where full-length sequences are available.

(a) Peptides are common to all three variants of VAL-2; isotig and Accession number given are for VAL-2.1. Isotig and accession numbers for VAL-2.2 and -2.3 are isotig07425/JF914907 and isotig03105/JF914908.

(b) 1 additional peptide matches isotig18996 coding for N-terminal segment of Enolase separated from isotig 06965 by gap of ~27 nt.

(c) Peptides are common to NAS-1.1 and -1.2; latter is isotig04179.

(d) Two additional peptide matches to each of variants NAS 3.1 (isotig01799) and NAS-3.4 (isotig 01793).

**Table 2 – Top 100 proteins identified in HES by LC-MS/MS.**

Rank	Identity	Conserved domains	Code	Score	SP
1	VAL-1.1	SCP (Double)	isotig06320	6546	SP
2	VAL-1.2	SCP (Double)	isotig03069	5788	SP*
3	VAL-2.2	SCP (Double)	isotig07425	5149	SP*
4	VAL-2.3	SCP (Double)	isotig03105	5039	SP
5	VAL-3.1	SCP (Double)	isotig05790	4754	SP
6	VAL-2.1	SCP (Double)	isotig03106	4003	SP
7	Vitellogenin	Vitellogenin, DUF1943	contig00471	3997	SP
8	LYS-1 Lysozyme	Muramidase GH25	isotig08802	3615	SP
9	VAL-7.2	Single SCP	isotig01525	3369	SP
10	VAL-7.5	Single SCP	isotig01526	3367	SP
11	VAL-7.3	Single SCP	isotig02284	3051	SP
12	LYS-2 Lysozyme	Muramidase GH25	isotig05074	2947	SP
13	VAL-1.3	Double SCP	isotig06456	2668	SP
14	VAL-7.4	Single SCP	isotig02282	2604	SP
15	VAL-7.1	Single SCP	isotig01524	2593	SP
16	TTR-1 Transthyretin-related	DUF290	isotig04612	2130	SP
17	VAL-1.4	Double SCP	isotig01653	1841	SP
18	APY-1.1 Apyrase	Apyrase	isotig02250	1833	SP
19	APY-1.2 Apyrase	Apyrase	isotig07986	1818	SP
20	VAL-13	Double SCP	isotig06642	1809	SP
21	VAL-4	Single SCP	isotig10387	1770	SP
22	MEP-1 Zinc metalloprotease	Peptidase M13	isotig02967	1624	SP*
23	VAL-1.5	Double SCP	isotig01652	1596	SP*
24	NSP-1 Novel Secreted Protein with SP	None	isotig11973	1584	SP
25	APY-1.3 Apyrase	Apyrase	isotig05261	1583	SP
26	Myoglobin	Globin	isotig04274	1579	SP
27	Peritrophin-A-like protein	Chitin-binding type 2	Isotig05677	1561	SP*
28	Vitellogenin	DUF1943	contig00207	1552	SP*
29	ACE-1 Acetylcholinesterase	Esterase lipase	isotig05694	1527	SP
30	NPA-1 Nematode polyprotein allergen	None	isotig02438	1498	SP
31	APY-2 Apyrase	Apyrase	isotig07051	1474	SP
32	Myoglobin	Globin	isotig11742	1446	SP
33	VAL-9	Double SCP	isotig05765	1431	SP
34	Myoglobin	Globin	isotig04273	1390	SP
35	VAL-11	Single SCP	isotig05330	1303	SP*
36	MSP-1 Major Sperm Protein	Motile Sperm	isotig01565	1278	NO
37	MEP-2 Zinc metalloprotease	Pep0, Peptidase M13 N	isotig05366	1232	SP
38	Vitellogenin	Vitellogenin N, LPN N	contig00203	1205	SP*
39	Vitellogenin	VWD	contig00477	1204	SP*
40	VAL-8.1	Double SCP	isotig02308	1200	SP
41	TTR-2 Transthyretin-related	DUF290	isotig13558	1156	SP
42	NAS-1.1 Nematode Astacin protease	ZnMc astacin-like	isotig04178	1148	SP*
43	ACE-2 Acetylcholinesterase	Esterase lipase	isotig00868	1122	SP
44	PHP-1 PHA domain protein	PHA02954	isotig03547	1120	SP
45	NAS-1.2 Nematode Astacin protease	ZnMc astacin-like	isotig04179	1095	SP*
46	Myoglobin	Globin	isotig05169	1094	SP
47	VAL-8.2	Double SCP	isotig02308	1040	SP*
48	MEP-3 Zinc metalloprotease	Peptidases M13, M13N	isotig05155	1007	SP*
49	ACE-3 Acetylcholinesterase	Esterase lipase	isotig00869	1006	SP
50	Deoxyribonuclease II	Dnase_II	isotig08122	1006	SP*
51	VAL-6	Double SCP	isotig03505	996	SP
52	CHI-1 Chitinase	Chitinase GH19	isotig04840	988	SP
53	Astacin protease (fragment)	None	isotig01219	956	SP*
54	Trypsin family protein	Trypsin SPc	isotig03474	919	NO
55	Vitellogenin	VWD	contig00212	918	SP*
56	Aspartyl protease (necepsin)	Pepsin/retropepsin-like	isotig06497	884	SP
57	Trypsin family protein	Trypsin SPc	isotig03473	883	NO
58	VAL-1.6	Double SCP	isotig01655	867	SP
59	NAS-2.1 Nematode Astacin protease	ZnMc astacin-like	isotig01336	858	SP*
60	MFH-1 Ascaris MFP2b homologue	MFP2b	isotig07283	845	NO
61	NAS-2.2 Nematode Astacin protease	ZnMc astacin-like	isotig01334	836	SP*
62	NAS-3.1 Nematode Astacin protease	ZnMc astacin-like	isotig01799	832	SP*
63	NAS-4 Nematode Astacin protease	ZnMc astacin-like	isotig00214	830	SP*
64	Enolase	Enolase	isotig06965	826	NO
65	VAL-17	Double SCP	isotig05765	825	SP*



Table 2 (continued)

Rank	Identity	Conserved domains	Code	Score	SP
66	TTR-3 Transthyretin-related	DUF290	isotig14171	821	SP
67	MEP-4 Zinc metalloprotease	Peptidases M13, M13 N	isotig05402	806	SP
68	VAL-12	Double SCP	isotig06637	803	SP
69	NSP-2 Novel Secreted Protein with SP	None	isotig12022	802	SP
70	NSP-3 Novel Secreted Protein with SP	None	isotig12137	802	SP
71	NAS-5.1 Nematode Astacin protease	ZnMc astacin-like, CUB	isotig01596	779	SP*
72	NSN-1 Novel Secreted Non-signal-peptide protein	None	isotig18405	777	NFL
73	VAL-5	None	isotig04839	746	SP
74	Chondroitin-like protein	None	isotig01493	710	SP
75	Vitellogenin	Vitellogenin N	contig00199	696	SP
76	NAS-5.2 Nematode Astacin protease	ZnMc astacin-like, CUB	isotig01598	691	SP*
77	VAL-14	Double SCP	isotig04896	689	SP
78	LYS-3 Lysozyme	Muramidase GH25	isotig08606	683	SP*
79	NAS-6 Nematode Astacin protease	ZnMc astacin-like, CUB	isotig07866	681	SP*
80	Myoglobin	Globin	isotig11666	683	SP
81	NSP-4 Novel Secreted Protein	None	isotig05257	681	SP
82	MFH-2 Ascaris MFP2b homologue	MFP2b	isotig07373	647	NO
83	APY-3 Apyrase-3	Apyrase	isotig03589	644	SP
84	Complement regulatory protein	CCP	isotig13827	632	SP
85	Aldolase	FBP Aldolase 1a	isotig04915	627	NO
86	Complement regulatory protein	CCP	isotig12552	611	SP*
87	NSP-5 Novel Secreted Protein	None	isotig12800	599	SP
88	NSP-6 Novel Secreted Protein	None	isotig12832	596	SP
89	NSP-7 Novel Secreted Protein	None	isotig12576	592	SP
90	NSN-2 Novel Secreted Non-signal-peptide protein	None	isotig14195	589	NFL
91	FAR-1 Fatty acid/retinol binding protein	Gp-FAR-1	isotig11475	577	SP*
92	HEX-1 Hexokinase	Hexokinases 1, 2	isotig05973	563	NO
93	NSP-8 Novel Secreted Protein	None	isotig16252	557	SP
94	NSP-9 Novel Secreted Protein	None	isotig09200	529	SP
95	SCP-1 Serine carboxypeptidase	Esterase lipase	isotig05975	529	SP
96	CSP-1 Conserved nematode Secreted protein	None	isotig13235	528	SP*
97	SPN-1 Serpin	Serpin	isotig08233	512	NFL
98	NAS-3.2 Nematode Astacin protease	ZnMc astacin-like	isotig01792	509	SP*
99	CSP-2 Conserved nematode Secreted Protein	None	isotig09016	502	SP
100	NAS-3.3 Nematode Astacin protease	ZnMc astacin-like	isotig01791	498	SP*

SCP denotes the PFAM Sperm Coat Protein domain found in either single or double formation in VAL proteins. SS denotes presence of predicted signal sequence; SS\* indicates that the sequence is truncated but homologous to known proteins containing signal sequences; NFL denotes transcripts which are not full length and for which no database homologues were found with signal sequences (in the case of novel proteins, no homologues were found at all). Note that VAL-8.1 corresponds to a sequence similar but not identical to isotig02308, while VAL-8.2 matches exactly.

domain proteins appear to have diversified independently, although the single-domain VAL-19 is most likely to have evolved from a double-domain ancestor (Fig. 5). Double-domain VAL proteins include an inter-domain linker “hinge” region, which in the case of VAL-1, 2 and 5 comprise multiple Ser/Thr residues bearing a common antigenic O-glycan [64]. Immunodominant serum antibodies target this glycan early in infection, although anti-peptide antibodies such as those tested in Fig. 4 are also represented [64]. Similar stretches of predicted O-glycosylation are present in other secreted VAL proteins, particularly VAL-11, 17 and 20 (Fig. 5).

Analysis of individual *H. polygyrus* VAL amino acid sequences reveals extensive variation between genes: for example, within the C-terminal domains of the 13 double-domain proteins, only 15/189 amino acid residues are completely conserved (8%), and a further 10 altered in only 1 gene sequence, while in the N-terminal SCP-1 domain there are only 7 identical positions (of which 6 are cysteines). Similarly, for the single domain VAL proteins there are only 8 fully conserved amino acids, and 7 differing at only 1 position. Additionally, transcriptomic studies have identified

significant micro-sequence variation within individual VAL proteins (Harcus et al. manuscript in preparation), and this was evident with respect to proteomic data. For example, five alternative VAL-1 sequences verified by proteomic data are presented in Fig. 6. Note the extensive sequence diversity, particularly in the first SCP domain, suggesting that domain 1 is either under diversifying selection, or that intra-genic recombination is taking place. Similar sequence variation, with multiple variants matching the peptide data, was observed for VAL-2 and VAL-7. As the full genome of this parasite has yet to be assembled, we cannot yet determine whether the micro-variation is due to either allelic polymorphism or recent duplication of the relevant gene loci. Sequence variation has previously been reported for VALs of *A. caninum* [66], *Cooperia punctata* [67] and *H. contortus* [22].

### 3.7. Non-VAL gene families represented in HES

In addition to the 25 VAL proteins, a number of conserved gene family proteins are well represented in HES with particular prominence of the following functional groups.

**Table 3 – Top 100 Proteins identified in HEx by LC-MS/MS.**

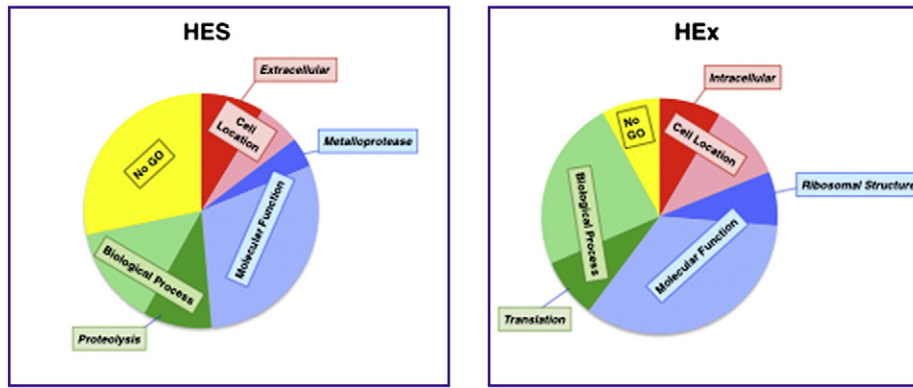
Rank	Identity	Code	Score	SS
1	Vitellogenin	contig00471	3369	
2	Myoglobin	isotig11742	3366	
3	Myoglobin	isotig13336	3098	
4	Actin	isotig01642	2585	
5	Actin	isotig01640	2251	
6	Myoglobin	isotig04274	2208	
7	Myoglobin	isotig11356	2113	
8	Myoglobin	isotig05169	1977	
9	Myoglobin	isotig04273	1833	
10	HSP90	isotig04714	1831	
11	Major sperm protein	isotig01565	1820	
12	Phosphoenolpyruvate carboxykinase	isotig05532	1713	
13	Aldolase	isotig04915	1611	
14	Myoglobin	isotig11666	1607	
15	Nematode polyprotein allergen NPA-1	isotig02438	1562	
16	Enolase	isotig06965	1476	
17	Protein disulphide isomerase	isotig05787	1465	✓
18	Vitellogenin	contig00207	1301	
19	Vitellogenin	contig00212	1295	
20	Eukaryotic translation elongation factor 1A	isotig06220	1223	
21	14-3-3 family member	isotig07975	1218	
22	Actin	isotig08179	1212	
23	Eukaryotic translation elongation factor 1A	isotig02523	1187	
24	Beta-tubulin	isotig06421	1179	
25	HSP70	isotig07134	1167	
26	Vitellogenin	contig00203	1040	
27	Vitellogenin	EST Hp_ADY001C04	1025	
28	Eukaryotic translation elongation factor 2	isotig05363	1015	
29	Ribosomal protein 60S P0	isotig01358	965	
30	HSP60	isotig06093	958	
31	Protein disulphide isomerase	isotig01916	902	✓
32	Vitellogenin	contig00199	884	
33	AAA family ATPase	isotig05378	857	
34	HSP90	isotig05404	851	
35	Serpin	isotig02225	788	
36	Ribosomal protein 40S S8	isotig10482	785	
37	Thioredoxin peroxidase	isotig05188	771	
38	Alpha-tubulin	isotig06550	736	
39	Arginine kinase	isotig06694	736	
40	Fumarase	isotig02533	717	
41	Myoglobin	isotig12632	698	
42	Actin	isotig08340	695	
43	Cystathionine beta-synthase	isotig05473	681	
44	Glyceraldehyde 3-phosphate dehydrogenase	isotig03503	646	
45	Calreticulin	Hpb-CRT	636	
46	Glutamate dehydrogenase	isotig06500	634	
48	HSP70	isotig05458	585	
49	Tropomyosin	Hpb-TRP	585	
50	Nucleosome assembly protein	isotig04959	573	
51	Chondroitin family member	isotig01493	571	
52	Malate dehydrogenase	isotig05076	561	
53	Phosphoglycerate_kinase	isotig06484	533	
54	Myoglobin	isotig13442	531	
55	Ribosomal protein 60S L6	isotig11208	524	
56	Vitellogenin	EST Hp_ADY_001G11	523	
57	Myoglobin (fragment)	isotig19571	521	

**Table 3 (continued)**

Rank	Identity	Code	Score	SS
58	Fatty acid and retinol binding protein	isotig11475	515	
59	Alpha-tubulin	isotig06165	512	
60	C-type lectin	isotig02337	499	✓
61	Phosphatidylethanolamine-binding protein	isotig05126	469	✓
62	Ribosomal protein 60S L18	isotig10630	465	
63	Ribosomal protein 40S S4	isotig05116	463	
64	CSN-2 Conserved secreted protein, No signal sequence	isotig04423	449	
65	Chondroitin family member	isotig06559	433	
66	Eukaryotic translation initiation factor 4A	isotig06491	432	
67	Retinol binding protein	Hpb-RBP	419	
68	Enolase (fragment)	isotig18996	404	
69	Nucleoside diphosphate kinase	isotig05236	401	
70	Ribosomal protein 60S L7a	isotig09109	397	
71	CSN-5 Conserved secreted protein, No signal sequence	isotig03664	393	
72	Glutathione S-transferase	isotig10292	389	
73	Macrophage migration inhibitory factor	isotig14093	383	
74	Vitellogenin	contig00211	372	
75	Cyclophilin	isotig11124	364	
76	Dehydrogenase	isotig03780	363	
77	Cytochrome C	isotig11144	361	
78	Ribosomal protein 60S L14	isotig14543	359	
79	RACK family member	isotig07754	358	
80	Aspartyl protease inhibitor	isotig05063	355	
81	Glutathione S-transferase	isotig07073	351	
82	VAL-3.1	isotig05790	349	✓
83	Ribosomal protein 60S L10	isotig12696	346	
84	Transketolase	isotig02953	346	
85	HSP70	isotig13132	345	
86	Ribosomal protein 40S S3	isotig09688	345	
87	Aldehyde dehydrogenase family member	isotig04738	339	
88	Oxidoreductase	isotig08193	323	
89	Glutathione S-transferase	isotig11588	322	
90	Adenosylhomocysteinase	isotig08907	308	
91	TTR-13 Transthyretin-related	isotig13284	304	
92	Glutathione S-transferase	isotig04128	302	
93	Ureidopropionase	isotig08079	300	
94	HSP70	isotig08182	296	
95	CSN-6 Conserved secreted protein, No signal sequence	isotig04424	288	
96	Fructose-1,6-bisphosphatase	isotig07764	284	
97	NAC domain-containing protein	isotig10507	283	
98	Alpha-tubulin	isotig06090	282	
99	Isocitrate lyase-malate synthase	isotig05351	282	
100	Ribosomal protein 60S L12	isotig11621	278	

### 3.7.1. Acetylcholinesterases

Three acetylcholinesterases are among the 100 most abundant HES proteins (Table 2), with Hpb-ACE-1 represented by 3 distinct spots (Table 1). Two additional proteins, ACE-2 and -3, differ by only 18 amino acids (3.1%) from each other but show ~32% divergence from ACE-1. The secretion of AChE by adult



**Fig. 2 – GO distribution of the proteins in HES and HEx.** All identified proteins in HES and HEx were analysed by Gene Ontology and categorised firstly into 4 broad categories: cell location (Red), molecular function (Blue), biological process (Green), and no recognised GO similarity (Yellow). Within each category, the most frequent term is shown (darker colours). Some proteins are included in more than one category. A more detailed listing of GO identifications is given in Supplementary Table A3.

*H. polygyrus* has been previously reported [68] and is a general feature of most nematode ES products [69]. In *N. brasiliensis*, the expression of three AChE isoforms has been shown to be differentially regulated according to the immune status of the host [70–73], and the 3 *H. polygyrus* proteins show highest similarity to isoform A of *N. brasiliensis*.

### 3.7.2. Apyrases

Four distinct apyrases were identified, one of which (Hpb-APY-1) was found as three minor sequence variants, all but one in the most abundant 100 HES proteins. Apyrases are adenosine diphosphatases, similar to mammalian CD73-like proteins, which catalyse the hydrolysis of ATP/ADP to AMP, and can often act also on other NDPs. Two recent reports have identified arthropod (*Cimex*)-like apyrases from related trichostrongylid nematodes *T. circumcincta* [74] and *Ostertagia ostertagi* [75] sharing 92% amino acid identity; in the latter case the enzyme was localised to the oesophageal glands. While all 4 *H. polygyrus* apyrases are homologous to these enzymes, levels of amino acid identity are less than 60%, indicating considerable diversification since divergence of the murine and ruminant parasites.

### 3.7.3. Lipid-binding proteins

Several structurally unrelated lipid-binding protein families are represented in HES. HES contains two distinct homologues of the Fatty acid/Retinol-binding (FAR) protein that has been recorded in ES of many other parasitic nematodes [76,77] and which in the case of *A. caninum* FAR-1 has been shown to functionally bind fatty acids and retinol [78]. A nematode polyprotein allergen (NPA) is also secreted by *H. polygyrus*, homologues of which bind the same ligands in *Ascaris* [79] and *Dictyocaulus viviparus* [80]. In addition, a total of 12 transthyretins are represented in HES, members of a widespread and diverse family of small proteins believed to recognise small hydrophobic ligands such as thyroid hormone, retinol or phosphatidylserine [81–83].

### 3.7.4. Lysozymes

Lysozymes or muramidases are present in multiple forms in HES, with 8 distinct gene products identified. All are related to

*C. elegans* and the 7 full-length sequences contain a potential signal peptide. In other organisms, lysozymes degrade the glycosidic bond linking *N*-acetylglucosamine and *N*-acetylmuramic acid in the murein proteoglycan of bacterial cell walls. Since adult *H. polygyrus* cohabit with microbial flora, and Gram-positive bacteria are more susceptible to lysozyme-mediated lysis, it is possible that these lysozymes modify the bacterial population sharing the intestinal niche of the worm.

### 3.7.5. Proteases

Five aspartyl proteases (PF00026) in HES correspond to a major enzymatic class from parasitic nematodes. Aspartyl proteases play a key role in the ability of *A. caninum* hookworms to degrade haemoglobin [84], and are the target of protective antibodies against the human parasite *Necator americanus* [85]. We identified no fewer than 20 astacins, Zn-metalloproteases distributed across the animal kingdom with particular frequency in nematodes [86,87].

Cysteine proteases have a major role in nematode parasites, particularly within the hookworm family. However, while 7 ES cathepsin B products have been defined in *H. contortus* [88], part of a larger gene family which show a particularly high level of transcription in intestinal tissue [89], only 4 are found in HES, a cathepsin B-like cysteine protease, a homologue of *N. americanus* necpain and two legumains (asparaginyl endopeptidases). Serine proteases are also less prominent in HES than aspartyl or metallo-enzymes, but include Cathepsin A, aminopeptidase, serine carboxypeptidases, trypsin family proteases and dipeptidyl peptidase four (DPF) proteins.

### 3.7.6. Protease inhibitors

Three broad classes of protease inhibitor are found in HES. Cystatins are a broadly conserved family of cysteine protease inhibitors found across plant and animal phyla, with an especial role in immune modulation by parasitic nematodes [90,91]. Inhibition of the protease active site involves a QVVG sequence which is perfectly conserved in the *H. polygyrus* homologue, as also in Nippocystatin from *N. brasiliensis* ES [92] and in *B. malayi* Bm-CPI-2 [93]. However, while Bm-CPI-2 has a second inhibitory motif (SND) that blocks the legumain

**Table 4 – Proteins common to both HES and HEx.**

Rank	Name	CODE	HES score	HES emPAI	HEx score	HEx emPAI	HES/HEx emPAI
1	VAL-2.1	isotig03106	5149	8.35	103	0.12	69.6
2	TTR-1 Transthyretin-related	isotig04612	2130	29.24	144	0.86	34.0
3	VAL-1.3	isotig06456	2668	1.97	47	0.06	32.8
4	VAL-7.5	isotig01526	3367	11.23	100	0.35	32.1
5	VAL-1.1	isotig06320	6546	4.05	77	0.18	22.5
6	NAS-3.1 Astacin protease	isotig01799	832	2.02	47	0.12	16.8
7	VAL-3.1	isotig05790	4754	3.66	349	0.25	14.6
8	VAL-7.1	isotig01524	2593	8.84	129	0.64	13.8
9	NSN-1 Novel secreted protein, No SP	isotig18405	777	4.03	40	0.31	13.0
10	Astacin protease family member	isotig01791	419	1.41	67	0.12	11.8
11	CSP-4 Conserved secreted protein with SP	isotig13290	482	1.82	37	0.16	11.4
12	VAL-4	isotig10387	1770	4.08	53	0.42	9.7
13	Aspartyl protease (necepsin)	isotig06497	884	1.17	136	0.18	6.5
14	Kunitz inhibitor	isotig01217	273	0.76	88	0.12	6.3
15	VAL-12	isotig06637	803	1.02	130	0.19	5.4
16	Astacin protease family member (fragment)	isotig01219	956	3.05	89	0.59	5.2
17	Superoxide dismutase	isotig09104	235	0.47	48	0.10	4.7
18	C-type lectin mannose receptor-like	isotig00496	366	0.97	61	0.23	4.2
19	MFH-1 Ascaris MFP2b homologue	isotig07283	845	1.17	137	0.29	4.0
20	Cystatin	isotig02700	261	0.41	71	0.12	3.4
21	CSP-2 Conserved secreted protein with SP	isotig09016	502	0.57	434	0.20	2.9
22	Motile sperm domain containing protein	isotig16097	175	0.50	144	0.22	2.3
23	Galectin	isotig09082	120	0.20	44	0.09	2.2
24	TIL domain-containing protein	isotig00779	98	0.27	42	0.13	2.1
25	Arginine kinase	isotig05009	212	0.65	193	0.33	2.0
26	Motile sperm domain containing protein	isotig16375	220	0.89	160	0.53	1.7
27	ERM family member	isotig02815	79	0.08	100	0.05	1.6
28	Cyclophilin	isotig09088	111	0.31	71	0.20	1.6
29	Actin depolymerising factor	isotig07522	145	0.23	144	0.15	1.5
30	Macrophage migration inhibitory factor	isotig14093	256	0.96	383	0.65	1.5
31	TTR-5 Transthyretin-like	isotig13207	397	0.89	88	0.61	1.5
32	MFH-3 Ascaris MFP2b homologue	isotig07980	463	0.84	148	0.58	1.4
33	TTR-10 Transthyretin-like	isotig11792	108	0.67	84	0.47	1.4
34	NSP-30 Novel secreted protein with SP	isotig12701	194	0.99	166	0.73	1.4
35	Chondroitin family member	isotig01493	710	0.48	571	0.38	1.3
36	Ferritin	isotig15727	117	0.74	140	0.73	1.0
37	NSP-39 Novel secreted protein with SP	isotig16333	108	0.78	187	0.77	1.0
38	Vitellogenin	contig00207	1552	5.73	1301	5.70	1.0
39	Myoglobin (fragment)	isotig19571	415	3.10	521	3.09	1.0
40	Calmodulin	isotig07710	53	0.07	57	0.07	1.0
41	Cathepsin-B like cysteine protease	isotig07017	147	0.06	136	0.06	1.0
42	Galectin	isotig09713	81	0.10	69	0.10	1.0
43	Kunitz inhibitor	isotig02721	82	0.13	82	0.13	1.0
44	Kunitz inhibitor	isotig02378	35	0.20	36	0.20	1.0
45	ML-domain containing protein	isotig10098	110	0.23	88	0.23	1.0
46	Motile sperm domain containing protein	isotig16212	138	0.22	92	0.22	1.0
47	NSP-59 Novel secreted protein with SP	isotig13319	38	0.16	35	0.16	1.0
48	Triose phosphate isomerase	isotig05016	276	0.44	198	0.44	1.0
49	VAL-15	isotig02149	133	0.20	200	0.20	1.0
50	Chondroitin family member	isotig08745	82	0.21	92	0.22	1.0
51	NPA-1 Nematode polyprotein allergen	isotig02438	1498	0.67	1562	0.75	0.9
52	Vitellogenin	contig00471	3997	3.05	3369	3.74	0.8
53	Nucleoside diphosphate kinase	isotig05236	462	1.18	401	1.55	0.8
54	TTR-6 Transthyretin related	isotig05212	304	0.62	210	0.83	0.7
55	Rab GDP dissociation inhibitor	isotig06129	54	0.16	168	0.22	0.7
56	Chondroitin family member	isotig06559	284	0.48	433	0.66	0.7
57	Vitellogenin	contig00199	696	2.38	884	3.29	0.7
58	Myoglobin	isotig11742	1446	3.31	3366	4.60	0.7
59	Enolase	isotig06965	826	2.13	1476	3.02	0.7
60	Transaldolase	isotig07176	59	0.14	62	0.21	0.7
61	Vitellogenin	contig00203	1205	4.87	1040	7.34	0.7
62	Aldolase	isotig04899	67	0.13	114	0.20	0.7
63	Vitellogenin (fragment)	sing00711	131	1.60	179	2.57	0.6
64	Myoglobin	isotig05169	1094	1.64	1977	2.64	0.6



Table 4 (continued)

Rank	Name	CODE	HES score	HES emPAI	HEX score	HEX emPAI	HES/HEX emPAI
65	Myoglobin	isotig04274	1579	3.13	2208	5.07	0.6
66	TTR-11 Transthyretin related	isotig04258	83	0.29	95	0.47	0.6
67	Ribosomal protein 60S L40/Ubiquitin	isotig13199	108	0.34	116	0.56	0.6
68	Serpin	isotig02225	281	0.33	788	0.59	0.6
69	Vitellogenin	contig00212	918	2.24	1295	4.24	0.5
70	Vitellogenin	contig00211	263	1.29	374	2.46	0.5
71	Chondroitin family member	isotig04525	51	0.13	75	0.27	0.5
72	Fatty acid/retinol binding protein	isotig11475	577	1.47	515	3.12	0.5
73	Fatty acid/retinol binding protein (fragment)	sing12372	71	0.27	78	0.61	0.4
74	Phosphatidylethanolamine-binding protein	isotig05126	339	0.64	469	1.67	0.4
75	Independent phosphoglycerate mutase	isotig01699	48	0.05	136	0.14	0.4
76	Vitellogenin	Hp_ADY_001G11	144	2.05	523	5.75	0.4
77	Cyclophilin	isotig11124	256	0.87	364	2.48	0.4
78	Thioredoxin peroxidase	isotig05188	498	1.79	771	5.19	0.3
79	Enolase (fragment)	isotig18996	326	1.08	404	3.30	0.3
80	Glutathione S-transferase	isotig11588	219	0.47	322	1.45	0.3
81	HSP70 (fragment)	isotig13132	148	0.35	345	1.11	0.3
82	Myoglobin	isotig11356	398	0.63	2113	2.00	0.3
83	Calreticulin	Hpb-CRT	74	0.21	636	0.67	0.3
84	Lipocalin domain-containing protein	isotig05316	104	0.34	229	1.09	0.3
85	Myoglobin	isotig11666	683	4.48	1607	14.53	0.3
86	CSN-5 Conserved secreted protein, No SP	isotig03664	43	0.08	393	0.27	0.3
87	Aldolase	isotig04915	627	0.77	1611	3.00	0.3
88	Major sperm protein	isotig01565	1278	2.10	1820	8.59	0.2
89	HSP60	isotig06093	195	0.43	958	1.76	0.2
90	Piwi domain-containing protein	isotig04875	77	0.06	167	0.25	0.2
91	14-3-3 family member	isotig07975	324	0.94	1218	3.98	0.2
92	CSN-2 Conserved secreted protein, No SP	isotig04423	127	1.08	449	4.75	0.2
93	Actin	isotig08340	102	0.27	695	1.38	0.2
94	Cytochrome C	isotig11144	50	0.13	361	0.86	0.2
95	Actin	isotig01640	340	0.47	2251	3.21	0.1
96	Eukaryotic translation elongation factor 2	isotig05363	86	0.09	1015	0.87	0.1
97	Protein disulfide isomerase	isotig01916	81	0.16	902	1.62	0.1
98	Eukaryotic translation elongation factor 1A	isotig02523	44	0.12	1187	1.64	0.1
99	Protein disulfide isomerase	isotig05787	133	0.24	1465	3.42	0.1
100	HSP70	isotig07134	100	0.22	1167	3.18	0.1
101	Glutathione S-transferase	isotig10292	47	0.24	389	3.88	0.1
102	Malate dehydrogenase	isotig05076	44	0.07	561	1.31	0.1
103	Ribosomal protein 60S P0	isotig01358	38	0.04	965	0.77	0.1
104	Retinol binding protein	Hpb-RBP	29	0.16	419	5.15	0.0

is altered in *H. polygyrus* (SNA) and may therefore not function in an identical fashion. Seven Kunitz type serine protease inhibitors are represented in HES, and again this gene is found in multiple forms in many helminth products [94]. Notably, a related secreted product from adult *A. ceylanicum* (AceK1) acts against a broad range of serine proteases, including trypsin, chymotrypsin and pancreatic elastase [95,96]. In *C. elegans*, a Kunitz-type inhibitor is important in collagen processing for cuticle formation, with mutations causing the blister-5 (*bli-5*) phenotype [97]. A structurally unrelated family of serine inhibitors are the serpins, large globular proteins that include most dominant ES antigen of *B. malayi* MF [98,99]. Three HES serpins were identified but are only distantly related to Bm-SPN-2, all being most similar to a serpin from *Trichostrongylus vitrinus*.

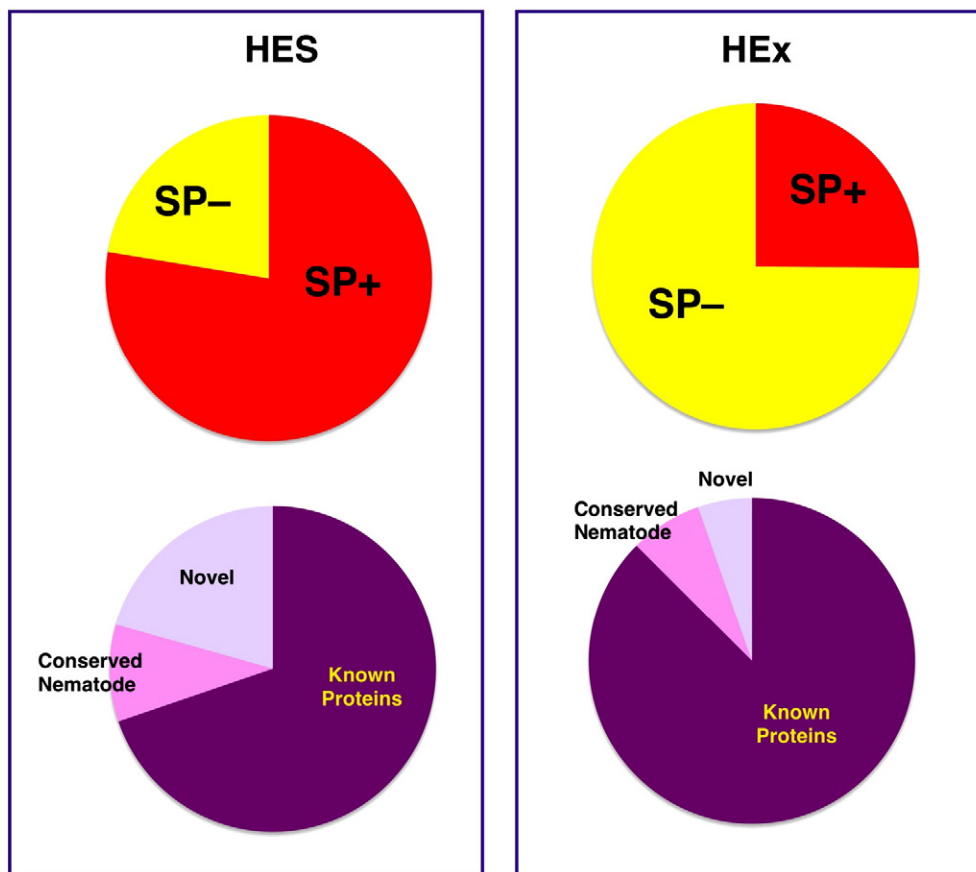
### 3.8. Similarity to other Strongylid nematodes

The overall profile of gene sets represented in HES shows many similarities to those reported in Trichostrongylid nematode

*T. circumcincta*[30] as well as the more distantly related hookworm *Ancylostoma caninum*[28], as shown in Table 5, and a number of specific homologies are noted above. In each species, VAL family members predominate, and similar findings have been reported for other members of the Trichostrongylid taxon such as *Ostertagia ostertagi* [100].

## 4. Discussion

The secretome of extracellular pathogens provides fascinating insights into the biological strategy of infectious organisms, in particular those such as long-lived helminth species that must attain an optimal physiological and immunological balance with their hosts [7]. The spectrum of parasite secreted proteins represents the elaborate adaptations demanded by the parasitic mode of life, over and above nematode-specific functions evident in free-living relatives such as *C. elegans*, many of which are likely to have evolved to interact with a precise host pathway, often with a specific ligand or receptor, and each of



**Fig. 3 – Novel genes and signal peptides of the proteins in HES and HEx. Upper diagrams: Distribution of signal peptide-containing protein sequences. Lower diagrams: Proportions of novel and nematode-conserved genes containing signal peptides.**

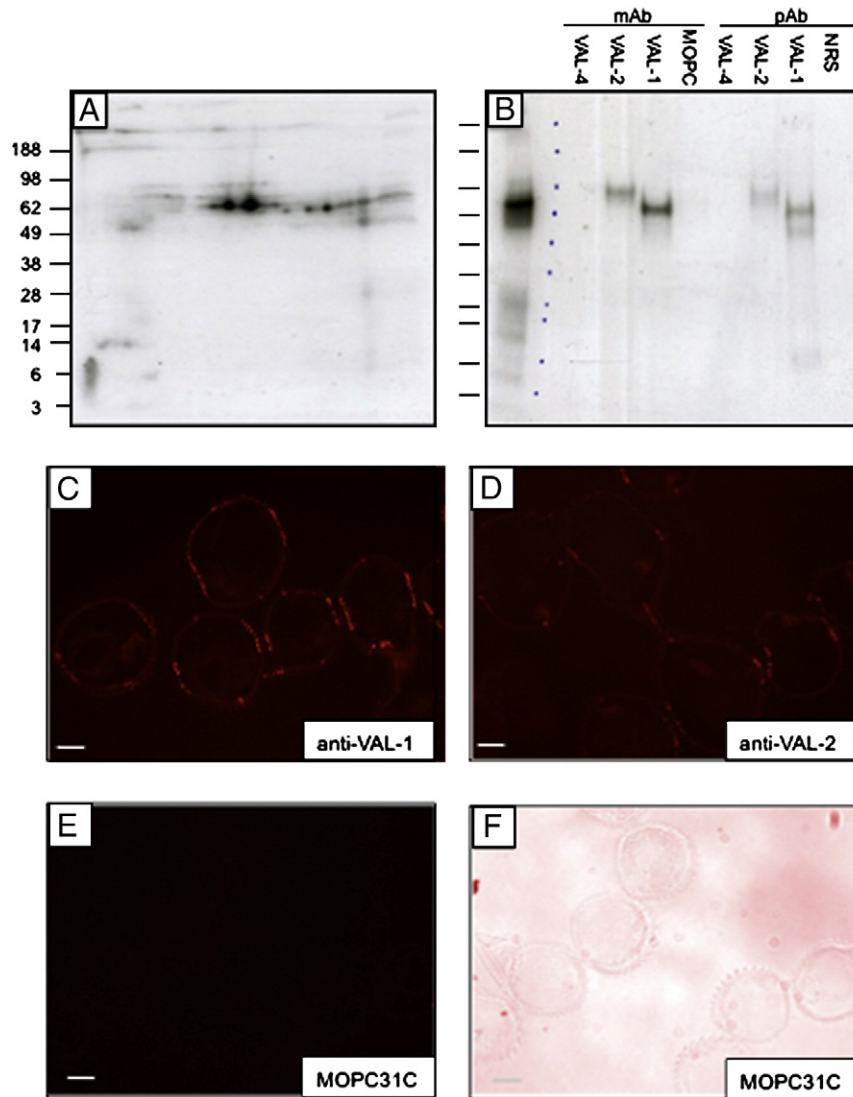
which may offer a novel and effective route to intervention and therapy.

The analytical power of modern proteomics enabled us to identify 374 ES products of adult *H. polygyrus*, holding open the prospect of a complete “worm pharmacopoeia” of potential host-modulating proteins [101]. Comparisons with the whole worm somatic protein extracts indicate a high level of selectivity, with many proteins only detectable in the secretions of the parasite, and conversely most somatic products being absent from HES. The likelihood that HES collected *in vitro* faithfully reflects release *in vivo*, is supported by recent work detecting circulating HES antigen in the serum of infected mice (Hewitson, unpublished), and demonstrating that host antibody responses are predominantly directed against HES, rather than HEx, antigens [64]. Validation of the methodological focus on ES products is further offered by the contrast in signal peptide-positive sequences within HES compared to HEx. Moreover, the concept that ES proteins may have evolved more rapidly to interact with host systems is supported by the higher number of novel (and novel signal peptide-positive) gene sequences among HES proteins than in the general body components, as previously noted in *N. brasiliensis* [13].

We also established that two of the major HES proteins, VAL-1 and -2, are represented on the surface of adult parasites. This finding reiterates work with other nematode species that

found concordance between surface and secreted proteins of adult *B. malayi* [63,102] and larval *T. canis* [103,104]. These studies raise interesting questions regarding the route of secretion by live parasitic nematodes: while they possess specialised secretory apparatus such as oesophageal glands [103,105], there is also evidence of direct trans-cuticular secretion deriving from the syncytial hypodermal tissue underlying the extracellular cuticle [106]. Now that the adult ES proteins are well defined, and with a number of monoclonal antibodies to these proteins [64], these issues can be directly addressed at the microscopic level.

The most striking feature of the HES analysis is the dominance of multiple VAL proteins. The VAL family is an extraordinary one in being so widely distributed in nature, with alternative names from diverse systems including Cysteine-Rich Secretory Protein (CRISP), Sperm Coat Protein (SCP) in mammals and plant Pathogenesis-Related protein (PR), without any clear indication as to their functional role(s) [107,108]. Within the nematodes, the archetypal VAL protein is the *Ancylostoma* secreted protein (ASP), a homologue of which (ASP-2) has been taken forward for human hookworm vaccine trials [109]. Interestingly, in *Ancylostoma* species from dogs and humans, multiple and divergent ASPs are known. For example, some 25 distinct VAL family transcripts were identified by Mitreva and colleagues from *A. caninum* [16]. For potential vaccines, it is important to target products of the initial



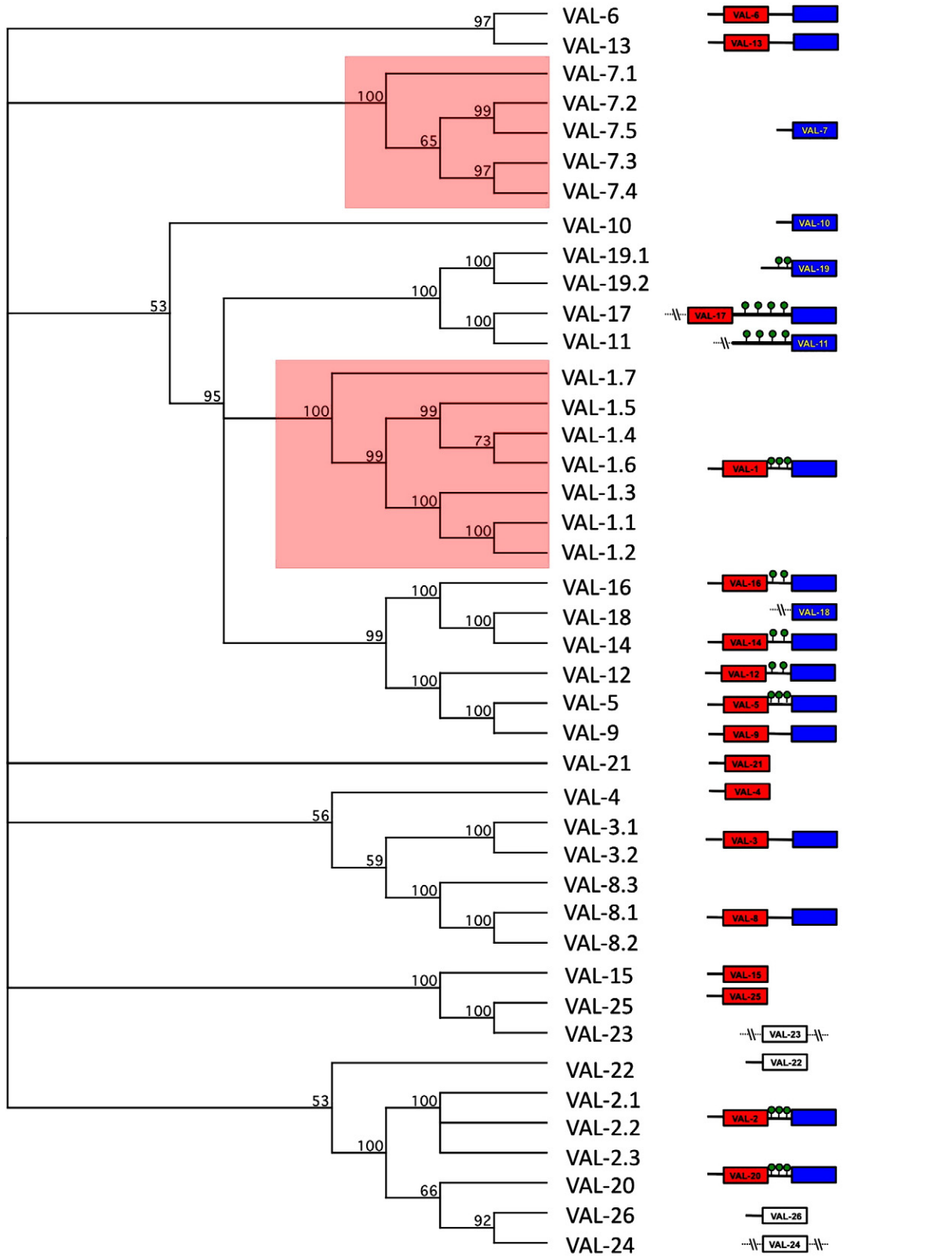
**Fig. 4 – Surface labelling of adult *H. polygyrus* reveals VAL-1 and VAL-2 are surface associated. A. Surface iodination, 2D gel. B. Immunoprecipitation of surface labelled VAL-1 and -2 with specific antibodies. C. Anti-VAL-1 monoclonal antibody on adult worm section. D. Anti-VAL-2 monoclonal antibody on adult worm section. E and F. MOPC control antibody on adult worm section with corresponding bright field image.**

infective stages as well as of established adults; thus future work with *H. polygyrus* will investigate VAL gene expression in immature as well as adult stages of the parasite. Additionally, we are currently assessing the protective potential of both the total adult secretions and individual VAL proteins.

VAL genes are also expressed in non-parasitic organisms, but with an intriguing association with the interface between different species, as for example in snake and insect venoms and haematophagous insect saliva, as well as in the response of plants to microbial infection. Moreover, numerous homologues exist in the free-living nematode *C. elegans* [108], arguing that these genes must have many functions outside the frame of host–parasite interactions. Indeed, *lon-1* is such a gene that controls body length downstream of TGF- $\beta$  signalling [110]. It is plausible, therefore, that the SCP domain is simply an adaptable protein framework that facilitates the evolution of

diverse specialised functions, and that such diversity is accentuated by inter-species interactions; certainly the extensive radiation within the *H. polygyrus* lineage is consistent with this notion. Intriguingly, VAL proteins found in HES differ from those of *C. elegans* in terms of both the enrichment of double domain proteins and in the presence of a serine/threonine rich linker region between SCP domains, the site of highly antigenic O-glycans in the HES proteins [64].

The comparison with related nematode species (Table 5) is instructive, not only in confirming that dominant VAL secretion is shared among different members of the taxon, but also identifying many common, and some particular, gene sets associated with intestinal parasitism. Most conspicuous across all these species is the level of protease production, predominantly astacins and other metalloproteases, but also including the aspartyl, cysteine and serine protease classes. Together with the





Signal Peptide cleavage

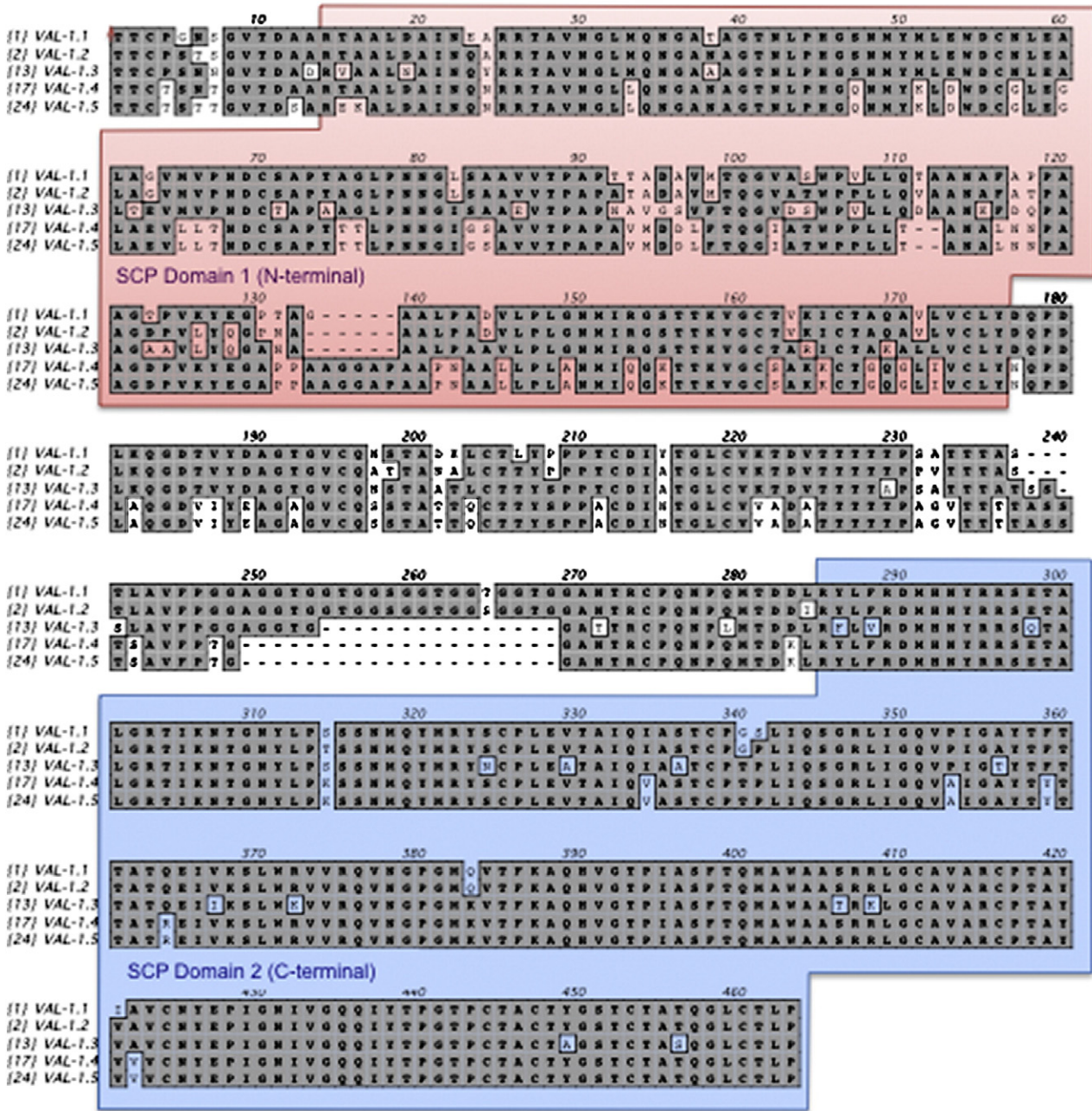


Fig. 6 – Sequence variation within *H. polygyrus* VAL-1 proteins. Mature proteins with signal cleavage site and N and C-terminal SCP domains indicated.

release of a series of protease inhibitors (cystatins, serpins, Kunitz inhibitors etc.) this suggests that nematodes can reset the hydrolytic and proteolytic environment of the gastrointestinal tract, possibly to avoid enzymatic attack but also quite probably to

degrade host mediators and obstacles such as mucins, immunoglobulins and innate defence molecules.

A further set of parallels are seen with the production of apyrases, with four different proteins identified. Whilst

Fig. 5 – Schematic of relatedness of *H. polygyrus* HES VAL-1-25, and VAL-26 identified in HEx. The domain structure of Hpb-VAL-1 to -26 are depicted including signal sequences, linker regions (predicted O-glycosylation is indicated with green circles) and SCP homology domains (N-terminal red, C-terminal blue). Single SCP domain proteins are coloured according to whether they are related to N (red) or C-terminal (blue) SCP domains. Divergent sequences equally distinct from both are white. Sequence truncation is indicated by (---).

**Table 5 – Table of common ES proteins between *H. polygyrus*, *Ancylostoma caninum*, *Haemonchus contortus* and *Teladorsagia circumcincta*.**

	<i>A. caninum</i>	<i>H. contortus</i>	<i>H. polygyrus</i>	<i>T. circumcincta</i> (L4 not adult)
	[28]	[22]	Hewitson, this paper	[30]
n	105	107	374	32
VAL/ASP	24	13	25	8
Metalloproteases/Astacins/Aminopeptidases	4	14	25	2
Transthyretins	5	2	12	0
Lysozymes	2	0	8	0
Serine proteases	0	8	8	0
Galectins	2	0	5	0
Aspartyl proteases	1	1	5	0
Cysteine proteases/necpains	3	0	4	13
CTLs	3	0	3	0
15-kDa ES	8	21	2	0
Cyclophilins	0	2	2	0
Protein disulphide isomerases	1	0	2	2
Nucleoside diphosphate kinases	0	1	1	0
GA1	0	21	0	0

apyrases have been associated with the inhibition of blood clotting by blood feeding insects [111], they do not appear to be secreted by blood feeding hookworms [28]. An alternative role may be immunomodulatory through conversion of pro-inflammatory ATP/ADP to anti-inflammatory AMP [112]. ATP, released by intestinal bacteria, can activate dendritic cells (DC) to secrete the pro-inflammatory cytokines IL-6 and IL-23, resulting in the induction of inflammatory Th17 cells [113]. Similarly, ATP-dependent DC activation is essential for Th2-dependent lung pathology in asthma models, and this can be inhibited through the administration of an apyrase [114]. In this context, adenosine generation may both induce Foxp3+ regulatory T cells and be used by this cell type as a mechanism of suppression [115]. Finally, in *C. elegans*, an endo-apyrase, APY-1, fulfils a more homeostatic role in the stress response of the organism [116], a role that cannot be excluded for *H. polygyrus* homologues.

Whilst we have catalogued the identity of several hundred of the most abundant proteins in the search for helminth immunomodulators, we did not detect expression of the TGF $\beta$  family member encoded by *H. polygyrus*, and expressed by all mammalian stages [60]. Such activity appears linked to the ability of HES to induce de novo Foxp3-positive regulatory T cells [53]. This may be because TGF $\beta$  is active at sub-nM levels, beyond the level of detection of our MS analysis. Alternatively, it is not known whether the TGF $\beta$  activity in HES is due to a true-TGF $\beta$  homolog. We are currently fractionating HES to determine this. Another potential immunomodulator previously considered is calreticulin, which has been reported to drive the Th2 response that is strongly provoked in *H. polygyrus* infection [58]; as calreticulin is present at only low levels in HES, but is readily detectable in worm extract, it remains to be determined if this product emanates from active secretion or leakage from compromised parasites.

In conclusion, we have embarked on a fine-detail molecular dissection of an important nematode parasite, which serves as an excellent model for both human and veterinary helminth diseases. Whilst RNAi has not yet been successfully demonstrated with *H. polygyrus* [117], more recent studies have

established that secretory proteins may be optimally positioned for silencing in this way [118], which should allow functional testing of the various potentially immunomodulatory proteins described here. It is important to note that so far we have analysed only adult HES, but that infection is initiated by L3 which embed in the intestinal submucosa and clearly elaborate a set of equally fascinating mediators; hence in due course we hope to analyse the ES of immature stages and complete the proteomic characterization of this organism.

Supplementary materials related to this article can be found online at [doi:10.1016/j.jprot.2011.06.002](https://doi.org/10.1016/j.jprot.2011.06.002)

## Acknowledgements

We thank the Wellcome Trust for Programme Grant support. KJF is supported by an MRC CASE studentship with UCB Celltech. JRG was supported by a Wellcome Trust PhD studentship.

## REFERENCES

- [1] Hotez PJ, Brindley PJ, Bethony JM, King CH, Pearce EJ, Jacobson J. Helminth infections: the great neglected tropical diseases. *J Clin Invest* 2008;118:1311–21.
- [2] Miller JE, Horohov DW. Immunological aspects of nematode parasite control in sheep. *J Anim Sci* 2006;84: E124–32 suppl.
- [3] Maizels RM, Yazdanbakhsh M. Regulation of the immune response by helminth parasites: cellular and molecular mechanisms. *Nat Rev Immunol* 2003;3:733–43.
- [4] Elliott DE, Summers RW, Weinstock JV. Helminths as governors of immune-mediated inflammation. *Int J Parasitol* 2007;37:457–64.
- [5] Harnett W, Harnett MM. Therapeutic immunomodulators from nematode parasites. *Expert Rev Mol Med* 2008;10:e18.
- [6] Adisakwattana P, Saunders SP, Nel HJ, Fallon PG. Helminth-derived immunomodulatory molecules. *Adv Exp Med Biol* 2009;666:95–107.
- [7] Hewitson JP, Grainger JR, Maizels RM. Helminth immunoregulation: the role of parasite secreted proteins in



- modulating host immunity. *Mol Biochem Parasitol* 2009;167:1–11.
- [8] Ghedin E, Wang S, Spiro D, Caler E, Zhao Q, Crabtree J, et al. Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* 2007;317:1756–60.
- [9] Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EG, Deleury E, et al. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat Biotechnol* 2008;26:909–15.
- [10] Opperman CH, Bird DM, Williamson VM, Rokhsar DS, Burke M, Cohn J, et al. Sequence and genetic map of *Meloidogyne hapla*: a compact nematode genome for plant parasitism. *Proc Natl Acad Sci USA* 2008;105:14802–7.
- [11] Mitreva M, Jasmer DP, Zarlenga DS, Wang Z, Abubucker S, Martin J, et al. The draft genome of the parasitic nematode *Trichinella spiralis*. *Nat Genet* 2011;43:228–35.
- [12] Tetteh KKA, Loukas A, Tripp C, Maizels RM. Identification of abundantly-expressed novel and conserved genes from infective stage larvae of *Toxocara canis* by an expressed sequence tag strategy. *Infect Immun* 1999;67:4771–9.
- [13] Harcus YM, Parkinson J, Fernández C, Daub J, Selkirk ME, Blaxter ML, et al. Signal sequence analysis of expressed sequence tags from the nematode *Nippostrongylus brasiliensis* and the evolution of secreted proteins in parasites. *Genome Biol* 2004;5:R39.
- [14] Mitreva M, Jasmer DP, Appleton J, Martin J, Dante M, Wylie T, et al. Gene discovery in the adenophorean nematode *Trichinella spiralis*: an analysis of transcription from three life cycle stages. *Mol Biochem Parasitol* 2004;137:277–91.
- [15] Parkinson J, Mitreva M, Whitton C, Thomson M, Daub J, Martin J, et al. A transcriptomic analysis of the phylum Nematoda. *Nat Genet* 2004;36:1259–67.
- [16] Mitreva M, McCarter JP, Arasu P, Hawdon J, Martin J, Dante M, et al. Investigating hookworm genomes by comparative analysis of two *Ancylostoma* species. *BMC Genomics* 2005;6:58.
- [17] Nisbet AJ, Redmond DL, Matthews JB, Watkins C, Yaga R, Jones JT, et al. Stage-specific gene expression in *Teladorsagia circumcincta* (Nematoda: Strongylida) infective larvae and early parasitic stages. *Int J Parasitol* 2008;38:829–38.
- [18] Yin Y, Martin J, Abubucker S, Scott AL, McCarter JP, Wilson RK, et al. Intestinal transcriptomes of nematodes: comparison of the parasites *Ascaris suum* and *Haemonchus contortus* with the free-living *Caenorhabditis elegans*. *PLoS Negl Trop Dis* 2008;2:e269.
- [19] Abubucker S, Zarlenga DS, Martin J, Yin Y, Wang Z, McCarter JP, et al. The transcriptomes of the cattle parasitic nematode *Ostertagia ostertagi*. *Vet Parasitol* 2009;162:89–99.
- [20] Cantacessi C, Mitreva M, Jex AR, Young ND, Campbell BE, Hall RS, et al. Massively parallel sequencing and analysis of the *Necator americanus* transcriptome. *PLoS Negl Trop Dis* 2010;4:e684.
- [21] Cantacessi C, Mitreva M, Campbell BE, Hall RS, Young ND, Jex AR, et al. First transcriptomic analysis of the economically important parasitic nematode, *Trichostrongylus colubriformis*, using a next-generation sequencing approach. *Infect Genet Evol* 2010;10:1199–207.
- [22] Yatsuda AP, Krijgsveld J, Cornelissen AWCA, Heck AJ, De Vries E. Comprehensive analysis of the secreted proteins of the parasite *Haemonchus contortus* reveals extensive sequence variation and differential immune recognition. *J Biol Chem* 2003;278:16941–51.
- [23] Craig H, Wastling JM, Knox DP. A preliminary proteomic survey of the in vitro excretory/secretory products of fourth-stage larval and adult *Teladorsagia circumcincta*. *Parasitology* 2006;132:535–43.
- [24] Morgan C, LaCourse EJ, Rushbrook BJ, Greetham D, Hamilton JV, Barrett J, et al. Plasticity demonstrated in the proteome of a parasitic nematode within the intestine of different host strains. *Proteomics* 2006;6:4633–45.
- [25] Hewitson JP, Harcus YM, Curwen RS, Dowle AA, Atmadja AK, Ashton PD, et al. The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory–secretory products. *Mol Biochem Parasitol* 2008;160:8–21.
- [26] Moreno Y, Geary TG. Stage- and gender-specific proteomic analysis of *Brugia malayi* excretory–secretory products. *PLoS Negl Trop Dis* 2008;2:e326.
- [27] Bennuru S, Semnani R, Meng Z, Ribeiro JM, Veenstra TD, Nutman TB. *Brugia malayi* excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling. *PLoS Negl Trop Dis* 2009;3:e410.
- [28] Mulvenna J, Hamilton B, Nagaraj SH, Smyth D, Loukas A, Gorman JJ. Proteomics analysis of the excretory/secretory component of the blood-feeding stage of the hookworm, *Ancylostoma caninum*. *Mol Cell Proteomics* 2009;8:109–21.
- [29] Smith SK, Nisbet AJ, Meikle LI, Inglis NF, Sales J, Beynon RJ, et al. Proteomic analysis of excretory/secretory products released by *Teladorsagia circumcincta* larvae early post-infection. *Parasite Immunol* 2009;31:10–9.
- [30] Nisbet AJ, Smith SK, Armstrong S, Meikle LI, Wildblood LA, Beynon RJ, et al. *Teladorsagia circumcincta*: activation-associated secreted proteins in excretory/secretory products of fourth stage larvae are targets of early IgA responses in infected sheep. *Exp Parasitol* 2010;125:329–37.
- [31] Murray J, Gregory WF, Gomez-Escobar N, Atmadja AK, Maizels RM. Expression and immune recognition of *Brugia malayi* VAL-1, a homologue of vespid venom allergens and *Ancylostoma* secreted proteins. *Mol Biochem Parasitol* 2001;118:89–96.
- [32] Asojo OA, Goud GN, Dhar K, Loukas A, Zhan B, Deumic V, et al. X-ray structure of Na-ASP-2, a pathogenesis related-1 protein from the nematode parasite, *Necator americanus*, and a vaccine antigen for human hookworm infection. *J Mol Biol* 2005;346:801–14.
- [33] Hawdon JM, Jones BF, Hoffman DR, Hotez PJ. Cloning and characterization of *Ancylostoma*-secreted protein. A novel protein associated with the transition to parasitism by infective hookworm larvae. *J Biol Chem* 1996;271:6672–8.
- [34] Hotez PJ, Zhan B, Bethony JM, Loukas A, Williamson A, Goud GN, et al. Progress in the development of a recombinant vaccine for human hookworm disease: the Human Hookworm Vaccine Initiative. *Int J Parasitol* 2003;33:1245–58.
- [35] Monroy FG, Enriquez FJ. *Heligmosomoides polygyrus*: a model for chronic gastrointestinal helminthiasis. *Parasitol Today* 1992;8:49–54.
- [36] Finkelman FD, Shea-Donohue T, Morris SC, Gildea L, Strait R, Madden KB, et al. Interleukin-4- and interleukin-13-mediated host protection against intestinal nematode parasites. *Immunol Rev* 2004;201:139–55.
- [37] Anthony RM, Rutitzky LI, Urban Jr JF, Stadecker MJ, Gause WC. Protective immune mechanisms in helminth infection. *Nat Rev Immunol* 2007;7:975–87.
- [38] McCoy KD, Stoel M, Stettler R, Merky P, Fink K, Senn BM, et al. Polyclonal and specific antibodies mediate protective immunity against enteric helminth infection. *Cell Host Microbe* 2008;4:362–73.
- [39] Herbert DR, Yang J-Q, Hogan SP, Groschwitz K, Khodoun MV, Munitz A, et al. Intestinal epithelial cell secretion of RELM- $\beta$  protects against gastrointestinal worm infection. *J Exp Med* 2009;206:2947–57.
- [40] Wojciechowski W, Harris DP, Sprague F, Mousseau B, Makris M, Kusser K, et al. Cytokine-producing effector B cells regulate type 2 immunity to *H. polygyrus*. *Immunity* 2009;30:421–33.
- [41] Bashir ME, Andersen P, Fuss IJ, Shi HN, Nagler-Anderson C. An enteric helminth infection protects against an allergic response to dietary antigen. *J Immunol* 2002;169:3284–92.

- [42] Wilson MS, Taylor M, Balic A, Finney CAM, Lamb JR, Maizels RM. Suppression of allergic airway inflammation by helminth-induced regulatory T cells. *J Exp Med* 2005;202: 1199–212.
- [43] Saunders KA, Raine T, Cooke A, Lawrence CE. Inhibition of autoimmune type 1 diabetes by gastrointestinal helminth infection. *Infect Immun* 2006;75:397–407.
- [44] Liu Q, Sundar K, Mishra PK, Mousavi G, Liu Z, Gaydo A, et al. Helminth infection can reduce insulinitis and type 1 diabetes through CD25- and IL-10-independent mechanisms. *Infect Immun* 2009;77:5347–58.
- [45] Khan WI, Blennerhasset PA, Varghese AK, Chowdhury SK, Omsted P, Deng Y, et al. Intestinal nematode infection ameliorates experimental colitis in mice. *Infect Immun* 2002;70:5931–7.
- [46] Elliott DE, Setiawan T, Metwali A, Blum A, Urban Jr JF, Weinstock JV. *Heligmosomoides polygyrus* inhibits established colitis in IL-10-deficient mice. *Eur J Immunol* 2004;34:2690–8.
- [47] Sutton TL, Zhao A, Madden KB, Elfrey JE, Tuft BA, Sullivan CA, et al. Anti-inflammatory mechanisms of enteric *Heligmosomoides polygyrus* infection against trinitrobenzene sulfonic acid-induced colitis in a murine model. *Infect Immun* 2008;76:4772–82.
- [48] Hang L, Setiawan T, Blum AM, Urban J, Stoyanoff K, Arihiro S, et al. *Heligmosomoides polygyrus* infection can inhibit colitis through direct interaction with innate immunity. *J Immunol* 2010;185:3184–9.
- [49] Metwali A, Setiawan T, Blum AM, Urban J, Elliott DE, Hang L, et al. Induction of CD8<sup>+</sup> regulatory T cells in the intestine by *Heligmosomoides polygyrus* infection. *Am J Physiol Gastrointest Liver Physiol* 2006;291:G253–9.
- [50] Finney CAM, Taylor MD, Wilson MS, Maizels RM. Expansion and activation of CD4<sup>+</sup>CD25<sup>+</sup> regulatory T cells in *Heligmosomoides polygyrus* infection. *Eur J Immunol* 2007;37:1874–86.
- [51] Rausch S, Huehn J, Kirchhoff D, Rzepecka J, Schnoeller C, Pillai S, et al. Functional analysis of effector and regulatory T cells in a parasitic nematode infection. *Infect Immun* 2008;76:1908–19.
- [52] Wilson MS, Taylor MD, O’Gorman MT, Balic A, Barr TA, Filbey K, et al. Helminth-induced CD19<sup>+</sup>CD23<sup>hi</sup> B cells modulate experimental allergic and autoimmune inflammation. *Eur J Immunol* 2010;40:1682–96.
- [53] Grainger JR, Smith KA, Hewitson JP, McSorley HJ, Harcus Y, Filbey KJ, et al. Helminth secretions induce de novo T cell Foxp3 expression and regulatory function through the TGF- $\beta$  pathway. *J Exp Med* 2010;207:2331–41.
- [54] Segura M, Su Z, Piccirillo C, Stevenson MM. Impairment of dendritic cell function by excretory–secretory products: a potential mechanism for nematode-induced immunosuppression. *Eur J Immunol* 2007;37:1887–904.
- [55] Belkaid Y, Liefenfeld O, Maizels RM. Induction and control of regulatory T cells in the GI tract: consequences for local and peripheral immune responses. *Clin Exp Immunol* 2009;160: 35–41.
- [56] Hoselton S, Piche L, Gustad T, Robinson M. Production of a recombinant version of a *Heligmosomoides polygyrus* antigen that is preferentially recognized by resistant mouse strains. *Parasite Immunol* 2002;24:429–35.
- [57] Rzepecka J, Lucius R, Doligalska M, Beck S, Rausch S, Hartmann S. Screening for immunomodulatory proteins of the intestinal parasitic nematode *Heligmosomoides polygyrus*. *Parasite Immunol* 2006;28:463–72.
- [58] Rzepecka J, Rausch S, Klotz C, Schnöller C, Kornprobst T, Hagen J, et al. Calreticulin from the intestinal nematode *Heligmosomoides polygyrus* is a Th2-skewing protein and interacts with murine scavenger receptor-A. *Mol Immunol* 2008;46:1109–19.
- [59] Harcus Y, Nicoll G, Murray J, Filbey K, Gomez-Escobar N, Maizels RM. C-type lectins from the nematode parasites *Heligmosomoides polygyrus* and *Nippostrongylus brasiliensis*. *Parasitol Int* 2009;58:461–70.
- [60] McSorley HJ, Grainger JR, Harcus YM, Murray J, Nisbet A, Knox DP, et al. *daf-7*-related TGF- $\beta$  homologues from trichostrongyloid nematodes show contrasting life cycle expression patterns. *Parasitology* 2010;137:159–71.
- [61] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004;340: 783–95.
- [62] Pritchard DI, Maizels RM, Behnke JM, Appleby P. Stage-specific antigens of *Nematospiroides dubius*. *Immunology* 1984;53: 325–35.
- [63] Maizels RM, Gregory WF, Kwan-Lim G-E, Selkirk ME. Filarial surface antigens: the major 29,000 mol.wt. glycoprotein and a novel 17,000–200,000 mol.wt. complex from adult *Brugia malayi* parasites. *Mol Biochem Parasitol* 1989;32:213–27.
- [64] Hewitson JP, Filbey KJ, Grainger JR, Dowle AA, Pearson M, Murray J, Harcus Y, Maizels RM. *Heligmosomoides polygyrus* elicits a dominant nonprotective antibody response directed at restricted glycan and peptide epitopes. Manuscript submitted for publication 2011.
- [65] Zhan B, Liu Y, Badamchian M, Williamson A, Feng J, Loukas A, et al. Molecular characterisation of the *Ancylostoma*-secreted protein family from the adult stage of *Ancylostoma caninum*. *Int J Parasitol* 2003;33:897–907.
- [66] Qiang S, Bin Z, Shu-hua X, Zheng F, Hotez P, Hawdon JM. Variation between ASP-1 molecules from *Ancylostoma caninum* in China and the United States. *J Parasitol* 2000;86:181–5.
- [67] Yatsuda AP, Eysker M, Vieria-Bressan MCR, De Vries E. A family of activation associated secreted protein (ASP) homologues of *Cooperia punctata*. *Res Vet Sci* 2002;73:297–306.
- [68] Lawrence CE, Pritchard DI. Differential secretion of acetylcholinesterase and proteases during the development of *Heligmosomoides polygyrus*. *Int J Parasitol* 1993;23:309–14.
- [69] Selkirk ME, Lazari O, Hussein AS, Matthews JB. Nematode acetylcholinesterases are encoded by multiple genes and perform non-overlapping functions. *Chem Biol Interact* 2005;157–158:263–8.
- [70] Blackburn CC, Selkirk ME. Characterisation of the secretory acetylcholinesterases from adult *Nippostrongylus brasiliensis*. *Mol Biochem Parasitol* 1992;53:79–88.
- [71] Grigg ME, Tang L, Hussein AS, Selkirk ME. Purification and properties of monomeric (G<sub>1</sub>) forms of acetylcholinesterase secreted by *Nippostrongylus brasiliensis*. *Mol Biochem Parasitol* 1997;90:513–24.
- [72] Hussein AS, Chacón MR, Smith AM, Tosado-Acevedo R, Selkirk ME. Cloning, expression, and properties of a nonneuronal secreted acetylcholinesterase from the parasitic nematode *Nippostrongylus brasiliensis*. *J Biol Chem* 1999;274:9312–9.
- [73] Hussein A, Harel M, Selkirk M. A distinct family of acetylcholinesterases is secreted by *Nippostrongylus brasiliensis*. *Mol Biochem Parasitol* 2002;123:125–34.
- [74] Nisbet AJ, Zarlenga DS, Knox DP, Meikle LI, Wildblood LA, Matthews JB. A calcium-activated apyrase from *Teladorsagia circumcincta*: an excretory/secretory antigen capable of modulating host immune responses? *Parasite Immunol* 2011.
- [75] Zarlenga DS, Nisbet AJ, Gasbarre LC, Garrett WM. A calcium-activated nucleotidase secreted from *Ostertagia ostertagi* 4th-stage larvae is a member of the novel salivary apyrases present in blood-feeding arthropods. *Parasitology* 2011;138:333–43.
- [76] Kennedy MW, Garside LH, Goodrick LE, McDermott L, Brass A, Price NC, et al. The Ov20 protein of the parasitic nematode *Onchocerca volvulus*. A structurally novel class of small helix-rich retinol-binding proteins. *J Biol Chem* 1997;272: 29442–8.



- [77] Garofolo A, Kläger SL, Rowlinson M-C, Nirmalan N, Klion A, Allen JE, et al. The FAR proteins of filarial nematodes: secretion, glycosylation and lipid binding characteristics. *Mol Biochem Parasitol* 2002;122:161–70.
- [78] Basavaraju S, Zhan B, Kennedy MW, Liu Y, Hawdon J, Hotez PJ. Ac-FAR-1, a 20 kDa fatty acid- and retinol-binding protein secreted by adult *Ancylostoma caninum* hookworms: gene transcription pattern, ligand binding properties and structural characterisation. *Mol Biochem Parasitol* 2003;126:63–71.
- [79] Kennedy MW, Brass A, McCrudden AB, Price NC, Kelly SM, Cooper A. The ABA-1 allergen of the parasitic nematode *Ascaris suum*: fatty acid and retinoid binding function and structural characterization. *Biochemistry* 1995;34:6700–10.
- [80] Kennedy MW, Britton C, Price NC, Kelly SM, Cooper A. The DvA-1 polyprotein of the parasitic nematode *Dictyocaulus viviparus*. A small helix-rich lipid-binding protein. *Journal of Biological Chemistry* 1995;270:19277–81.
- [81] Eneqvist T, Lundberg E, Nilsson L, Abagyan R, Sauer-Eriksson AE. The transthyretin-related protein family. *Eur J Biochem* 2003;270:518–32.
- [82] Henneby SC. Evolutionary changes to transthyretin: structure and function of a transthyretin-like ancestral protein. *FEBS J* 2009;276:5367–79.
- [83] Wang X, Li W, Zhao D, Liu B, Shi Y, Chen B, et al. *Caenorhabditis elegans* transthyretin-like protein TTR-52 mediates recognition of apoptotic cells by the CED-1 phagocyte receptor. *Nat Cell Biol* 2010;12:655–64.
- [84] Brinkworth RI, Prociw P, Loukas A, Brindley PJ. Hemoglobin-degrading, aspartic proteases of blood-feeding parasites: substrate specificity revealed by homology models. *J Biol Chem* 2001;276:38844–51.
- [85] Pearson MS, Bethony JM, Pickering DA, de Oliveira LM, Jariwala A, Santiago H, et al. An enzymatically inactivated hemoglobinase from *Necator americanus* induces neutralizing antibodies against multiple hookworm species and protects dogs against heterologous hookworm infection. *FASEB J* 2009;23:3007–19.
- [86] Hawdon JM, Jones BF, Perregaux MA, Hotez PJ. *Ancylostoma caninum*: metalloprotease release coincides with activation of infective larvae *in vitro*. *Exp Parasitol* 1995;80:205–11.
- [87] Williamson AL, Lustigman S, Oksov Y, Deumic V, Plieskatt J, Mendez S, et al. *Ancylostoma caninum* MTP-1, an astacin-like metalloprotease secreted by infective hookworm larvae, is involved in tissue migration. *Infect Immun* 2006;74:961–7.
- [88] Yatsuda AP, Bakker N, Krijgsveld J, Knox DP, Heck AJ, de Vries E. Identification of secreted cysteine proteases from the parasitic nematode *Haemonchus contortus* detected by biotinylated inhibitors. *Infect Immun* 2006;74:1989–93.
- [89] Jasmer DP, Mitreva MD, McCarter JP. mRNA sequences for *Haemonchus contortus* intestinal cathepsin B-like cysteine proteases display an extreme in abundance and diversity compared with other adult mammalian parasitic nematodes. *Mol Biochem Parasitol* 2004;137:297–305.
- [90] Hartmann S, Lucius R. Modulation of host immune responses by nematode cystatins. *Int J Parasitol* 2003;33:1291–302.
- [91] Klotz C, Ziegler T, Figueiredo AS, Rausch S, Hepworth MR, Obsivac N, et al. A helminth immunomodulator exploits host signaling events to regulate cytokine production in macrophages. *PLoS Pathog* 2011;7:e1001248.
- [92] Dainichi T, Maekawa Y, Ishii K, Zhang T, Nashed BF, Sakai T, et al. Nippocystatin, a cysteine protease inhibitor from *Nippostrongylus brasiliensis*, inhibits antigen processing and modulates antigen-specific immune response. *Infect Immun* 2001;69:7380–6.
- [93] Murray J, Manoury B, Balic A, Watts C, Maizels RM. Bm-CPI-2, a cystatin from *Brugia malayi* nematode parasites, differs from *C. elegans* cystatins in a specific site mediating inhibition of the antigen-processing enzyme AEP. *Mol Biochem Parasitol* 2005;139:197–203.
- [94] Gonzalez S, Flo M, Margenat M, Duran R, Gonzalez-Sapienza G, Grana M, et al. A family of diverse Kunitz inhibitors from *Echinococcus granulosus* potentially involved in host–parasite cross-talk. *PLoS One* 2009;4:e7009.
- [95] Milstone AM, Harrison LM, Bungiro RD, Kuzmic P, Cappello M. A broad spectrum Kunitz type serine protease inhibitor secreted by the hookworm *Ancylostoma ceylanicum*. *J Biol Chem* 2000;275:29391–9.
- [96] Chu D, Bungiro RD, Ibanez M, Harrison LM, Campodonico E, Jones BF, et al. Molecular characterization of *Ancylostoma ceylanicum* Kunitz-type serine protease inhibitor: evidence for a role in hookworm-associated growth delay. *Infect Immun* 2004;72:2214–21.
- [97] Page AP, McCormack G, Birnie AJ. Biosynthesis and enzymology of the *Caenorhabditis elegans* cuticle: identification and characterization of a novel serine protease inhibitor. *Int J Parasitol* 2006;36:681–9.
- [98] Zang XX, Yazdanbakhsh M, Kiang H, Kanost MR, Maizels RM. A novel serpin expressed by the blood-borne microfilariae of the parasitic nematode *Brugia malayi* inhibits human neutrophil serine proteinases. *Blood* 1999;94:1418–28.
- [99] Zang XX, Atmadja AK, Gray P, Allen JE, Gray CA, Lawrence RA, et al. The serpin secreted by *Brugia malayi* microfilariae, Bm-SPN-2, elicits strong, but short-lived, immune responses in mice and humans. *J Immunol* 2000;165:5161–9.
- [100] Saverwyns H, Visser A, Nisbet AJ, Peelaers I, Gevaert K, Vercruyse J, et al. Identification and characterization of a novel specific secreted protein family for selected members of the subfamily Ostertagiinae (Nematoda). *Parasitology* 2008;135:63–70.
- [101] Johnston MJG, Macdonald JA, McKay DM. Parasitic helminths: a pharmacopeia of anti-inflammatory molecules. *Parasitology* 2009;136:125–47.
- [102] Kwan-Lim G-E, Gregory WF, Selkirk ME, Partono F, Maizels RM. Secreted antigens of filarial nematodes: survey and characterisation of *in vitro* excretory/secretory (E/S) products of adult *Brugia malayi* filarial parasites. *Parasite Immunol* 1989;11:629–54.
- [103] Page AP, Hamilton AJ, Maizels RM. *Toxocara canis*: monoclonal antibodies to carbohydrate epitopes of secreted (TES) antigens localize to different secretion-related structures in infective larvae. *Exp Parasitol* 1992;75:56–71.
- [104] Page AP, Rudin W, Fluri E, Blaxter ML, Maizels RM. *Toxocara canis*: a labile antigenic coat overlying the epicuticle of infective larvae. *Exp Parasitol* 1992;75:72–86.
- [105] Wu Y, Egerton G, Pappins DJC, Harrison RA, Wilkinson M, Underwood A, et al. The secreted larval acidic proteins (SLAPs) of *Onchocerca* spp. are encoded by orthologues of the alt gene family of *Brugia malayi* and have host protective potential. *Mol Biochem Parasitol* 2004;134:213–24.
- [106] Selkirk ME, Gregory WF, Yazdanbakhsh M, Jenkins RE, Maizels RM. Cuticular localisation and turnover of the major surface glycoprotein (gp29) of adult *Brugia malayi*. *Mol Biochem Parasitol* 1990;42:31–44.
- [107] Gibbs GM, Roelants K, O'Bryan MK. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins—roles in reproduction, cancer, and immune defense. *Endocr Rev* 2008;29:865–97.
- [108] Cantacessi C, Campbell BE, Visser A, Geldhof P, Nolan MJ, Nisbet AJ, et al. A portrait of the “SCP/TAPS” proteins of eukaryotes — developing a framework for fundamental research and biotechnological outcomes. *Biotechnol Adv* 2009;27:376–88.
- [109] Bethony J, Loukas A, Smout M, Brooker S, Mendez S, Plieskatt J, et al. Antibodies against a secreted protein from hookworm larvae reduce the intensity of hookworm infection in

- humans and vaccinated laboratory animals. *FASEB J* 2005;19:1743–5.
- [110] Morita K, Flemming AJ, Sugihara Y, Mochii M, Suzuki Y, Yoshida S, et al. A *Caenorhabditis elegans* TGF-beta, DBL-1, controls the expression of LON-1, a PR-related protein, that regulates polyploidization and body length. *EMBO J* 2002;21:1063–73.
- [111] Valenzuela JG, Belkaid Y, Rowton E, Ribeiro JM. The salivary apyrase of the blood-sucking sand fly *Phlebotomus papatasi* belongs to the novel Cimex family of apyrases. *J Exp Biol* 2001;204:229–37.
- [112] Gounaris K, Selkirk ME. Parasite nucleotide-metabolizing enzymes and host purinergic signalling. *Trends Parasitol* 2005;21:17–21.
- [113] Atarashi K, Nishimura J, Shima T, Umesaki Y, Yamamoto M, Onoue M, et al. ATP drives lamina propria T(H)17 cell differentiation. *Nature* 2008;455:808–12.
- [114] Idzko M, Hammad H, van Nimwegen M, Kool M, Willart MA, Muskens F, et al. Extracellular ATP triggers and maintains asthmatic airway inflammation by activating dendritic cells. *Nat Med* 2007;13:913–9.
- [115] Fletcher JM, Lonergan R, Costelloe L, Kinsella K, Moran B, O'Farrelly C, et al. CD39+ Foxp3+ regulatory T Cells suppress pathogenic Th17 cells and are impaired in multiple sclerosis. *J Immunol* 2009;183:7602–10.
- [116] Uccelletti D, Pascoli A, Farina F, Alberti A, Mancini P, Hirschberg CB, et al. APY-1, a novel *Caenorhabditis elegans* apyrase involved in unfolded protein response signalling and stress responses. *Mol Biol Cell* 2008;19:1337–45.
- [117] Lendner M, Doligalska M, Lucius R, Hartmann S. Attempts to establish RNA interference in the parasitic nematode *Heligmosomoides polygyrus*. *Mol Biochem Parasitol* 2008;161:21–31.
- [118] Samarasinghe B, Knox DP, Britton C. Factors affecting susceptibility to RNA interference in *Haemonchus contortus* and in vivo silencing of an H11 aminopeptidase gene. *Int J Parasitol* 2011;41:51–9.

## Appendix 3B

Declaration from James Hewitson attesting to my contribution to paper 3.



UNIVERSITY  
*of York*

## Declaration of reviewee's contribution

### Paper

Hewitson, J. P., Harcus, Y., Murray, J., van Agtmaal, M., Filbey, K. J., Grainger, J. R., Bridgett, S., Blaxter, M. L., Ashton, P. D., Ashford, D. A., Curwen, R. S., Wilson, R. A., **DOWLE, A. A.** and Maizels, R. M. (2011) Proteomic analysis of secretory products from the model gastrointestinal nematode *Heligmosomoides polygyrus* reveals dominance of Venom Allergen-Like (VAL) proteins. *J. Proteomics* **74**, 1573-1594.

### Candidate's contribution

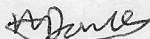
The submitting candidate performed the mass spectrometry component of the paper. Gel spots and protein extracts were provided by James Hewitson. The reviewee's contribution to the work included: protein digestion, MALDI-MS/MS, LC-MS/MS, database searching and associated data analysis. The aim of the proteomics analysis was to compare both within and between fractions using emPAI. The work was provided as fee for service through the Bioscience Technology Facility, Department of Biology, University of York. Contributions to data analysis and input into writing the paper were provided on a collaborative basis.

### Declaration

I attest that the summation above is a true reflection of the candidate's (Adam A. Dowle) contribution to the paper.

### Candidate

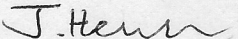
Name: Adam Dowle

Signature: 

Date: 1st June 2016

### Co-author

Name: James Hewitson

Signature: 

Date: 01.06.16

## Appendix 4A

Submitted paper 4 – Secretion of protective antigens by tissue-stage nematode larvae revealed by proteomic analysis and vaccination-induced sterile immunity.



# Secretion of Protective Antigens by Tissue-Stage Nematode Larvae Revealed by Proteomic Analysis and Vaccination-Induced Sterile Immunity

James P. Hewitson<sup>1</sup>, Al C. Ivens<sup>2</sup>, Yvonne Harcus<sup>1</sup>, Kara J. Filbey<sup>1</sup>, Henry J. McSorley<sup>1</sup>, Janice Murray<sup>1</sup>, Stephen Bridgett<sup>3</sup>, David Ashford<sup>4</sup>, Adam A. Dowle<sup>4</sup>, Rick M. Maizels<sup>1\*</sup>

**1** Institute of Immunology and Infection Research, Ashworth Laboratories, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom, **2** Centre for Immunity, Infection and Evolution, Ashworth Laboratories, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom, **3** Gene Pool, Ashworth Laboratories, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom, **4** Technology Facility, University of York, York, United Kingdom

## Abstract

Gastrointestinal nematode parasites infect over 1 billion humans, with little evidence for generation of sterilising immunity. These helminths are highly adapted to their mammalian host, following a developmental program through successive niches, while effectively down-modulating host immune responsiveness. Larvae of *Heligmosomoides polygyrus*, for example, encyst in the intestinal submucosa, before emerging as adult worms into the duodenal lumen. Adults release immunomodulatory excretory-secretory (ES) products, but mice immunised with adult *H. polygyrus* ES become fully immune to challenge infection. ES products of the intestinal wall 4th stage (L4) larvae are similarly important in host-parasite interactions, as they readily generate sterile immunity against infection, while released material from the egg stage is ineffective. Proteomic analyses of L4 ES identifies protective antigen targets as well as potential tissue-phase immunomodulatory molecules, using as comparators the adult ES proteome and a profile of *H. polygyrus* egg-released material. While 135 proteins are shared between L4 and adult ES, 72 are L4 ES-specific; L4-specific proteins correspond to those whose transcription is restricted to larval stages, while shared proteins are generally transcribed by all life cycle forms. Two protein families are more heavily represented in the L4 secretome, the Sushi domain, associated with complement regulation, and the ShK/SXC domain related to a toxin interfering with T cell signalling. Both adult and L4 ES contain extensive but distinct arrays of Venom allergen/Ancylostoma secreted protein-Like (VAL) members, with acetylcholinesterases (ACEs) and apyrase APY-3 particularly abundant in L4 ES. Serum antibodies from mice vaccinated with L4 and adult ES react strongly to the VAL-1 protein and to ACE-1, indicating that these two antigens represent major vaccine targets for this intestinal nematode. We have thus defined an extensive and novel repertoire of *H. polygyrus* proteins closely implicated in immune modulation and protective immunity.

**Citation:** Hewitson JP, Ivens AC, Harcus Y, Filbey KJ, McSorley HJ, et al. (2013) Secretion of Protective Antigens by Tissue-Stage Nematode Larvae Revealed by Proteomic Analysis and Vaccination-Induced Sterile Immunity. *PLoS Pathog* 9(8): e1003492. doi:10.1371/journal.ppat.1003492

**Editor:** Thomas B. Nutman, National Institutes of Health, United States of America

**Received:** January 25, 2013; **Accepted:** May 28, 2013; **Published:** August 15, 2013

**Copyright:** © 2013 Hewitson et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The work was supported by the Wellcome Trust (grant number 090281), and by an MRC CASE studentship with UCB Celltech. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: rick.maizels@ed.ac.uk

## Introduction

Gastro-intestinal nematode parasites are among the most prevalent pathogens in the world, afflicting over 1 billion people [1] and causing widespread disease in livestock [2]. Control through drug therapy is compromised by rapid reinfection [3], reflecting the lack of protective immunity generated during natural exposure. Helminth infection is also associated with a wide suite of immunological down-modulatory effects [4], which have evolved to promote parasite survival. Although it is possible to vaccinate animals against helminthiases, few protective antigens have been defined and immunity reduces rather than eliminates worm loads [1,5].

*Heligmosomoides polygyrus* is an ideal model species to study host-parasite interactions in gastro-intestinal helminth infection [6–8]. The parasite follows an entirely intestinal course of infection, entering orally, developing through larval stages before establishing as long-lived adults in the lumen of the small intestine,

releasing eggs that are transmitted through feces for onward transmission. Infected mice show multiple immuno-modulatory changes including expansion of regulatory T cells [9–12], B cells [13] and dendritic cells [14–16]. The immunoregulatory environment engendered by *H. polygyrus* extends to dampening bystander immune responses to allergens, autoantigens and intestinal antigens [4,17,18].

A critical point in infection is during the first 8–10 days, when incoming larvae invade the intestinal tissue and become encysted in the submucosa adjacent to the serosal membrane. Depending on the genetic background of the murine host, a localised immune reaction can envelop the larva, forming a macrophage-rich granuloma. In primary infection however, parasites can escape these inflammatory foci and successfully migrate to the lumen to continue their life cycle [19].

How parasites evade immune attack in the tissues has yet to be determined, but it is known that systemic immune suppression is associated not only with the long-lived adult stage but also the

## Author Summary

Intestinal helminth parasites are highly prevalent in humans and animals, and survive for long periods by deviating the host immune system. No vaccines are currently available to control these infections. Many helminths invade through barrier surfaces (such as the skin or the digestive tract) and develop through tissue larval stages before reaching their final niche such as the intestinal lumen. We studied the tissue larval stage of a mouse parasite, *Heligmosomoides polygyrus*, to test whether proteins released by this stage could elicit protective immunity, and found that they indeed constitute very effective vaccine antigens. Proteomic analysis to identify the individual proteins released by the larvae demonstrated that while many products are shared between tissue-dwelling larvae and adults occupying the intestinal lumen, larvae express higher levels of two gene families linked to immunomodulation, namely the Sushi protein family and the ShK toxin family. Antibody analysis of serum from vaccinated mice identified two major antigens recognised by the protective immune response as VAL-1 and ACE-1, which are respectively members of the venom allergen and acetylcholinesterase families. This work establishes that tissue larvae are the source of protective antigens for future vaccines, and highlights their production of two potentially immunomodulatory gene families.

larval stage in the intestinal wall [20]. Moreover, mice given curative anthelmintic treatment prior to adult maturation [21], or heavily irradiated infective larvae [22] develop protective immunity to challenge infection. Hence, as well as being a source of protective antigens, immature tissue-dwelling parasites actively contribute to immunological down-regulation, and we hypothesise that in the non-immune setting they are able to deflect or disable immune mechanisms within the gut tissue.

The ability of parasites to modulate immunity is likely to be dependent on the molecular components secreted into their mammalian host, reproduced by the “excretory-secretory” (ES) products collected from parasites maintained *in vitro* [23]. Global proteomic analyses of the parasite “secretome” are proving to be important and illuminating steps towards defining the host-parasite interaction at the molecular level [24]. Thus, the secretome of adult *H. polygyrus* has been found to contain more than 300 proteins, including many enzymes and homologues of host immune system genes, as well as novel proteins whose precise function has yet to be defined [25,26].

Parasite ES products have also proven to be effective vaccines in many settings [1,5], not least in the case of *H. polygyrus* in which adult HES immunization elicits sterilizing immunity in mice [27]. An attractive hypothesis for this protective effect is that vaccination generates neutralising antibodies that counter the immunomodulatory molecules secreted from the host. If such vaccines can also be directed against the immature stage of helminths, such as the tissue-phase larvae of *H. polygyrus*, then parasites may be eliminated before egg production and transmission can be attained.

We have therefore studied the released products of *H. polygyrus* tissue phase larvae, both to identify individual protein components, and to test their ability to induce immunity through vaccination. As detailed below, this investigation identifies a shift in secreted protein composition between the tissue and luminal phases, with expansion and contraction of different gene sets, in particular in the representation of the major multi-gene protein families associated with nematode infection. Most importantly, the

larval secretions are fully immunogenic, induce complete sterilising immunity, and have allowed us to identify key candidate protein antigens that may form the basis of a subunit vaccine for protective immunity.

## Results

### Secretome of 4th-stage *H. polygyrus* larvae

To identify the secretory products of the tissue-dwelling phase of *H. polygyrus*, individual L4 larvae were isolated from the submucosa of mice 5 days (120 hours) following oral infection with third-stage (L3) infective larvae, at a point soon after the third molt (90–96 hours) but prior to the fourth and final molt to the adult stage (144–166 hours) [28]. Using serum-free media for cultivation, in conditions previously reported for adult *H. polygyrus* [25], larvae were cultured for 3 days and supernatants collected, concentrated using centrifugal filters and analysed for recovered proteins. We also collected material released by *H. polygyrus* eggs (ERM; egg released material) harvested *in vitro* from adult worms collected from the lumen of infected mice to act as an outgroup as the egg stage has not previously been associated with immunity or immunomodulation.

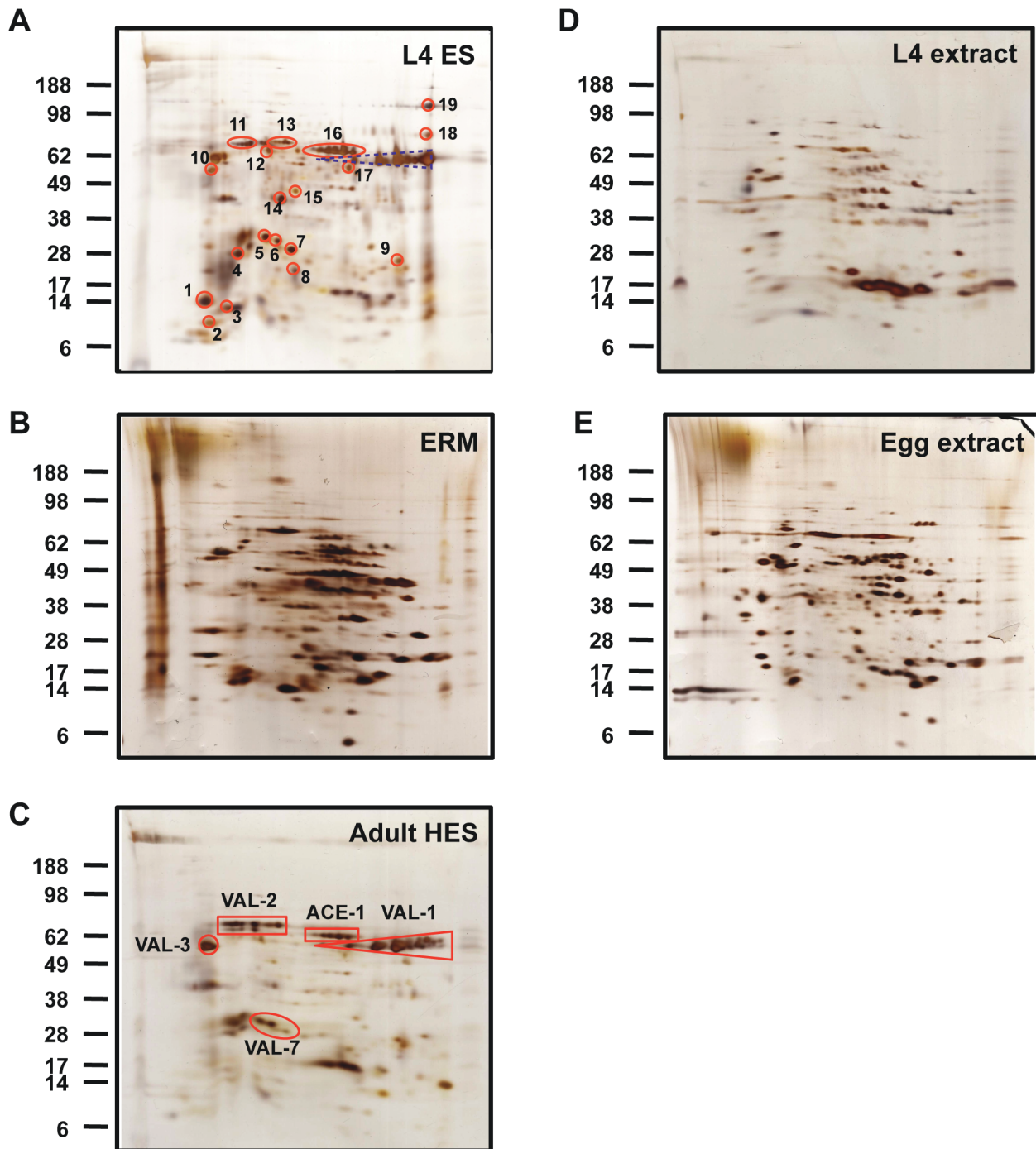
Proteins in ES or released material from the three life cycle stages were first compared by 2-dimensional SDS-PAGE and silver staining. As shown in **Fig. 1**, L4 ES (**Fig. 1 A**) was clearly very distinct from egg-released material (ERM) (**Fig. 1 B**), but bore a number of similarities with adult HES (**Fig. 1 C**). When L4 ES was compared to somatic extract from the same stage (**Fig. 1 D**), a number of proteins enriched or only visible in the secreted fraction were apparent. In contrast, most spots appeared to be shared between ERM (**Fig. 1 B**) and egg somatic extract (**Fig. 1 E**), suggesting in this case passive release rather than selective secretion. In the comparison between L4 and adult ES, a number of co-localising spots have previously been identified in HES including members of the Venom allergen/*Ancylostoma* secreted protein-Like (VAL) family and acetylcholinesterases (ACE) [25].

MS analysis of individual L4 ES spots, detailed in **Table 1**, confirmed the presence of two forms of ACE (ACE-1 and -2), as well as several VAL proteins, particularly variants of VAL-7, the lysozymes LYS-1 and LYS-2, apyrase APY-3, and the previously described Th2-skewing protein calreticulin [29]. In addition, proteins with no known homologs in other species were present, including variants of a Novel Secreted Protein-3 (NSP-3) found in adult HES [25] and a newly described Larval Secreted Protein (LSP) not detected in HES.

### Excretory-secretory products from 4th-stage larvae, but not eggs, induce sterilising immunity to challenge infection with *H. polygyrus*

Our first objective was to test whether L4 ES antigens were able and sufficient to generate protective immunity in mice that are normally susceptible to primary infection with *H. polygyrus*. We have previously shown that vaccination with adult HES in alum adjuvant generates sterile immunity against challenge [27]. We therefore compared immunization with L4 ES, HES and egg-derived ERM.

Immunity in vaccinated mice was measured both by fecal egg output and worm burden at autopsy. Egg burdens measured between 14 and 28 days of infection rapidly approached zero in mice vaccinated with either L4 ES or HES (**Fig. 2 A**), while adult worm burdens at day 28 showed that both conferred sterile immunity on nearly all mice (**Fig. 2 B**). In contrast, ERM immunization did not affect adult worm burdens or parasite fecundity (**Fig. 2 A, B**). HES-immunized mice produced high levels



**Figure 1. 2-D gel electrophoresis of ES released material and somatic extracts from different stages of *H. polygyrus*, visualised by silver staining.** A. L4 ES. Identities of numbered spots are given in Table 1. B. Egg released material (ERM). C. Adult HES. Solid red boxes correspond to the indicated protein products. D. L4 somatic extract. E. Egg somatic extract. Positions of molecular weight markers (kDa) are indicated. doi:10.1371/journal.ppat.1003492.g001

of anti-HES IgG1 antibodies (Fig. 2 C), the isotype most important in antibody-mediated immunity following secondary infection [30]. Notably, anti-HES IgG1 titers were similar in ERM- and HES-immunized mice, showing that the failure of ERM immunization is not due to any intrinsic lack of immunogenicity (Fig. 2 C). Instead, this emphasised the importance of identifying specific molecular targets following protective immunisation, a process that required we first gain a deeper understanding of the individual protein components in each ES preparation.

#### LC-MS identification and quantification of L4 and egg secreted proteins

Taking advantage of an expanded *H. polygyrus* nucleotide database based on >1 million cDNA sequences from 5 different life-cycle stages of *H. polygyrus* (L3, day 3, day 5, adult and eggs) including each of those from which the secreted proteins were collected, we subjected L4 ES and ERM to LC-MS/MS analysis and compared them to adult HES [25]. This analysis identified



**Table 1.** L4 ES protein spot identities.

Spot	Identity	Isotig and Isogroup	Peptides	Score
1	<b>NSP-3.3</b>	Hp_I01045_IG00068 Hp_I01047_IG00068	6	464
2	<b>MPP-like</b>	Hp_I10549_IG03359 Hp_I10550_IG03359	1	98
3	<b>NSP-3.2</b>	Hp_I17967_IG09911	3	241
	<b>NSP-42</b>	Hp_I02051_IG00167	1	69
4	<b>Lysozyme-1</b>	Hp_I08665_IG02417	5	522
5	<b>VAL-7.4</b>	Hp_I01449_IG00104	6	662
	<b>VAL-7.3</b>	Hp_I01450_IG00104	6	592
6	<b>VAL-7.5</b>	Hp_I01454_IG00104	5	394
	<b>VAL-7.3</b>	Hp_I01450_IG00104	4	293
	<b>VAL-7.2</b>	Hp_I01453_IG00104	4	265
7	<b>VAL-7.1</b>	Hp_I01451_IG00104 Hp_I01452_IG00104	7	551
8	<b>Sushi-like</b>	Hp_I08843_IG02506	2	246
	<b>NSP-62</b>	Hp_I30075_IG22019	1	134
9	<b>Lysozyme-2</b>	Hp_I05758_IG01053	3	242
	<b>Deoxyribonuclease II</b>	Hp_I15874_IG07818	2	181
	<b>VAL-25</b>	Hp_I10899_IG03534 Hp_I19276_IG11220	1	78
10	<b>Calreticulin</b>	Hp_C00231_IG00001	3	213
11	<b>ACE-2</b>	Hp_I04629_IG00719	7	470
	<b>VAL-2.3</b>	Hp_I07952_IG02060	1	77
12	<b>VAL-9</b>	Hp_I14221_IG06165	8	604
13	<b>VAL-2.1/2.2/2.3</b>	Hp_I07952_IG02060 Hp_I07953_IG02060 Hp_I15068_IG07012	3	214
14	<b>APY-3</b>	Hp_I06737_IG01430	5	622
15	<b>Astacin protease</b>	Hp_I13832_IG05776	2	144
16	<b>ACE-1</b>	Hp_I12803_IG04747	6	618
17	<b>VAL-13</b>	Hp_I13967_IG05911	4	329
18	<b>LSP-1</b>	Hp_I03144_IG00345	8	664
19	<b>Kunitz inhibitor</b>	Hp_I12299_IG04234	1	165

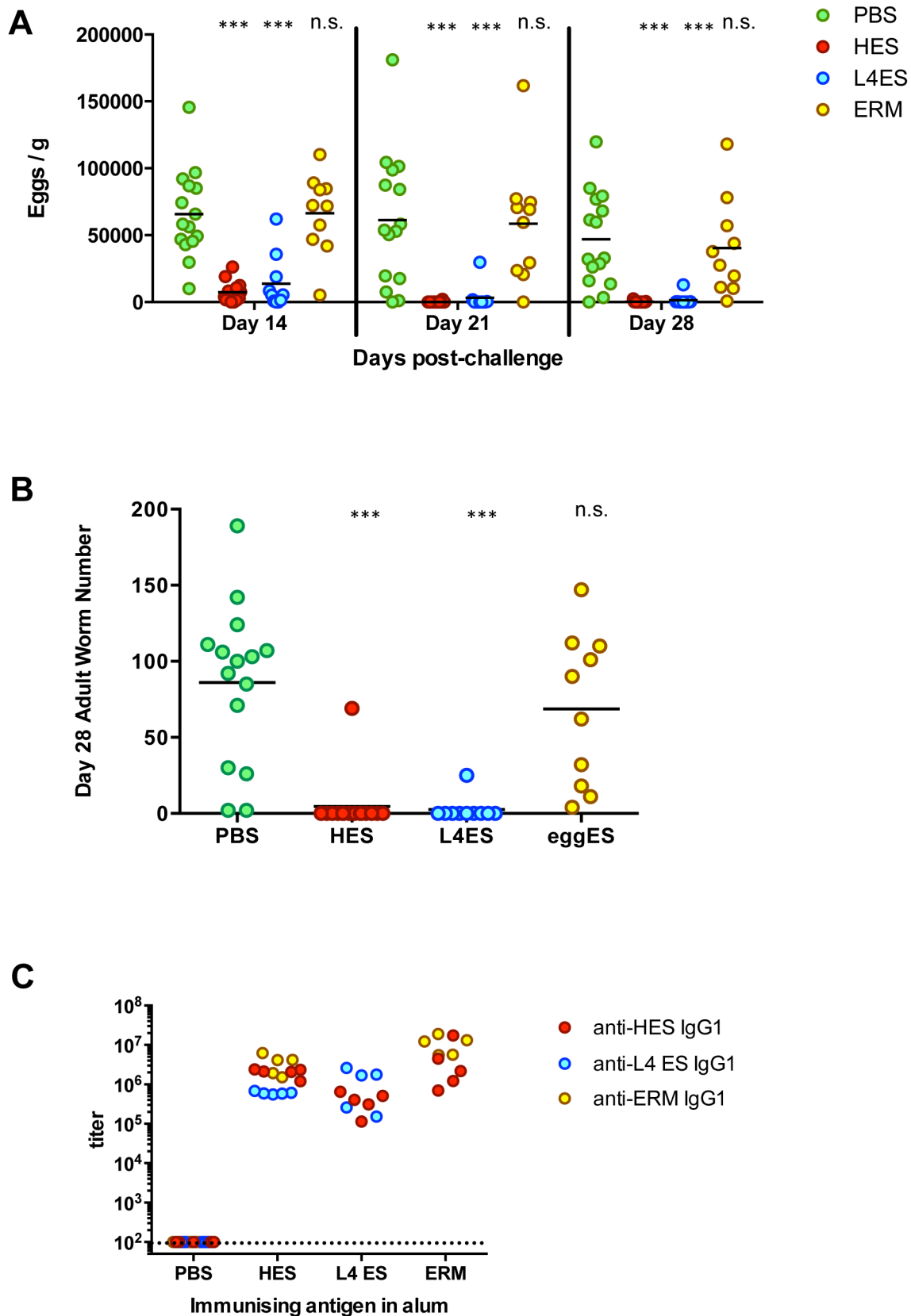
Spot numbers refer to those in **Fig. 1 A**. One identity is shown where the same peptides match multiple isotigs (i.e. variants). Additional identities are shown where distinct peptides match either variants of the same protein or different proteins that co-localise within the same gel area. Number of unique peptides identifying each protein is shown, as is the mascot score.

doi:10.1371/journal.ppat.1003492.t001

214 proteins present in L4 ES, and 209 in ERM, in comparison to the secretion of 364 in adult HES. The number of proteins shared between adult and L4 (63.1% = 135/214 L4 ES proteins), and adult and ERM (59.3% = 124/209 ERM proteins) were broadly similar (**Fig. 3 A**). Both L4 (33.6% = 72/214) and ERM (37.3% = 78/209) contained a similar proportion of proteins detected only in that stage. Very few proteins (7) were shared between L4 ES and ERM but were not detected in HES, while a core of 60 ES proteins was shared between the three different lifecycle stages. Additionally, the proportion of novel proteins (i.e. those proteins lacking homologs in other species) that were stage-specific was enriched compared to the total secretions (**Fig. 3 B**; HES 44.3% = 27/61, L4 ES 42.5% = 19/45, ERM 60.5% = 23/38). The vast majority of L4 ES proteins had a predicted signal peptide (92.5% = 198/214), greater than that seen for adult HES

(79.7% = 290/364), whereas ERM was much lower (59.8% = 125/209), again suggesting that the latter contains constituents that are not actively secreted.

Two parallel approaches were taken to quantify the relative levels of proteins present in the different ES material and determine selective expression, protein abundance and gene expression profiling. Firstly, protein abundance was estimated by emPAI (exponentially modified Protein Abundance Index [31]), which provides an approximate protein ranking based on the number of observed peptides relative to the known full-length sequence. Because not all peptides are equally amenable to MS detection, and not all sequences are full length, this method is not infallible but provides a benchmark for further analysis and allows direct comparison of the relative abundance of an individual protein across different life-cycle stages.



**Figure 2. Protective immunity following L4 ES or HES, but not ERM, immunization of C57BL/6 mice.** **A.** Fecal egg burdens in mice at days 14, 21 and 28 post-challenge. **B.** Adult worm recoveries at day 28 post-challenge. **C.** Anti-ES IgG1 titers in immunized mice prior to challenge. Anti-HES (red), anti-L4 ES (blue) and anti-ERM (yellow) titers are shown following HES, L4 ES or ERM immunisation. Data in (A–B) are pooled from two

individual experiments with 5–10 mice per group, and data in (C) are representative of two individual experiments. Statistical significance refers to ANOVA comparisons with PBS/alum group (\*\*\*)  $p < 0.001$ ; n.s. = non-significant). doi:10.1371/journal.ppat.1003492.g002

Heat maps ranking the top 50 highest emPAI-scoring L4 ES, HES and ERM are shown in **Figure 4 A–C** respectively, while complete alphabetical lists are presented in **Tables S1, S2, S3**. On the criterion of emPAI score, the most abundant L4 ES constituents were two members of the Shk/SXC gene family (PF01549) containing a conserved six-cysteine motif (SXC-like 1 and 2) and three isoforms of VAL-7 corresponding to 2D gel spots (**Fig. 1**). Most of these were also prominently expressed by the adult stage, although SXC-like proteins were highly upregulated in L4 ES. Members of the novel NSP-3 family were also expanded in the L4 stage. Most notably, a number of Sushi-domain containing proteins were relatively highly represented in L4 ES but absent from adult ES.

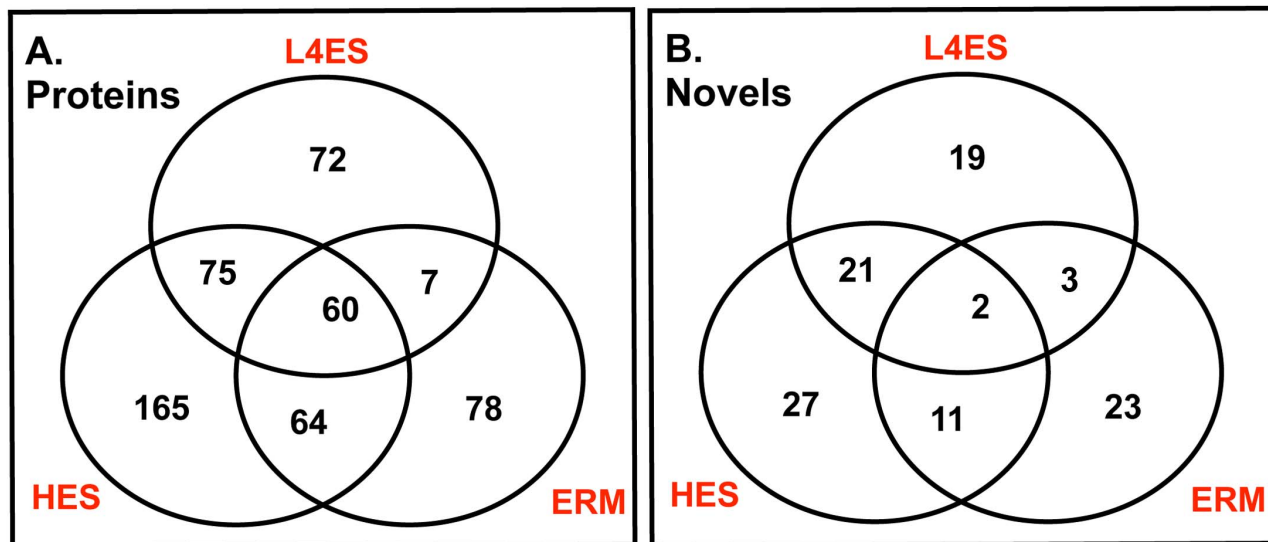
To form a broader picture of the functional pattern of each ES compartment, we also analysed the constituent proteins for domains corresponding to defined protein families in Pfam (**Fig. 4 D and Tables S4, S5**). Statistical comparison with Pfam domains detected in the total CDS of the *H. polygyrus* transcriptome (Harcus et al, manuscript in preparation) confirmed the selective release of protein families secreted by all parasite stages (e.g. PF00188 CAP indicating widespread release of VAL proteins). Furthermore, it also revealed regulation in the secretion of key gene families through parasite development (e.g. expansion of PF00084 Sushi and PF01549 ShK/SXC in L4 ES, and PF00635 motile\_sperm in ERM). Two-way clustering of abundant proteins from the three ES preparations again shows close association in proteins released by L4 and adult worms, but not ERM, confirming the global similarity between larval and adult released material (**Fig. 4E**).

It is interesting to note that visual inspection of the 2D SDS-PAGE comparisons (**Fig. 1**) indicates a closer similarity between L4 and adult HES than emerges from the full LC-MS proteomic analysis, while ERM appears distinctly different. This reflects the relative abundance of the proteins shared by the L4 and adult

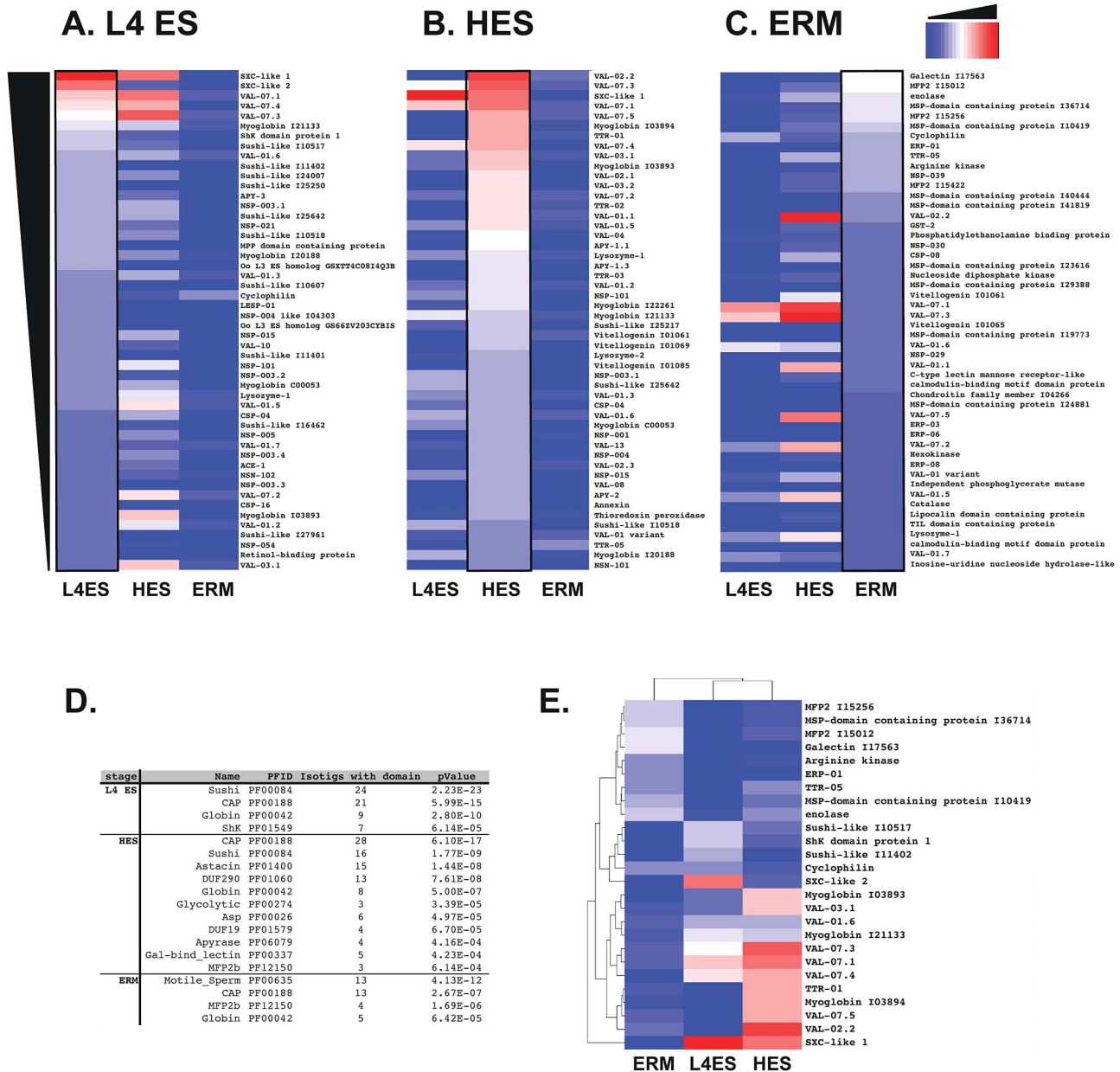
stages. Moreover, a significant positive correlation was observed between the emPAI values of shared L4 ES and HES proteins ( $p < 0.0001$ ; Spearman  $r^2 = 0.3478$ ) but not between those of shared ERM and HES proteins (n.s.; Spearman  $r^2 = 0.1660$ ; **Fig. S1 A, B**); hence although certain proteins are present in both HES and ERM, their abundance may be very different. This is well illustrated by the heat map of emPAI values of the 60 proteins shared between all three ES preparations showing clear differences in expression level between the stages (**Fig. S1 C**). The identities and relative abundances of proteins shared between HES and L4 ES (which both include protective antigens), as well as HES and ERM (in which the latter is non-protective), are shown in **Fig. S1 D and E**.

### Transcriptomic analysis of ES protein-encoding genes

An independent method of determining the stage-specific expression of ES proteins was based on an extensive transcriptomic database, using a normalised read count (RPKM; reads per kb per million mapped reads [32]) to create a heat map corresponding to relative gene expression across the five different lifecycle stages (**Fig. 5 A–C**). This revealed most L4 and adult ES proteins are encoded by genes that are relatively quiescent in the free-living L3 stage, and transcription is either initiated or upregulated following host entry. Generally, gene expression analysis supports the proteomic data; expression of L4-specific ES proteins is mostly at its highest in day 3 or 5 larvae, before downregulation in the adult (**Fig. 5 A**), expression of shared L4 ES and HES proteins is maintained in day 3, 5 and adult worms (**Fig. 5 B**), and expression of HES-specific proteins is highest in adult worms (**Fig. S2 A**). The expression of the “core 60” ES proteins common to L4 ES, HES and ERM splits into two main clusters, those that are expressed by all 5 stages and those mainly expressed by d3, d5 and adult worms (**Fig. 5 C**). A small cluster of ubiquitously expressed genes was also seen in ERM-specific



**Figure 3. Shared and stage-specific ES proteins.** **A.** Venn diagram of total secretome proteins from three stages, enumerating specific and shared identities. **B.** Venn diagram of subset of total proteins showing no database similarity to known proteins in other species. Such proteins in Adult HES (irrespective of presence in other stages) are designated as previously described [25] as Novel Secreted Proteins (NSPs), while those found in L4 ES but not the adult are Larval Secreted Proteins (LSPs). doi:10.1371/journal.ppat.1003492.g003



**Figure 4. Comparison of protein abundance (emPAI) in each ES product.** **A.** The 50 most abundant proteins in L4 ES, ranked by emPAI score (left column) with corresponding emPA scores for HES (centre) and ERM (right column). Scores are coloured on a log<sub>2</sub> scale with red maximum and blue minimum. **B.** The 50 most abundant proteins in HES ranked as above (centre column) with corresponding scores for L4 ES and ERM. **C.** The 50 most abundant proteins in ERM ranked as above (right column) with corresponding scores for L4 ES (left column) and HES (centre). **D.** Protein family (Pfam) enrichment within each ES, ranked by statistical significance (cut-off of p<0.001) compared to Pfam in total transcriptomic assembly. "Isotigs with domain" indicates number of times domain present in indicated ES. Full analyses are provided in **Table S5**. **E.** Two-way clustering of the 10 most abundant proteins in the 3 ES preparations. doi:10.1371/journal.ppat.1003492.g004

proteins, but surprisingly the majority of ERM-specific proteins were not actually expressed by the egg, and instead must be derived from the adult worm where expression is greatest (**Fig. S2 B**). This includes sperm-derived proteins present in ERM, such as the MSP-domain proteins (13x) and MFP2 (4x), as well as the egg sac vitellogenins (5 variants).

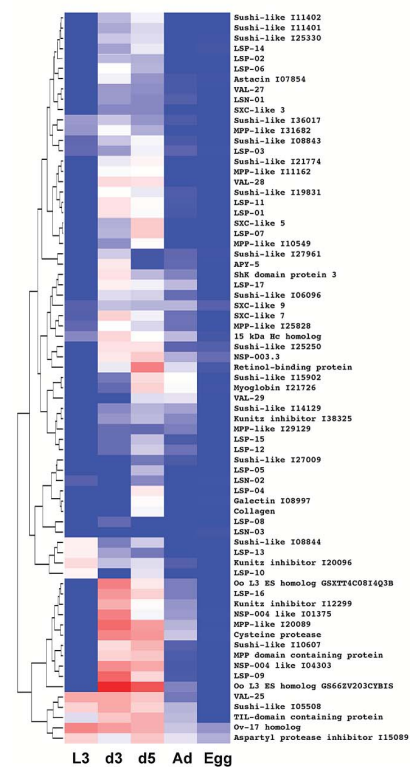
Comparison of protein abundance (emPAI) with gene expression (RPKM) for all L4 ES proteins showed that protein levels positively correlate with gene expression of day 5, but surprisingly not day 3, larvae (**Fig. S3 A-E**). This reflects the major changes in

gene expression either side of the L3 to L4 molt, with new transcription of many new genes being initiated (Harcus Y. *et al*, manuscript in preparation). Notably, the significant correlation between protein emPAI and mRNA RPKM within the L4 stage lends further validation to these parameters as quantitative measures of expression in the nematode worm.

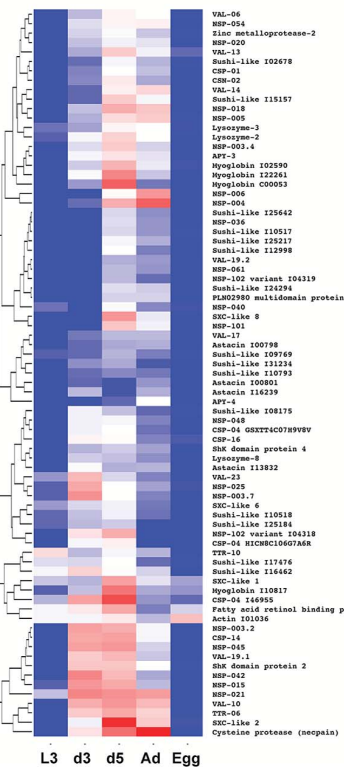
#### Predominant and enriched protein families in L4 ES

The most highly represented proteins in both L4 and adult ES are VAL (Venom allergen/Ancylostoma secreted protein-Like)

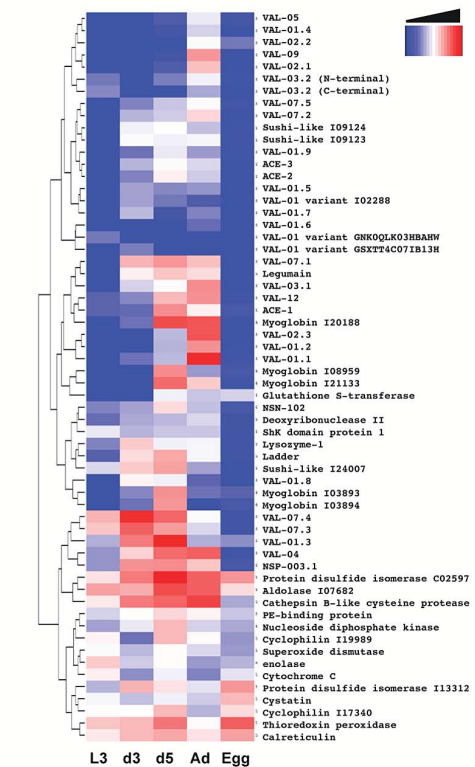
## A. L4 ES alone



## B. L4 ES / HES



## C. L4 ES / HES / ERM



**Figure 5. Gene expression (RPKM) heat-maps of stage-specific and shared ES proteins.** A. L4 ES stage-specific proteins, clustered by gene expression profiles in a transcriptomic dataset based on 5 life cycle stages (infective L3, d3 post-infection L3, d5 post-infection L4, Adult, and Egg). RPKM = Reads Per Kilobase Mapped. Scores are coloured on a log<sub>2</sub> scale with red maximum and blue minimum. B. As above for proteins shared between L4 ES and HES. C. As above for proteins shared between L4 ES, HES and ERM. The corresponding heat-maps for adult HES stage-specific proteins, and for ERM-specific proteins, are presented in **Suppl. Fig. 2 A and B** respectively. doi:10.1371/journal.ppat.1003492.g005

family members; these are part of a larger CAP protein superfamily (Pfam PF00188), which includes mammalian sperm-coating protein (SCP) that lends its name to the canonical domain structure. The CAP superfamily shows extensive diversity in most species, with over 30 genes in mice and humans [33]; in this respect, *H. polygyrus* is not exceptional in expressing over 20 different VAL family members (numbered in descending order of abundance in adult HES) although their representation differs between adult and L4 ES. Generally, L4 ES contains lower levels of most VAL proteins compared to HES, with just trace amounts of VAL-2 present, and there is no abundant L4 ES VAL protein not previously described in adult HES. Only the low abundance VAL proteins VAL-23, 25, 27, 28 and 29, as well as the moderately abundant VAL-10, are enriched in L4 ES compared to HES, a pattern replicated at the transcript level for the low abundance VAL proteins (**Fig. 6 and Table S1**).

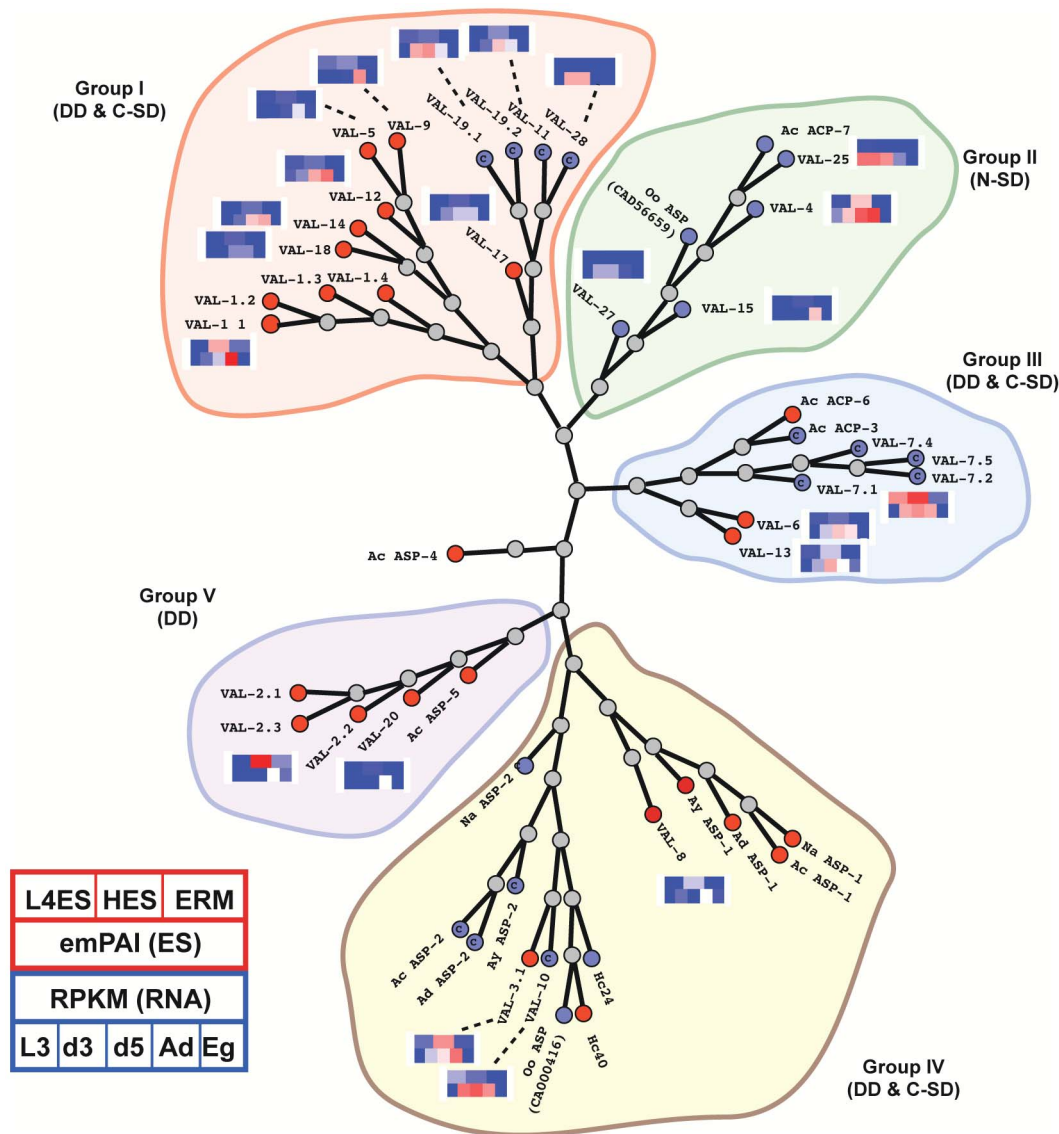
Using specific mAb, we confirmed by ELISA that L4 ES contains little VAL-2, but that other abundant VAL proteins (VAL-1, 3 and 4) are present at levels similar to adult HES (**Fig. S4 A–D**). In contrast, little VAL-1 was detected in ERM (**Fig. S4 A**) despite LC-MS analysis showing the presence of several variants (**Table S3**). This is suggestive of partial degradation, consistent with the 2-D gel of ERM lacking the characteristic stretch of VAL-1 isoforms readily identifiable in both HES and L4 ES (**Fig. 1 A–C** and [25]).

As has been noted in studies with other helminth species [34,35], the VAL gene family shows extensive evolutionary diversity within *H. polygyrus*, in both sequence and domain structure (**Fig. 6**). Additionally, proteins expressed in a stage-regulated manner do not form a clear structural subset within this family as within Group 1 alone (**Fig. 6**) VAL-1 is preferentially expressed by the adult, VAL-28 is restricted to the d3/d5 larval transcriptome, whereas others (such as VAL-12) show a broader pattern of expression. Such contrasting profiles may reflect fine adaptation of different family members to specialised functional roles. Notably, comparison with the closely related *Ancylostoma* and *Necator* hookworm species reveals an intricate pattern of gene expansion and contraction: Group 1 *H. polygyrus* VAL proteins are clearly separated from the human and dog hookworm ASP-1-like VAL-8, ASP-2-like VAL-3 and ASP-5-like VAL-2. Hence, adaptation to the murine host may have resulted in an accelerated radiation of VAL family members.

While the VAL family is less prominent in L4 ES than HES, other protein families are distinguished by maximal representation in the L4 stage. These include two conserved families, the Sushi-like proteins and the ShK/SXC-like proteins, as well as a newly defined set of Novel Secreted Protein-3 (NSP-3) variants.

Sushi domains (PF00084) are 60-amino acid consensus sequences with 4 conserved cysteine residues, and are prevalent in mammalian proteins that regulate complement activation, hence the alternative domain name of complement control protein





**Figure 6. Phylogeny and expression of VAL family members in *H. polygyrus*.** Phylogenetic tree of *H. polygyrus* secreted VAL proteins and those from selected other intestinal nematodes (Na = *Necator americanus*; Ac = *Ancylostoma caninum*; Ad = *Ancylostoma duodenale*; Ay = *Ancylostoma ceylanicum*; Oo = *Ostertagia ostertagi*; Hc = *Haemonchus contortus*). Blue nodes represent single SCP domain VAL/ASP proteins whereas red nodes indicate double SCP domain proteins; blue nodes are indicated as more similar to N-terminal (N) or C-terminal domains (C) as appropriate. Protein (emPAI, top in same order as Fig. 4) and gene (RPKM, bottom, in same order as Fig. 5) expression profiles are indicated for *Hp* VAL proteins. Expression profiles are for most abundant variant where more than one is present. *H. polygyrus* VAL-21 to 24 and VAL-29 have been omitted as their sequences were either incomplete or did not unambiguously align with a single branch of the tree. On the basis of this phylogenetic analysis, 5 groups of VAL proteins are indicated. Note that this does not conform with a 3-group classification recently proposed [69] as we found the Groups 1 and 3 defined by these authors to be polyphyletic. doi:10.1371/journal.ppat.1003492.g006

(cd0003; [36]). Of note, Sushi domains in complement regulatory proteins generally occur in repeated units (e.g. complement receptor 1A, 30 repeats; factor H, 20 repeats), which is also evident in an *Ascaris* complement factor H homolog (15 repeats; ADY39830.1) and similar proteins in *Loa loa* (12 repeats; XP\_003143211.1) and *C. elegans* (11–16 repeats; Ce F36H2.3 isoforms A-G). Some 34 L4 ES proteins contained sequences related to the Sushi domains (pfam 00084), which we have collectively termed Sushi-like proteins as detailed in Fig. S5. Of these, 18 are also present in HES albeit at lower levels, together with an additional 5 not present in L4 ES. ERM contained only two Sushi-like proteins at low levels. The majority of *H. polygyrus*

sushi-like proteins however, have only 1 or 2 sushi domains (Fig. S5 A) more similar to those of cytokine receptors such as the alpha subunits of the IL-2 (2 domains) and IL-15 (1 domain) receptors [37].

ShK/SXC proteins have a 36 amino acid domain with 6 conserved cysteine residues, and are widely expressed by parasitic and non-parasitic nematodes [38,39]. L4 ES contains 15 proteins with ShK domains, with an additional one detected at low levels in HES. Several of these domains (7/16) were fused to other sequence motifs; 3 astacin metalloproteases and 4 homologs of hypothetical (non-annotated) *C. elegans* proteins. In the latter cases, proteins contained either 2 or 4 tandemly repeated ShK domains

(**Fig. S6 A**), in a pattern very similar to that observed in other parasitic nematodes, such as *T. canis* mucins [39]. More unusually, many ShK proteins (9/16) in *H. polygyrus* ES are found as short single-domain proteins (full-length 63–67 aa, mature protein 41–49 aa) consisting of no more than a signal sequence and the 6-cysteine domain (**Fig. S6 B**). As this is reminiscent of *Ostertagia ostertagi* SXC-1 (accession number CAC17797), we have termed these proteins SXC-like. The parasite appears primed to produce this group of proteins even before infection of the mammalian host as gene expression is detected for 5/9 in the L3 stage, and peaks in day 3–5 larvae.

L4 ES contains relatively high levels of several enzymes including apyrases (particularly APY-3, but also low levels of a new L4-specific APY-5), lysozymes and the three acetylcholinesterase proteins (ACE-1, 2 and 3) previously identified in adult secretions [25]. ACE levels in L4 ES were comparable to those in HES, as determined by emPAI, an expression pattern confirmed by RPKM transcript analysis, which revealed low or absent expression in L3 and eggs, and peak expression by day 5 L4 parasites, maintained in adults (**Tables S1, S2, S3**). In contrast, significant products which are preferentially secreted by the luminal adults, and which have previously been reported to be secreted by parasitic nematodes, include proteases such as astacins and zinc metalloproteases [40–42], cathepsins [43,44] and trypsin-like and other serine proteases [45–47], serpin [48] and Kunitz protease inhibitors [49], and predicted chitinases and chitin-binding proteins [50].

In common with most nematode transcriptomic analyses, a large proportion of expressed protein genes either lack identifiable homology to proteins of other species, or only match hypothetical *C. elegans* proteins. These included over 80 distinct sequences incorporating a canonical signal peptide and found in either or both L4 ES and HES (**Table S1**). Those unique to L4 ES were designated Larval Secreted Proteins (LSPs) while those present in HES adhered to the earlier designation of Novel Secreted Proteins (NSPs). Some 17 distinct LSPs were identified, while a further 17 L4 ES proteins were among a total of 50 NSPs present in the adult HES. The L4-specific proteins LSP-1 and LSP-2 were similar to the adult-specific ES constituents NSP-28 and NSP-53, indicating micro-adaptation of sequences to the different niches. Similarly, seven closely related variants of NSP-3 were also identified, one of which was found only in the L4 ES and two other only in HES. A further grouping was formed by three L4 ES NSP-4-like proteins, with homology to adult secreted NSP-4, 6, 10, 12, 18 and 44.

Adult HES contains an as yet unidentified TGF- $\beta$ -like molecule that is able to directly induce Foxp3<sup>+</sup> regulatory T cells in the presence of a mitogenic stimulus [51]. Using a TGF- $\beta$  reporter cell line, L4 ES was also shown to possess TGF- $\beta$  activity (59.4 $\pm$ 14% of the level present in adult HES over multiple batches), whereas this was undetectable in ERM (**data not shown**). Scrutiny of the mass spectrometric data did not reveal any peptides matching either mammalian TGF $\beta$ , or members of the TGF- $\beta$  superfamily previously identified within the genome and transcriptome of *H. polygyrus* [52]. Hence, it is possible that the ability of HES to ligate the mammalian TGF- $\beta$  receptor is mediated by one of the novel secreted proteins described above.

### Protective and non-protective antigens following ES immunisation

We noted earlier that ERM induced similarly high titers of anti-HES IgG1 antibodies to those generated by protective L4 ES or HES immunisation (**Fig. 2 C**) while failing to induce immunity, suggesting that non-protective but cross-reactive epitopes were well represented in these antigen preparations. We have previously

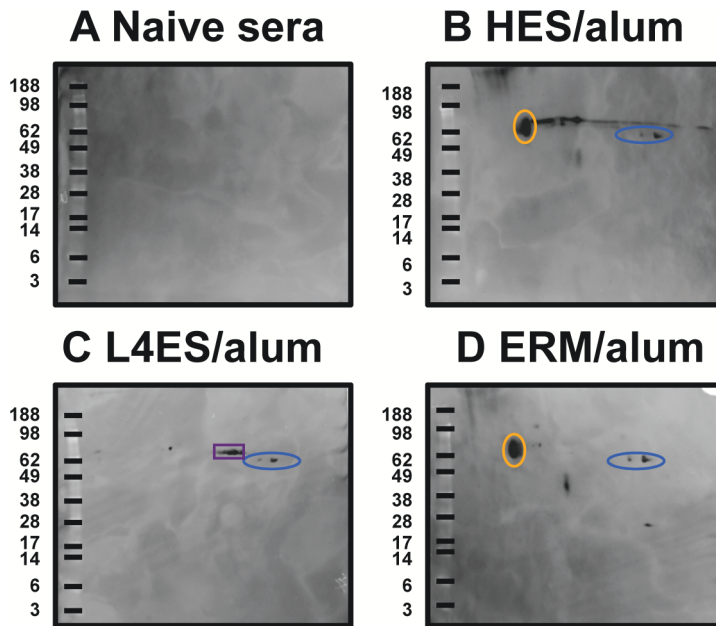
identified two non-protective immunodominant carbohydrate epitopes in *H. polygyrus* with monoclonal antibodies, namely Glycan A, an O-linked sugar attached to several VAL proteins (including VAL-1), and glycan B, a proteoglycan-like structure abundant in the adult worm soma [27]. Using the same monoclonal antibodies, we found that whilst L4 ES contained markedly less of both Glycans A and B than adult HES, confirming previous work [27], ERM contained substantially more glycan B than the other preparations (**Fig. S4 E–F**). Consistent with this, Western blot analysis revealed serum IgG1 from HES and ERM, but not L4 ES, immunised animals showed strong reactivity to a 65 kDa spot representative of glycan B (**Fig. 7 A–D**). All three preparations induced IgG1 reactivity against VAL-1 spots, suggesting class-switched antibodies binding to glycan A. Additionally L4 ES immunisation resulted in Western reactivity to spots corresponding to ACE-1 (**Fig. 7 C**).

To identify potential protective protein antigens, we analysed immunised sera through an immunoprecipitation protocol which captures responses to conformation-dependent protein epitopes [27]. Immunoprecipitation of biotin-labelled ES with sera from immunised mice revealed a more extensive profile of targeted antigens, with sera from L4 ES and HES immunised mice binding to a range of proteins in both L4 ES and HES (**Fig. 8 A**). In contrast, ERM immunisation did not elicit such antibodies, providing further evidence that the high anti-HES IgG1 titers generated bind almost exclusively to non-protective cross-reactive glycans. On 2-D analysis, it was found that while HES immunisation generated strong antibody recognition of VAL-1, 2 and 3, as well as ACE-1, L4 ES only induced antibodies against VAL-1 and ACE-1, and these appeared responsible for the vast majority of the observed immunological cross-reactivity between the two preparations (**Fig. 8 B–E**). These data argue that the minimal requirement for protective immunity is recognition of a single VAL protein (VAL-1) and ACE-1.

### Discussion

The larval and adult stages of parasitic nematodes often colonise different physiological niches. Distinct immune pressures exerted in each environment will likely be reflected in the preferential expression of key genes and protein families capable of modulating host immunity. In this regard, the tissue-phase ES material of the intestinal nematode *H. polygyrus* was characterised by increased expression of proteins associated with the regulation of complement (sushi-like) and T cell activation (SXC-like), as well novel proteins of unknown function (e.g. NSP-3 family). However, other proteins are common to both L4 ES and HES, and it is two of these (VAL-1 and ACE-1) that are major antigenic targets in mice rendered fully immune to challenge infection by vaccination with either of these preparations.

We originally hypothesised that protective immunity induced by ES immunisation would function by neutralising parasite immunomodulatory molecules. Although no common mechanistic function has yet been attributed to VAL proteins or the CAP superfamily, it is notable that they are most prominent at interfaces between organisms (virally-infected plants, insect venoms, helminth parasite secretions and the mammalian male gamete) and in some instances are associated with immunological down-modulation [53,54]. Alternatively, immunisation may inhibit proteins (such as ACE-1) with key but uncharacterised roles in parasite metabolism, migration, and coordination. It is also possible that vaccine-elicited antibodies protect through their Fc domains by recruiting innate immune cells to kill the invading larvae, hypotheses we are currently testing. Potential immunomodulatory



**Figure 7. ERM immunization elicits class-switched anti-glycan B antibodies.** A–D Western blot analysis of IgG1 antibodies from (A) naïve, or pre-challenge (B) HES, (C) L4 ES and (D) ERM immunised mice. Spots corresponding to VAL-1/glycan A (blue), glycan B (orange) and ACE-1 (purple) are shown. Molecular weight markers (kDa) are indicated.  
doi:10.1371/journal.ppat.1003492.g007

roles can also be ascribed to the Sushi and SXC-like proteins. Homologs of sushi proteins from vaccinia viruses are able to inhibit both the classical and alternative pathways of complement [55], whereas a SXC protein homolog from the sea anemone *Stichodactyla helianthus* was shown to block mammalian potassium channels, and hence was named ShK [56]. ShK proteins and related sequences can inhibit calcium-dependent lymphocyte activation [57], suggesting a direct immunomodulatory role for ShK homologues from nematode parasites.

Helminth pathogens regulate their physiological and immunological environment through a spectrum of released products that are being characterised by new high throughput technologies. Our study on the mouse model *H. polygyrus* directly complements many datasets with major human pathogens such as *Brugia malayi* [58–62] and the hookworm species [63,64], as well as economically important trichostrongylid nematodes [65–67]. Each of these parasites has evolved specialised strategies adapted to particular host species, and niches within those species, that will be mirrored in the repertoire of secreted products, and our study provides many examples of protein families which are differentially expressed, expanded or regulated during the course of a parasite life cycle. Importantly, in view of the lack of anti-helminth vaccines for human use, we have also tested ES products from larval and adult stages for their ability to induce sterilising immunity and in combination with the proteomic database, have now identified potential vaccine antigens for future appraisal. Taken together, these analyses have defined an extensive and novel repertoire of protein candidates from *H. polygyrus* that can be taken forward for functional and immunological testing in a tractable and naturally adapted host-parasite model system.

## Materials and Methods

### Ethics statement

All animal protocols adhered to the guidelines of the UK Home Office, complied with the Animals (Scientific Procedures) Act

1986, were approved by the University of Edinburgh Ethical Review Committee, and were performed under the authority of the UK Home Office Project Licence number 60/4105.

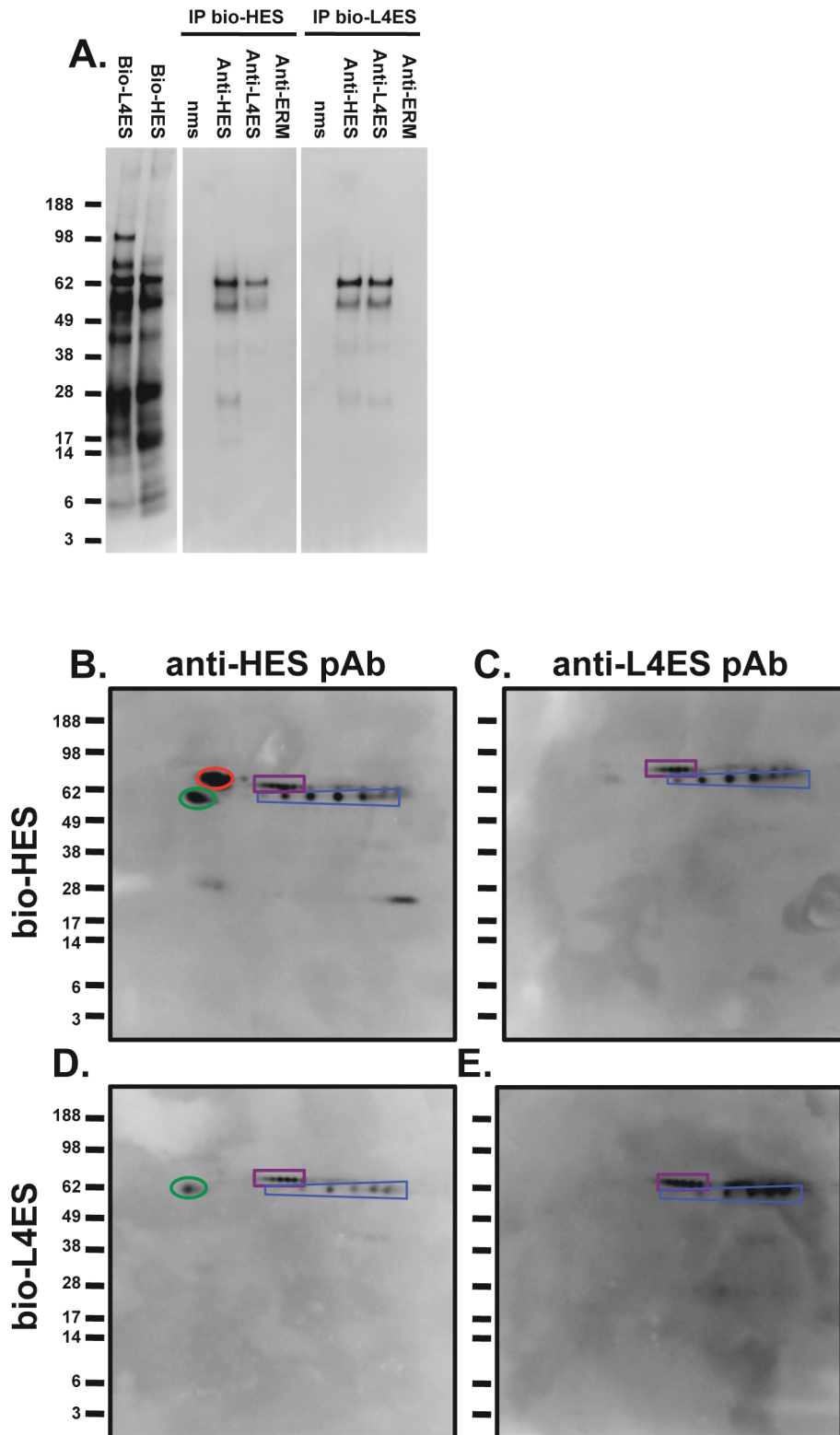
### Parasite maintenance and preparation of ES material

CBAXC57BL/6F1 mice were infected by gavage with 500 L3 stage larvae of *H. polygyrus bakeri* (originally provided by Professor J M Behnke, University of Nottingham, UK). Day 5 4<sup>th</sup> stage larvae and day 14 adults were washed, cultured in serum free media and the ES concentrated as before [25,27]. For egg released material (ERM), *in vitro* released eggs from adult worms were washed and cultured for 3 days in serum free media before concentration and diafiltration into PBS using 5000 MWCO spin columns (Vivaspin). L4 produced  $10 \pm 2$  ng ES protein/larva/day, compared to adult worm production of HES at  $18 \pm 2$  ng/adult/day. Eggs cultivated for 3 days *in vitro* released  $56 \pm 15$  pg protein/egg/day. Somatic extracts of L4 and egg were prepared with a Qiagen tissuelyser.

### 2-D gel electrophoresis, LC-MS/MS and bioinformatics

L4 ES, adult HES, ERM, L4 extract and egg extract (10  $\mu$ g) were separated by 2-D gel electrophoresis, protein spots of interest excised, trypsinised and analysed by MALDI, or trypsinised and assessed by LC-MS/MS as previously described [25,27]. MS spectra were submitted to a locally running copy of Mascot (v2.3, Matrix Science) and searched against an improved in-house BLASTx annotated database of >57K isotigs and >92K singletons obtained by 454 sequencing of *H. polygyrus* infective L3, day 3 larvae, day 5 larvae, adults and eggs, with additional full length *H. polygyrus* sequences from NCBI and our own Sanger sequencing (Harcus Y. *et al*, manuscript in preparation). Search parameters required trypsin specificity, cysteine modification, possible oxidation of methionine, and allowed one missed cleavage. Gel spot identifications had a peptide tolerance of 100 ppm and MS/MS tolerance of 0.5 Da, with peptide expect scores <0.05. MudPit scoring was used for LC-MS with a <0.05





**Figure 8. HES and L4 ES immunization both induce recognition of VAL-1 and ACE-1.** **A.** 1-D immunoprecipitation (IP) of biotin-labelled HES (Bio-HES) and L4 ES (Bio-L4ES) with naïve mouse serum (nms) or sera from mice vaccinated with HES, L4 ES or ERM taken immediately before challenge with live larvae. Total biotin-labelled L4 ES and HES are included for comparison. **B–E.** 2-D immunoprecipitation of biotin-labelled HES (**B–C**) and L4 ES (**D–E**) with anti-HES (**B, D**) or anti-L4 ES (**C, E**). Spots corresponding to VAL-1 (blue), VAL-2 (red), VAL-3 (green) and ACE-1 (purple) are shown. Molecular weight markers (kDa) are indicated.

doi:10.1371/journal.ppat.1003492.g008

significance threshold. Single peptides were more stringently filtered for expect values  $<0.01$ . Each protein had at least one peptide not present in higher ranked hits and was manually inspected for open reading frames. Peptide false discovery rates were estimated using a decoy database (0.88% L4 ES; 2.18% HES; 1.5% ERM). Identified ORF were assessed for protein domains (Pfam v26.0) and N-terminal signal peptides (SignalP v4.0). The statistical significance of Pfam domain enrichment was determined for selected subsets using the hypergeometric functions of the “Category” Bioconductor package. Protein abundance was estimated by emPAI (exponentially modified protein abundance index [31] which calculates the ratio of observed:observable peptides. Spectral counts were also included for comparison. Sequence alignments and phylogenetic trees were carried out with MacVector (v11.1.1) and COBALT (NCBI). Heat maps and sequence clustering were generated with Artemis (Wellcome Trust Sanger Institute). When analysing stage-specific expression, proteins were treated as isogroups (i.e. closely related micro-variants of the same protein, likely representing alleles), meaning proteins were classed as shared even if the specific isotig was not detected (“shared isogroups”). Such analysis ensured that the determination of stage-specific protein expression was not skewed by the presence or absence of closely related alleles, detection of which likely depends on the relative abundance of the protein in the sample rather than any biological significant difference.

## Vaccinations

C57BL/6 female mice were immunized with 5  $\mu\text{g}$  L4 ES, HES, ERM or PBS control i.p. in alum adjuvant, boosted on days 28 and 35 with 1  $\mu\text{g}$  in alum i.p., before oral challenge with 200 *H. polygyrus* L3 at day 42. Fecal egg counts were determined at days 14, 21 and 28 post-challenge, and adult worms counted at day 28.

## ELISAs, Western blots and IP

Serum was obtained by tail bleeds from vaccinated mice immediately prior to challenge and assessed for anti-L4 ES, HES or ERM IgG1 reactivity as described before [27]. VAL-1, 2, 3, and 4, glycan A and B levels in the different ES preparations was determined using specific mAb obtained from primary infection (anti-VAL-1, 4-M15; anti-VAL-2, 4-S4; anti-VAL-4, 2-11; anti-glycan A, 13.1), secondary infection (anti-VAL-3, 5-S1) and HES immunisation (anti-glycan B, 9.1.3) ([27] and Filbey K.J *et al*, manuscript in preparation]. Western blotting of unlabelled HES, biotin-labelling and IP of ES material, followed by blotting and strep-HPO visualisation was carried out as before [27].

## TGF- $\beta$ bioassay

MFB-F11 cells [68] were cultured as before [51]  $\pm 20$   $\mu\text{g}/\text{ml}$  ES material or with varying amounts of rhTGF- $\beta$ 1 standard (R & D systems). After 24 hours, supernatants were collected and alkaline phosphatase activity determined with SEAP reporter assay kit (InvivoGen) as per manufacturer’s instructions.

## Statistical analysis

Statistical significance was determined by ANOVA or Mann-Whitney test where indicated. Correlation was measured with Spearman’s rank as emPAI values are non-parametric. All statistical analysis were performed using Prism (v6.0)

## Supporting Information

**Figure S1 L4 ES protein composition is more similar to HES than is ERM.** **A.** Comparison of emPAI values for proteins shared between L4 ES and HES. Spearman  $r$  values indicate

correlation coefficients, line indicates linear regression (\*\*\*) =  $p < 0.001$ ). **B.** As above for proteins shared between ERM and HES (n.s. = non-significant). **C.** Heat map of emPAI values for core 60 proteins shared between L4 ES, HES and ERM indicating differential expression. **D.** As above, for LAES-HES shared proteins. **E.** As above, for HES-ERM shared proteins. (EPS)

**Figure S2 Gene expression (RPKM) heat-maps of HES-specific and ERM proteins.** **A.** HES -specific proteins, clustered by gene expression profiles in a transcriptomic dataset based on 5 life cycle stages (infective L3, d3 post-infection L3, d5 post-infection L4, Adult, and Egg). RPKM = Reads Per Kilobase Mapped. Scores are coloured on a  $\log_2$  scale with red maximum and blue minimum. **B.** As above for proteins found only in ERM. (EPS)

**Figure S3 Stage-specific gene expression of L4 ES proteins.** Comparison of emPAI values of L4 ES proteins with RPKM gene expression levels from: **A.** Infective stage L3 larvae **B.** Day 3 post-infection L3 larvae **C.** Day 5 post-infection L4 larvae **D.** Adult worms **E.** Eggs. Spearman  $r$  values indicate correlation co-efficients (\*\*\*) =  $p < 0.001$ ; n.s. = non-significant). (EPS)

**Figure S4 A. Levels of VAL-1 in L4 ES (blue), HES (red) and ERM (yellow) determined by reactivity with the mAb 4-M15 [27].** **B.** As above for VAL-2 (mAb 4-S4). **C.** As above for VAL-3 (mAb 5-S1). **D.** As above for VAL-4 (mAb 2-11). **E.** As above for Glycan A (mAb 13.1). **F.** As above for Glycan B (mAb 9.1.3). (EPS)

**Figure S5 Sushi-domain proteins.** **A.** Phylogenetic tree of Sushi-domain containing proteins in L4 ES and HES. Blue domains indicate significant Pfam matches (E-value  $< 0.01$ ) to Sushi domain (pf00084); green domains indicate lower level similarities (E-value 0.01–0.05) retaining recognisable homology to pf00084. Signal peptides are depicted in yellow, presumed N-terminal truncations by broken lines, and stretches of  $> 50$  amino acids without detectable homology are indicated by black bars. **B.** Heat maps showing protein (emPAI) and transcript (RPKM) expression of indicated sushi-like proteins. (EPS)

**Figure S6 ShK/SXC-like proteins.** **A.** Cartoon indicating domain structure of ShK/SXC-like proteins (top) compared to other ShK domain proteins (bottom). ShK domain indicated by orange, N-terminal signal sequence by yellow, astacin domains by purple and presumed N-terminal truncation by broken lines. Heat maps showing protein (emPAI) and transcript (RPKM) expression of indicated proteins also shown. **B.** Sequence alignment of ShK/SXC-like proteins indicating mature protein following removal of N-terminal signal peptide. Positions of the canonical 6 conserved cysteine residues are indicated in yellow. (EPS)

**Table S1 Full A-Z list of proteins identified in L4 ES.** “emPAI rank” represents ranked abundance, “spectral count” is total number of peptides mapped to the protein of interest, whereas “peptides” is the number of different peptide sequences detected. “SS?” shows +/– N-terminal signal sequence, with \* indicating the sequence is N-terminally truncated but its closest BLAST homolog is SS+ve. emPAI values for L4 ES, HES and ERM are indicated, as are RPKM transcript levels for L3, day 3, day 5, adult and egg. Proteins present only as “shared isogroups” (see materials and methods) are shared grey. One protein

previously identified in HES, CSP-4, is present in 3 fragments in the transcriptomic assembly, and is only counted once.

(XLSX)

**Table S2 Full A-Z list of proteins identified in HES.** As above for HES.

(XLSX)

**Table S3 Full A-Z list of proteins identified in ERM.** As above for ERM.

(XLSX)

**Table S4 Full list of Pfam domains in different ES preparations.** Pfam domains present in ES proteins listed by expression pattern. emPAI and RPKM values included to show relative expression.

(XLSX)

## References

- Bethony JM, Cole RN, Guo X, Kamhawi S, Lightowlers MW, et al. (2011) Vaccines to combat the neglected tropical diseases. *Immunol Rev* 239: 237–270.
- Sargison ND (2012) Pharmaceutical treatments of gastrointestinal nematode infections of sheep—future of anthelmintic drugs. *Vet Parasitol* 189: 79–84.
- Boatin BA, Basanez MG, Prichard RK, Awadzi K, Barakat RM, et al. (2012) A research agenda for helminth diseases of humans: towards control and elimination. *PLoS Negl Trop Dis* 6: e1547.
- McSorley HJ, Maizels RM (2012) Helminth infections and host immune regulation. *Clin Micro Rev* 25: 585–608.
- Maizels RM, Holland M, Falcone FH, Zang XX, Yazdanbakhsh M (1999) Vaccination against helminth parasites: the ultimate challenge for immunologists? *Immunol Rev* 171: 125–148.
- Monroy FG, Enriquez EJ (1992) *Heligmosomoides polygyrus*: a model for chronic gastrointestinal helminthiasis. *Parasitol Today* 8: 49–54.
- Behnke JM, Menge DM, Noyes H (2009) *Heligmosomoides bakeri*: a model for exploring the biology and genetics of resistance to chronic gastrointestinal nematode infections. *Parasitology* 136: 1565–1580.
- Maizels RM, Hewitson JP, Murray J, Harcus Y, Dayer B, et al. (2012) Immune modulation and modulators in *Heligmosomoides polygyrus* infection. *Exp Parasitol* 132: 76–89.
- Wilson MS, Taylor M, Balic A, Finney CAM, Lamb JR, et al. (2005) Suppression of allergic airway inflammation by helminth-induced regulatory T cells. *J Exp Med* 202: 1199–1212.
- Finney CAM, Taylor MD, Wilson MS, Maizels RM (2007) Expansion and activation of CD4<sup>+</sup>CD25<sup>+</sup> regulatory T cells in *Heligmosomoides polygyrus* infection. *Eur J Immunol* 37: 1874–1886.
- Setiawan T, Metwali A, Blum AM, Ince MN, Urban JF, Jr., et al. (2007) *Heligmosomoides polygyrus* promotes regulatory T-cell cytokine production in the murine normal distal intestine. *Infect Immun* 75: 4655–4663.
- Rausch S, Huehn J, Kirchhoff D, Rzepecka J, Schnoeller C, et al. (2008) Functional analysis of effector and regulatory T cells in a parasitic nematode infection. *Infect Immun* 76: 1908–1919.
- Wilson MS, Taylor MD, O’Gorman MT, Balic A, Barr TA, et al. (2010) Helminth-induced CD19<sup>+</sup>CD23<sup>hi</sup> B cells modulate experimental allergic and autoimmune inflammation. *Eur J Immunol* 40: 1682–1696.
- Li Z, Liu G, Chen Y, Liu Y, Liu B, et al. (2011) The phenotype and function of naturally existing regulatory dendritic cells in nematode-infected mice. *Int J Parasitol* 41: 1129–1137.
- Smith KA, Hochweller K, Hämmerling GJ, Boon L, Macdonald AS, et al. (2011) Chronic helminth infection mediates tolerance in vivo through dominance of CD11c<sup>lo</sup> CD103<sup>+</sup> DC population. *J Immunol* 186: 7098–7109.
- Blum AM, Hang L, Setiawan T, Urban JF, Jr., Stoyanoff KM, et al. (2012) *Heligmosomoides polygyrus bakeri* induces tolerogenic dendritic cells that block colitis and prevent antigen-specific gut T cell responses. *J Immunol* 189: 2512–2520.
- Bashir ME, Andersen P, Fuss IJ, Shi HN, Nagler-Anderson C (2002) An enteric helminth infection protects against an allergic response to dietary antigen. *J Immunol* 169: 3284–3292.
- Elliott DE, Setiawan T, Metwali A, Blum A, Urban JF, Jr., et al. (2004) *Heligmosomoides polygyrus* inhibits established colitis in IL-10-deficient mice. *Eur J Immunol* 34: 2690–2698.
- Reynolds LA, Filbey KJ, Maizels RM (2012) Immunity to the model intestinal helminth parasite *Heligmosomoides polygyrus*. *Semin Immunopathol* 34: 829–846.
- Ali NMH, Behnke JM (1984) Non-specific immunodepression by larval and adult *Nematospiroides dubius*. *Parasitology* 88: 153–162.
- Behnke JM, Robinson M (1985) Genetic control of immunity to *Nematospiroides dubius*: a 9-day anthelmintic abbreviated immunizing regime which separates weak and strong responder strains of mice. *Parasite Immunol* 7: 235–253.
- Hagan P, Behnke JM, Parish HA (1981) Stimulation of immunity to *Nematospiroides dubius* in mice using larvae attenuated by cobalt 60 irradiation. *Parasite Immunol* 3: 149–156.
- Hewitson JP, Grainger JR, Maizels RM (2009) Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. *Mol Biochem Parasitol* 167: 1–11.
- Loukas A, Gaze S, Mulvenna JP, Gasser RB, Brindley PJ, et al. (2011) Vaccinomics for the major blood feeding helminths of humans. *OMICS* 15: 567–577.
- Hewitson JP, Harcus Y, Murray J, van Agtmaal M, Filbey KJ, et al. (2011) Proteomic analysis of secretory products from the model gastrointestinal nematode *Heligmosomoides polygyrus* reveals dominance of Venom Allergen-Like (VAL) proteins. *J Proteomics* 14: 1573–1594.
- Moreno Y, Gros PP, Tam M, Segura M, Valanparambil R, et al. (2011) Proteomic analysis of excretory-secretory products of *Heligmosomoides polygyrus* assessed with next-generation sequencing transcriptomic information. *PLoS Negl Trop Dis* 5: e1370.
- Hewitson JP, Filbey KJ, Grainger JR, Dowlé AA, Pearson M, et al. (2011) *Heligmosomoides polygyrus* elicits a dominant nonprotective antibody response directed at restricted glycan and peptide epitopes. *J Immunol* 187: 4764–4777.
- Bryant V (1973) The life cycle of *Nematospiroides dubius*, Baylis, 1926 (Nematoda: Heligmosomidae). *J Helminthol* 47: 263–268.
- Rzepecka J, Rausch S, Klotz C, Schnöller C, Kornprobst T, et al. (2009) Calreticulin from the intestinal nematode *Heligmosomoides polygyrus* is a Th2-skewing protein and interacts with murine scavenger receptor-A. *Mol Immunol* 46: 1109–1119.
- McCoy KD, Stoel M, Stettler R, Merky P, Fink K, et al. (2008) Polyclonal and specific antibodies mediate protective immunity against enteric helminth infection. *Cell Host Microbe* 4: 362–373.
- Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, et al. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 4: 1265–1272.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5: 621–628.
- Gibbs GM, Roelants K, O’Byrne MK (2008) The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins—roles in reproduction, cancer, and immune defense. *Endocr Rev* 29: 865–897.
- Cantacessi C, Campbell BE, Visser A, Geldhof P, Nolan MJ, et al. (2009) A portrait of the “SCP/TAPS” proteins of eukaryotes - Developing a framework for fundamental research and biotechnological outcomes. *Biotechnol Adv* 27: 376–388.
- Chalmers IW, Hoffmann KF (2012) Platyhelminth Venom Allergen-Like (VAL) proteins: revealing structural diversity, class-specific features and biological associations across the phylum. *Parasitology* 139: 1231–1245.
- Reid KB, Day AJ (1989) Structure-function relationships of the complement components. *Immunol Today* 10: 177–180.
- Wang X, Lupardus P, Laporte SL, Garcia KC (2009) Structural biology of shared cytokine receptors. *Annu Rev Immunol* 27: 29–60.
- Gems DH, Ferguson CJ, Robertson BD, Nieves R, Page AP, et al. (1995) An abundant, trans-spliced mRNA from *Toxocara canis* infective larvae encodes a 26 kDa protein with homology to phosphatidylethanolamine binding proteins. *J Biol Chem* 270: 18517–18522.
- Loukas AC, Hintz M, Tetteh KKA, Mullin NP, Maizels RM (2000) A family of secreted mucins from the parasitic nematode *Toxocara canis* bear diverse mucin domains but share similar flanking six-cysteine repeat motifs. *J Biol Chem* 275: 39600–39607.
- Hawdon JM, Jones BF, Perregaux MA, Hotez PJ (1995) *Ancylostoma caninum*: metalloprotease release coincides with activation of infective larvae *in vitro*. *Exp Parasitol* 80: 205–211.
- Feng J, Zhan B, Liu Y, Liu S, Williamson A, et al. (2007) Molecular cloning and characterization of Ac-MTP-2, an astacin-like metalloprotease released by adult *Ancylostoma caninum*. *Mol Biochem Parasitol* 152: 132–138.

**Table S5 Enrichment of Pfam domains in different ES preparations.** Pfam domains present in ES proteins ranked by statistical significance compared to their frequency in the *H. polygyrus* transcriptome assembly.

(XLSX)

## Author Contributions

Conceived and designed the experiments: JPH RMM. Performed the experiments: JPH YH KJF HJM JM AAD. Analyzed the data: JPH ACI SB DA AAD RMM. Wrote the paper: JPH RMM.

42. Williamson AL, Lustigman S, Oksov Y, Deumic V, Plieskatt J, et al. (2006) *Ancylostoma caninum* MTP-1, an astacin-like metalloprotease secreted by infective hookworm larvae, is involved in tissue migration. *Infect Immun* 74: 961–967.
43. Redmond DL, Smith SK, Halliday A, Smith WD, Jackson F, et al. (2006) An immunogenic cathepsin F secreted by the parasitic stages of *Teladorsagia circumcincta*. *International Journal for Parasitology* 36: 277–286.
44. De Vries E, Bakker N, Krijgsveld J, Knox DP, Heck AJ, et al. (2009) An AC-5 cathepsin B-like protease purified from *Haemonchus contortus* excretory secretory products shows protective antigen potential for lambs. *Vet Res* 40: 41.
45. Hasnain SZ, McGuckin MA, Grecis RK, Thornton DJ (2012) Serine protease(s) secreted by the nematode *Trichuris muris* degrade the mucus barrier. *PLoS Negl Trop Dis* 6: e1856.
46. Romaris F, North SJ, Gagliardo LF, Butcher BA, Ghosh K, et al. (2002) A putative serine protease among the excretory-secretory glycoproteins of L1 *Trichinella spiralis*. *Mol Biochem Parasitol* 122: 149–160.
47. Balasubramanian N, Toubarro D, Simoes N (2010) Biochemical study and in vitro insect immune suppression by a trypsin-like secreted protease from the nematode *Steinernema carpocapsae*. *Parasite Immunol* 32: 165–175.
48. Zang X, Maizels RM (2001) Serine proteinase inhibitors from nematodes and the arms race between host and pathogen. *Trends Biochem Sci* 26: 191–197.
49. Milstone AM, Harrison LM, Bungiro RD, Kuzmic P, Cappello M (2000) A broad spectrum Kunitz type serine protease inhibitor secreted by the hookworm *Ancylostoma ceylanicum*. *J Biol Chem* 275: 29391–29399.
50. Wu Y, Preston G, Bianco AE (2008) Chitinase is stored and secreted from the inner body of microfilariae and has a role in exsheathment in the parasitic nematode *Brugia malayi*. *Mol Biochem Parasitol* 161: 55–62.
51. Grainger JR, Smith KA, Hewitson JP, McSorley HJ, Harcus Y, et al. (2010) Helminth secretions induce *de novo* T cell Foxp3 expression and regulatory function through the TGF- $\beta$  pathway. *J Exp Med* 207: 2331–2341.
52. McSorley HJ, Grainger JR, Harcus YM, Murray J, Nisbet A, et al. (2010) *daf-7*-related TGF- $\beta$  homologues from trichostrongyloid nematodes show contrasting life cycle expression patterns. *Parasitology* 137: 159–171.
53. Moyle M, Foster DL, McGrath DE, Brown SM, Laroche Y, et al. (1994) A hookworm glycoprotein that inhibits neutrophil function is a ligand of the integrin CD11b/CD18. *J Biol Chem* 269: 10008–10015.
54. Winkler B, Bolwig C, Seppälä U, Spangfort MD, Ebner C, et al. (2003) Allergen-specific immunosuppression by mucosal treatment with recombinant Ves v 5, a major allergen of *Vespa vulgaris* venom, in a murine model of wasp venom allergy. *Immunology* 110: 376–385.
55. Kotwal GJ, Isaacs SN, McKenzie R, Frank MM, Moss B (1990) Inhibition of the complement cascade by the major secretory protein of vaccinia virus. *Science* 250: 827–830.
56. Tudor JE, Pallaghy PK, Pennington MW, Norton RS (1996) Solution structure of ShK toxin, a novel potassium channel inhibitor from a sea anemone. *Nat Struct Biol* 3: 317–320.
57. Chi V, Pennington MW, Norton RS, Tarcha EJ, Londono LM, et al. (2012) Development of a sea anemone toxin as an immunomodulator for therapy of autoimmune diseases. *Toxicon* 59: 529–546.
58. Hewitson JP, Harcus YM, Curwen RS, Dowle AA, Atmadja AK, et al. (2008) The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory-secretory products. *Mol Biochem Parasitol* 160: 8–21.
59. Bennuru S, Semnani R, Meng Z, Ribeiro JM, Veenstra TD, et al. (2009) *Brugia malayi* excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling. *PLoS Negl Trop Dis* 3: e410.
60. Bennuru S, Meng Z, Ribeiro JM, Semnani RT, Ghedin E, et al. (2011) Stage-specific proteomic expression patterns of the human filarial parasite *Brugia malayi* and its endosymbiont *Wolbachia*. *Proc Natl Acad Sci U S A* 108: 9649–9654.
61. Jiang D, Malone J, Townsend R, Weil GJ, Li B (2012) Multiplex proteomics analysis of gender-associated proteins in *Brugia malayi*. *Int J Parasitol* 42: 841–850.
62. Li BW, Wang Z, Rush AC, Mitreva M, Weil GJ (2012) Transcription profiling reveals stage- and function-dependent expression patterns in the filarial nematode *Brugia malayi*. *BMC Genomics* 13: 184.
63. Mulvenna J, Hamilton B, Nagaraj SH, Smyth D, Loukas A, et al. (2009) Proteomics analysis of the excretory/secretory component of the blood-feeding stage of the hookworm, *Ancylostoma caninum*. *Mol Cell Proteomics* 8: 109–121.
64. Cantacessi C, Mitreva M, Jex AR, Young ND, Campbell BE, et al. (2010) Massively parallel sequencing and analysis of the *Necator americanus* transcriptome. *PLoS Negl Trop Dis* 4: e684.
65. Yatsuda AP, Krijgsveld J, Cornelissen AWCA, Heck AJ, De Vries E (2003) Comprehensive analysis of the secreted proteins of the parasite *Haemonchus contortus* reveals extensive sequence variation and differential immune recognition. *J Biol Chem* 278: 16941–16951.
66. Nisbet AJ, Smith SK, Armstrong S, Meikle LI, Wildblood LA, et al. (2010) *Teladorsagia circumcincta*: activation-associated secreted proteins in excretory/secretory products of fourth stage larvae are targets of early IgA responses in infected sheep. *Exp Parasitol* 125: 329–337.
67. Smith SK, Nisbet AJ, Meikle LI, Inglis NF, Sales J, et al. (2009) Proteomic analysis of excretory/secretory products released by *Teladorsagia circumcincta* larvae early post-infection. *Parasite Immunol* 31: 10–19.
68. Tesseur I, Zou K, Berber E, Zhang H, Wyss-Coray T (2006) Highly sensitive and specific bioassay for measuring bioactive TGF- $\beta$ . *BMC Cell Biol* 7: 15.
69. Osman A, Wang CK, Winter A, Loukas A, Tribollet L, et al. (2012) Hookworm SCP/TAPS protein structure—A key to understanding host-parasite interactions and developing new interventions. *Biotechnol Adv* 30: 652–657.

## Appendix 4B

Declaration from James Hewitson attesting to my contribution to paper 4.



UNIVERSITY  
*of York*

## Declaration of reviewee's contribution

### Paper

Hewitson, J.P., Ivens, A.C., Harcus, Y., Filbey, K.J., McSorley, H.J., Murray, J., Bridgett, S., Ashford, D., **DOWLE, A.A.**, and Maizels, R.M. (2013) Secretion of Protective Antigens by Tissue-Stage Nematode Larvae Revealed by Proteomic Analysis and Vaccination-Induced Sterile Immunity *PLoS Pathog.* **9(8)**

### Candidate's contribution

The submitting candidate performed the mass spectrometry component of the paper. Gel spots, bands and protein extracts were provided by James Hewitson. The reviewee's contribution to the work included: protein digestion, MALDI-MS/MS, LC-MS/MS, database searching and associated data analysis. The aim of the proteomics analysis was to compare both within and between fractions using emPAI. The work was provided as fee for service through the Bioscience Technology Facility, Department of Biology, University of York. Contributions to data analysis and input into writing the paper were provided on a collaborative basis.

### Declaration

I attest that the summation above is a true reflection of the candidate's (Adam A. Dowle) contribution to the paper.

### Candidate

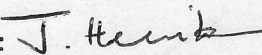
Name: Adam Dowle

Signature: 

Date: 1st June 2016

### Co-author

Name: James Hewitson

Signature: 

Date: 01.06.16

## Appendix 5A

Submitted paper 5 – Plasma membrane proteomes of differentially matured dendritic cells identified by LC-MS/MS combined with iTRAQ labelling.



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SciVerse ScienceDirect

[www.elsevier.com/locate/jprot](http://www.elsevier.com/locate/jprot)

# Plasma membrane proteomes of differentially matured dendritic cells identified by LC–MS/MS combined with iTRAQ labelling

Stéphanie Ferret-Bernard<sup>a,1</sup>, William Castro-Borges<sup>a,2</sup>, Adam A. Dowle<sup>a,b</sup>, David E. Sanin<sup>a</sup>, Peter C. Cook<sup>a</sup>, Joseph D. Turner<sup>a</sup>, Andrew S. MacDonald<sup>c</sup>, Jerry R. Thomas<sup>a,b</sup>, Adrian P. Mountford<sup>a,\*</sup>

<sup>a</sup>Department of Biology, University of York, York, YO10 5DD, UK

<sup>b</sup>Centre of Excellence in Mass Spectrometry, University of York, York, YO10 5DD, UK

<sup>c</sup>Institute of Immunology and Infection Research, School of Biological Sciences, University of Edinburgh, UK

## ARTICLE INFO

### Article history:

Received 8 August 2011

Accepted 17 October 2011

Available online 25 October 2011

### Keywords:

Dendritic cell

iTRAQ

Parasitic helminth

Plasma membrane proteomics

## ABSTRACT

Dendritic cells (DCs) play a pivotal role in polarising Th lymphocyte subsets but it is unclear what molecular events occur when DCs generate Th2-type responses. Here, we analysed plasma membrane-enriched fractions from immature, pro-Th1 and pro-Th2 DCs and used a combination of iTRAQ labelling and LC–MS/MS to quantify changes in the proteomes. Analysis was performed on triplicate biological samples and changes verified by flow cytometry. MHC class II molecules and CD29 were up-regulated in pro-Th1 DCs whilst CD18 and CD44 were up-regulated in pro-Th2 DCs. One of the most down-regulated molecules in pro-Th1 DCs was YM-1 whilst the greatest decrease in pro-Th2 DCs was NAP-22. Other molecules up-regulated in pro-Th2 DC compared to pro-Th1 DCs included some potentially involved in protein folding during antigen processing (clathrin and Rab-7), whilst other non-membrane proteins such as enzymes/transporters related to cell metabolism (malate dehydrogenase, pyruvate kinase, and ATPase Na<sup>+</sup>/K<sup>+</sup>) were also recorded. This suggests that pro-Th2 DCs are more metabolically active while pro-Th1 DCs have a mature ‘end state’. Overall, although several molecules were preferentially expressed on pro-Th2 DCs, our proteomics data support the view of a ‘limited maturation’ of pro-Th2 DCs compared to pro-Th1 DCs.

Crown Copyright © 2011 Published by Elsevier B.V. Open access under [CC BY license](http://creativecommons.org/licenses/by/3.0/).

## 1. Introduction

Dendritic cells (DCs) are critical in development of immunity against pathogens [1] and are able to interpret different

pathogen-inherent signals to play a pivotal role in polarising Th lymphocyte subsets [2]. In general, pathogen-associated molecular patterns (PAMPs) that drive DCs to promote Th1-type responses, such as bacterial lipopolysaccharide (LPS), bind to

**Abbreviations:** 0–3hRP, zero-to-three hours released proteins; Arp2/3, actin-related protein 2/3 complex; BM, bone marrow; CD, cluster of differentiation; DC, dendritic cell; E/S, excretory/secretory; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; GM-CSF, granulocyte-macrophage colony-stimulating factor; GNB, guanine nucleotide-binding protein; LPS, lipopolysaccharide; MFI, median fluorescence intensity; NAP-22, 22 kDa neuronal tissue-enriched acidic protein; PAMP, pathogen-associated molecular pattern; PRR, pattern recognition receptor; SEA, schistosoma egg antigen; Th, T helper.

\* Corresponding author at: Department of Biology (Area 5), University of York, York, YO10 5DD, UK. Tel.: +44 1904 328595; fax: +44 1904 328505.

E-mail address: [adrian.mountford@york.ac.uk](mailto:adrian.mountford@york.ac.uk) (A.P. Mountford).

<sup>1</sup> Current address: INRA, UMR 1079, SENAH, F-35590 Saint-Gilles, France.

<sup>2</sup> Current address: Departamento de Ciências Biológicas, Universidade Federal de Ouro Preto, Ouro Preto, Brasil.

1874-3919 Crown Copyright © 2011 Published by Elsevier B.V. Open access under [CC BY license](http://creativecommons.org/licenses/by/3.0/).

doi:10.1016/j.jprot.2011.10.010

pattern recognition receptors (PRRs) on the DC plasma membrane leading to subsequent activation and maturation of the DCs. Conversely, molecules from helminths, which are potent inducers of Th2-type responses, stimulate DCs with a 'modified' phenotype [3–7] but it is unclear how/why these DCs promote Th2-type immunity. Pro-Th2 DCs may have a distinct phenotype, defined by a unique profile of signature molecules [4,5], or may resemble immature DCs, stimulating Th2 polarisation via a 'default' pathway in the absence of Th1-inducing stimuli [8]. An 'inhibition model' in which Th2-stimuli inhibit Th1 polarisation by DCs through competitive signalling pathways has also been proposed [9].

Changes in DC gene expression correlate poorly with changes in the level of protein expression [10,11]. Therefore, as protein expression is a better indicator of cell phenotype and function, a number of studies have examined the proteomes of differentially matured DCs [12–15]. For example, 2-DE and MS/MS have revealed changes in the proteome of pro-Th1 DCs matured with LPS versus pro-Th2 DCs stimulated with excretory/secretory (E/S) material from a parasitic helminth *Schistosoma mansoni* [14]. This E/S material, released by the parasite during infection and known as '0–3hRP', is an important stimulant of innate immune cells in the skin [16,17] enabling DCs to promote Th2 responses *in vitro* and *in vivo* [7]. Proteomic analysis of pro-Th1 DCs revealed up-regulated expression of cytoskeletal proteins and chaperone molecules whereas pro-Th2 DCs, stimulated with 0–3hRP, exhibited a proteome intermediate between that of immature DCs and pro-Th1 DCs; thus termed a 'limited maturation' phenotype [14]. As soluble cytosolic molecules dominated the cell extracts used in the study, immune-associated proteins from the plasma membrane (*e.g.* PRRs, adhesion molecules, MHC complexes and costimulatory molecules) were not readily detected. Although such molecules are likely to be highly relevant with respect to differential maturation of DCs, their low abundance and hydrophobic nature makes them difficult to isolate for proteomic characterisation.

In order to 'home in' on the detection of specific plasma membrane proteins which are likely to be important in differential DC maturation, we compared proteins enriched from plasma membranes of immature DCs, pro-Th1 DCs stimulated with LPS and pro-Th2 DCs stimulated with schistosome egg antigen (SEA) which is a well characterised pro-Th2 helminth product [3]. First, DC plasma membrane-enriched fractions were analysed by shotgun LC-MS/MS to establish a list of proteins associated with DCs. Second, a gel-free technique using iTRAQ [18,19] was used to quantify changes in protein expression following differential DC maturation. By performing three biological replicates of each type of DCs and confirming proteomic data by flow cytometry, we identified a number of proteins that were differentially expressed by pro-Th1 versus pro-Th2 DCs.

## 2. Materials and methods

### 2.1. Generation and maturation of DCs from bone marrow

Bone marrow-derived dendritic cells (BM-DCs) from female C57BL/6 strain mice were cultured in RPMI medium containing 10% low endotoxin FCS plus 20 ng/mL GM-CSF (Peprotech,

London, UK) as previously described [7,14]. All experimental procedures were undertaken with the guidelines of the United Kingdom Animal's Scientific Procedures Act 1986 and approved by the University of York Ethics committee. On day 6, immature BM-DCs were seeded at  $1 \times 10^6$ /mL and cultured for 18 h alone (MED-DCs), or in the presence of 40  $\mu$ g/mL SEA [3] (SEA-DCs), or 10 ng/mL LPS (from *Escherichia coli* strain 0111:B4, Sigma-Aldrich, Poole, UK; LPS-DCs) [14]. After overnight culture, cells were harvested and prepared for proteomic analysis.

### 2.2. Preparation of DC plasma membrane-enriched fractions

Plasma membrane proteins from MED-DCs, SEA-DCs and LPS-DCs were extracted and purified using a plasma membrane protein extraction kit (BioVision, Mountain View, USA). All steps were performed at 4 °C. Briefly, BM-DCs were mechanically homogenised in an ice-cold glass cell grinder and then spun at 700 *g*. The resulting supernatants were spun at 10,000 *g* for 30 min to yield total membrane protein (*i.e.* plasma and cellular organelle membranes) enriched pellets which were re-suspended in 'Upper Phase solution' and mixed with an equal volume of 'Lower Phase solution' before centrifugation at 1000  $\times$  *g* for 5 min. The upper phase was spun at 25,000 *g* for 10 min, and the resulting plasma membrane-enriched pellet solubilised in 0.5% Triton X-100. Total protein content was assessed by densitometry of SYPRO Ruby stained 1-D electrophoresis gels (NuPAGE 4–12%) against known quantities of cytosolic fractions as standards separated on the same gel.

### 2.3. Digestion and iTRAQ labelling

Plasma membrane-enriched fractions (35–50  $\mu$ g) were reduced with 2 mM tris-(2-carboxyethyl)phosphine in 0.5 M triethylammonium bicarbonate (Sigma-Aldrich) pH8.5, at 60 °C for 1 h, alkylated with 10 mM methyl methanethiosulfonate (Sigma-Aldrich) at room temperature for 10 min and digested overnight with sequencing-grade porcine trypsin (Promega, Madison, USA) at 37 °C. The iTRAQ labelling reagents (114, 115, and 116; Applied Biosystems, Framingham, USA) were reconstituted in isopropanol and added to the digests. After 2 h, labelled peptides were combined and purified using cation-exchange and C18 cartridges. Although iTRAQ allows for the multiplexing of several samples in a single run, comparison is performed in a pair-wise manner. In this respect, a common reference sample between iTRAQ analyses is essential. However, a pooled standard is not universally required if the samples are suitably similar, as in our study, where MED-DC is taken as a common standard across all runs [20].

### 2.4. LC-MS/MS

Peptides were separated using a Dionex polystyrene-divinylbenzene column and fractions collected directly onto a target plate with addition of CHCA matrix. Positive-ion MALDI mass spectra were acquired using an Applied Biosystems 4700 Proteomics Analyzer in reflectron mode over a mass range of *m/z* 800–4000. Monoisotopic masses were obtained from centroids of raw, unsmoothed data. The 20 most intense peaks with a *S/N*  $\geq$  50 from each fraction were selected for CID-MS/MS

using collision energy of 1 keV, air as collision gas, precursor mass window set to a relative resolution of 50 and metastable suppressor enabled. MS/MS spectra were baseline-subtracted (peak width 50) and smoothed (Savitsky–Golay; three points; polynomial order 4); peak detection used  $S/N \geq 5$ , local noise window 50 m/z and minimum peak width 2.9 bins. Mascot peak list files were generated using the TS2Mascot utility (Matrix Science, version 1.0.0) with  $S/N \geq 10$ .

### 2.5. Protein identification and quantification

Peak lists were searched against the CDS (Celera Discovery System™, KBMS3.0.20040121) mouse database (65,307 sequences; 23,201,165 residues) using Mascot (Matrix Science Ltd., version 2.1). Search criteria specified: Enzyme, Trypsin; Maximum missed cleavages, 1; Fixed modifications, Methylthio (C); Variable modifications, Oxidation (M); Peptide tolerance, 0.3 Da; MS/MS tolerance, 0.3 Da; Instrument, MALDI-TOF-TOF. When searching iTRAQ peak lists the fixed modifications, iTRAQ4plex (N-term) and iTRAQ4plex (K), and the variable modification iTRAQ4plex (Y), were also specified. The Mascot significance threshold for protein identification was adjusted in each search to give a false discovery rate of approximately 1%. Only peptides that met or exceeded their identity score at this significance threshold and had an expect score less than 0.05 were accepted, including single-peptide protein matches. Where peptides match to multiple members of a protein family, the protein that tops the Mascot protein group and hence has the highest number of unique peptides, is reported. Mascot was used for iTRAQ quantification with these options: normalisation by median ratio; automatic outlier removal; median ratios for protein ratios. Fold changes were accepted only with three peptides. Each iTRAQ experiment included three biological replicates.

### 2.6. Sub-cellular and Gene Ontology classification of proteins

Sub-cellular classifications were performed with Gene Ontology classification (<http://www.ebi.ac.uk/GOA/>) according to the accession number of protein identities in Uniprot. If no classification was found, Mouse Genome Informatics website (<http://www.informatics.jax.org/>) and the Proteome Analyst v3.0 (<http://pa.cs.ualberta.ca:8080/pa/>) were used. Finally, if no prediction was available, we referred to published literature. Proteins identified as 'plasma membrane' included formations such as lipid rafts, endosomes, phagocytic cups and phagosomes. 'Cytosol' comprised cytosolic, cytoplasmic proteins and vesicles. Cytoskeletal proteins were given a particular column although most are also cytosolic. Ribosome proteins were classified as belonging to the 'ER', whilst proteins associated with the Golgi apparatus, late endosomes and lysosomes, were defined as 'Golgi'. However, classification of proteins into discrete compartments is arbitrary depending upon original allocation by bio-informatic interrogation and may be best allocated to more than one classification. Protein hits that were observed to be differentially regulated (i.e. SEA-DCs vs MED-DCs, and LPS-DCs vs MED-DCs) were analysed according to Gene Ontology (GO) classification using Visual Annotation Display (VLAD; [proto.informatics.jax.org/prototypes/vlad-1.0.3/](http://proto.informatics.jax.org/prototypes/vlad-1.0.3/)). Significant hits were selected on the basis of their p-value and the number of proteins(k) found within each category.

### 2.7. Validation of protein expression by flow cytometry

Differentially matured DCs were blocked with anti-CD16/32 mAb (BD Pharmingen, Oxford, UK) in PBS (supplemented with 1% FCS and 5 mM EDTA) and subsequently labelled with the following conjugated mAbs against: CD18-biotin (M18/2), CD29-PE (HmB1-1), CD44-FITC (1C10), CD98-FITC (RL388), IA/IE-FITC (M5/114.15.2), Galectin-3-biotin (M3/38; all from eBioscience, Hatfield, UK), I-A<sup>b</sup> biotin (28-16-8S; Caltag Medsystems, Buckingham, UK) and CD301b-AF647 (ER-MP23; AbD Serotec, Kidlington, UK). Biotin conjugated antibodies were probed with streptavidin-APC (Caltag Medsystems). Unlabelled rabbit polyclonal antibodies were used to probe for S100A10 (Abcam, Cambridge, UK), YM-1 (STEMCELL Technologies, Grenoble, France) and Rab-7 (Sigma-Aldrich), and were subsequently detected with anti-rabbit AF488 antibody (eBioscience). All antibody concentrations were optimised and labelling was performed alongside relevant isotype controls. Flow cytometric acquisition and analysis was performed using a Cyan ADP analyser with Summit v4.3 (DakoCytomation, UK). Data were plotted as means of the median fluorescence intensity (MFI) of three separate DC cultures for each maturation stimulus.

## 3. Results

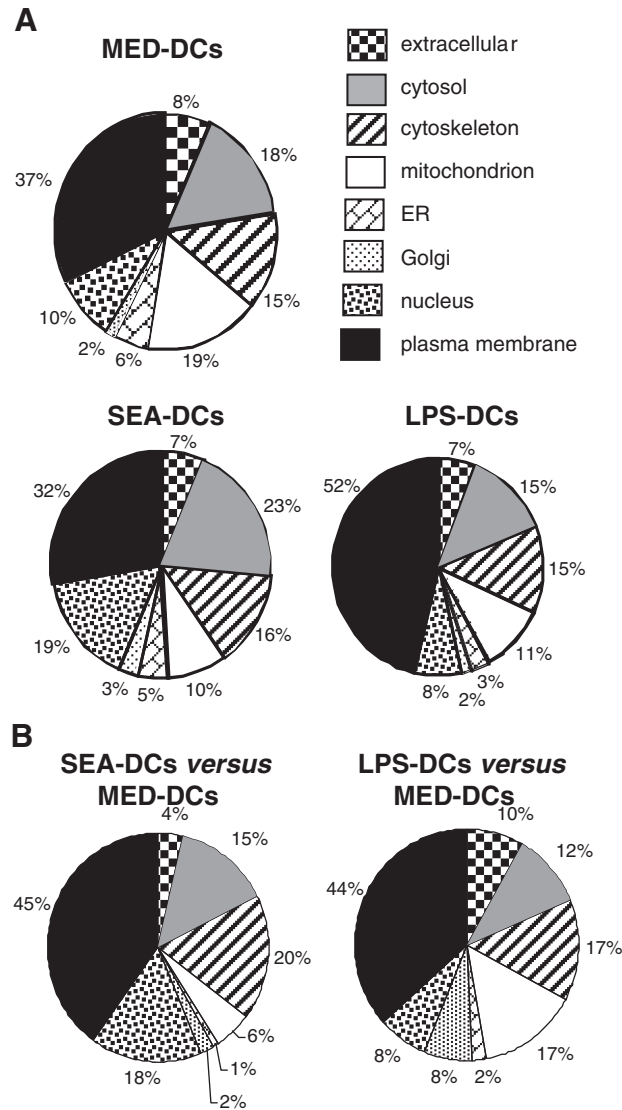
### 3.1. Sub-cellular classification of proteins identified by shotgun LC-MS/MS

Initial shotgun LC-MS/MS proteomic analyses of DC fractions were performed to validate the isolation technique. In MED-DCs, 119 significant hits were identified (Supp. Table 1) of which 37% were classified by Gene Ontology as plasma membrane (Fig. 1A). In SEA-DCs and LPS-DCs, 149 and 61 proteins respectively were identified (Supp. Tables 2 and 3) with 32% and 52% from the plasma membrane (Fig. 1A). The fractions also contained cytosolic (19%), cytoskeletal (15%) and nuclear proteins whilst extracellular, ER and Golgi proteins were only minor components. In six different purification experiments, the enrichment for plasma membrane proteins in the fractions, compared to the total cell, was  $23.4 \pm 3.1$ . Consequently, we concluded that the isolation technique yielded fractions greatly enriched in plasma membrane molecules.

Plasma membrane molecules (shaded rows in bold text of Supp. Tables 1, 2, and 3) comprised those with a CD prefix (18%) and ras-related proteins (rho, rac, rab and rap; 17%) which were highly represented from all three types of DC. Other abundant molecules were MHC class I and II molecules for LPS-DCs (13%) and guanine nucleotide-binding proteins (GNBPs) in MED-DCs (11%). Annexins (1, 2, 4 and 5) and S100 proteins represented 10% and 5% respectively, of plasma membrane proteomes of all DC types.

### 3.2. Plasma membrane proteomes of differentially matured DCs after iTRAQ labelling

The number of hits identified after iTRAQ labelling in SEA-DCs versus MED-DCs, and LPS-DCs versus MED-DCs increased compared to the number of hits in SEA-DCs and LPS-DCs identified solely by LC-MS/MS (+66% and +46% respectively; Supp.



**Fig. 1 – Sub-cellular classification of the total identified proteins. Pie-charts showing the classification of total protein identities in plasma membrane fractions obtained from MED-DCs, SEA-DCs and LPS-DCs determined by LC-MS/MS (A). Pie-charts showing the classification of the total hits obtained from three technical replicates of iTRAQ experiments comparing SEA-DCs or LPS-DCs versus MED-DCs (B).**

Tables 4, 5 versus 1, 2, 3). This resulted in expanded DC proteome coverage and database searching of multiple peptides per protein improved the confidence of the identifications. Nevertheless, the proportion of hits identified after iTRAQ labelling and classed as plasma membrane were similar to that revealed by shotgun LC-MS/MS.

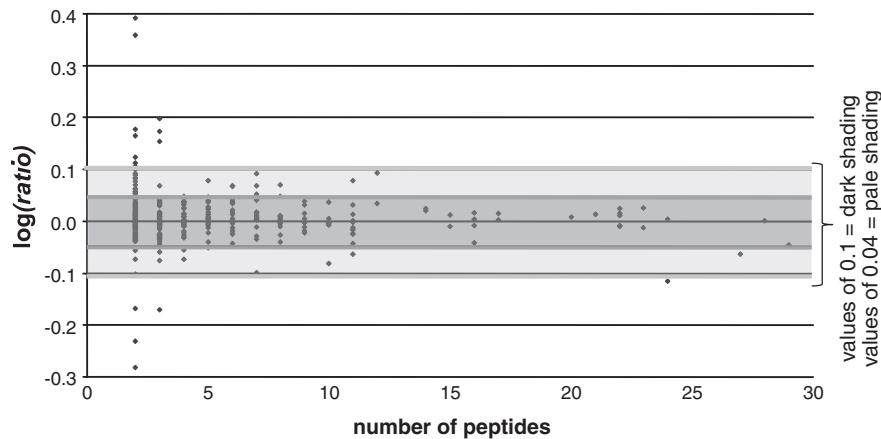
Compilation of hits identified in the three separate iTRAQ experiments of SEA-DCs versus MED-DCs showed that 222 proteins were significantly identified with 45% being classed as plasma membrane components (Fig. 1B, Supp. Table 4) while a similar proportion (44%) was identified in LPS-DCs versus MED-DCs (Fig. 1B, Supp. Table 5). Overall, a greater proportion of hits were identified as nuclear proteins in SEA-DCs versus MED-DCs (18%) compared to LPS-DCs versus MED-DCs (8%) whereas a greater proportion of proteins in LPS-DCs versus MED-DCs were mitochondrial, extracellular and associated with the Golgi apparatus.

### 3.3. Up- or down-regulated proteins from differentially matured DC membrane-enriched proteomes

In order to assess variance in iTRAQ quantification experiments, Spooncer et al. [21] devised a form of analysis to compare separate preparations of the same biological material [22]. Ratios obtained by comparing separate preparations of the same material plotted against the number of peptides identified (Fig. 2) should cluster around unity (i.e. protein levels should be the same between replicates). The extent to which values deviate from unity is an indication of the biological and technical variance. This approach provided a powerful visual indication of where to set the significance threshold.

Our iTRAQ experimental design consisted of six labelling experiments (Table 1), in which three preparations of SEA-DCs and three preparations of LPS-DCs were analysed. MED-DC fractions were labelled with 114 and 116 iTRAQ





**Fig. 2 – Whetton's plot. iTRAQ reporter-ion ratios for all proteins in the control versus control comparisons using the 114 and 116 reagents. A total of 355 protein ratios were plotted against the number of peptides that contributed to the ratio. Shading zones represent two levels of threshold: at  $\log(\text{ratio})$  values of  $\pm 0.1$  (ratios 0.79–1.26 in mid-grey) and a less stringent threshold of  $\pm 0.04$  (ratios 0.9–1.1 in light grey).**

reagents to provide technical replicates whilst SEA-DC or LPS-DC preparations were labelled with 115. All the protein ratios from the six control-control (114:116) comparisons (355 in total) are shown in Fig. 2. As a more stringent threshold, we chose  $\log(\text{ratio})$  values of  $\pm 0.1$ , which correspond to ratios of 1.26 and 0.79; a less stringent threshold of  $\pm 0.04$ , which correspond to ratios of 1.1 and 0.9 was also used. Only changes where the average  $\log(\text{ratio}) \pm \text{SEM} - 95\%$  exceeded these thresholds were considered significant as given for SEA-DCs versus MED-DCs (Fig. 3, Supp. Table 6) and LPS-DCs versus MED-DCs (Fig. 4, Supp. Table 7). The more stringent threshold is shown as shaded bars and rows.

In SEA-DCs versus MED-DCs, the greatest up-regulation was for nuclear proteins (mainly histones) while several proteins associated with antigen uptake, processing and presentation (chaperones, clathrin, enzymes and the small GTPase Rab-7) were also up-regulated (Fig. 3, Supp. Table 6). Other molecules which were significantly up-regulated included: S100A9, 14-3-3 $\beta/\alpha$  (part of TGF $\beta$  family), lymphocyte cytosolic protein 1 and lysozyme C. In contrast, decreased expression was observed for Fc $\epsilon$ RI, Rap-1B, hematopoietic cell specific lyn substrate 1, CD44, several annexins, MHC class II molecule (I-A), subunit 2 of actin-related protein 2/3 (Arp2/3) complex, lymphocyte-specific 1 protein (LSP1) and S100A10 (Fig. 3, Supp. Table 6). Down-regulation was also recorded for certain cytoskeletal proteins (ezrin, moesin, myosin and several isoforms of actin). The most pronounced down-regulation affected a protein not expected to be expressed by DCs: NAP-22 (22 kDa neuronal tissue-enriched acidic protein). Finally, several immune-related proteins were expressed at similar levels between SEA-DCs and MED-DCs (Fig. 3) including the integrins CD11b and CD18, the macrophage lectin CD301b, CD98 heavy chain, CD45, as well as MHC class I molecules (H-2D<sup>b</sup> and H-2K<sup>b</sup>) and their associated  $\beta$ 2-microglobulin. Also expressed at similar levels was the subunit 4 of Arp2/3 complex, proteins involved in general cell metabolism and the ras-related proteins Rac-1, -2, -3 together with cdc42.

Several proteins were highly up-regulated in LPS-DCs versus MED-DCs (Fig. 4, Supp. Table 7) including MHC class II molecules

(IA- $\alpha$  representing the highest increase), CD98 heavy chain,  $\beta$ 2-microglobulin and transgelin 2. Intriguingly, NAP-22 was up-regulated in LPS-DCs versus MED-DCs (it was down-regulated in SEA-DCs versus MED-DCs). The integrin CD29 ( $\beta$  subunit of VLA-4) was also up-regulated but not identified between SEA-DCs versus MED-DCs. Down-regulation affected S100A10, moesin and CD44, as also reported for SEA-DCs versus MED-DCs. Additionally, we observed specific decreased expression of CD18, both chains of ATP synthase and the precursor of cathepsin D. Significantly, expression of the secretory protein YM-1, associated with 'alternatively-activated' macrophages, was also highly down-regulated in LPS-DCs versus MED-DCs. Proteins whose expression was unchanged between LPS-DCs versus MED-DCs were GAPDH, several ras-related proteins (including Rab-7), several actin isoforms and galectin-3.

Analysis of protein classification according to GO terms revealed that SEA-DCs had numerous ( $k=17$ ) up-regulated proteins related to cellular metabolism, in particular Nucleoside phosphate metabolic processing (Table 2). These cells also exhibited up-regulated proteins associated with cytoskeleton remodelling specifically a number linked to the formation of cell projections ( $k=13$ ). On the other hand, SEA-DCs had down-regulated proteins associated with responses to chemical stimulation ( $k=10$ ) and GTPase activity ( $k=5$ ). Interestingly, down-regulated protein hits also mapped to those associated with antigen processing, and regulation of the adaptive immune response. In contrast, up-regulated proteins in LPS-DCs mapped clearly to antigen presentation and the MHC protein complex, whilst down-regulated proteins were associated with transmembrane transporter activity, catabolic processes and responses to wound healing (all  $k=3$ ).

#### 3.4. Extrapolated changes between pro-Th2 DCs and pro-Th1 DCs

As we could not directly compare SEA-DCs versus LPS-DCs (due to the constraints of the availability of only 3 iTRAQ labels), we made an extrapolation of the above data in order to

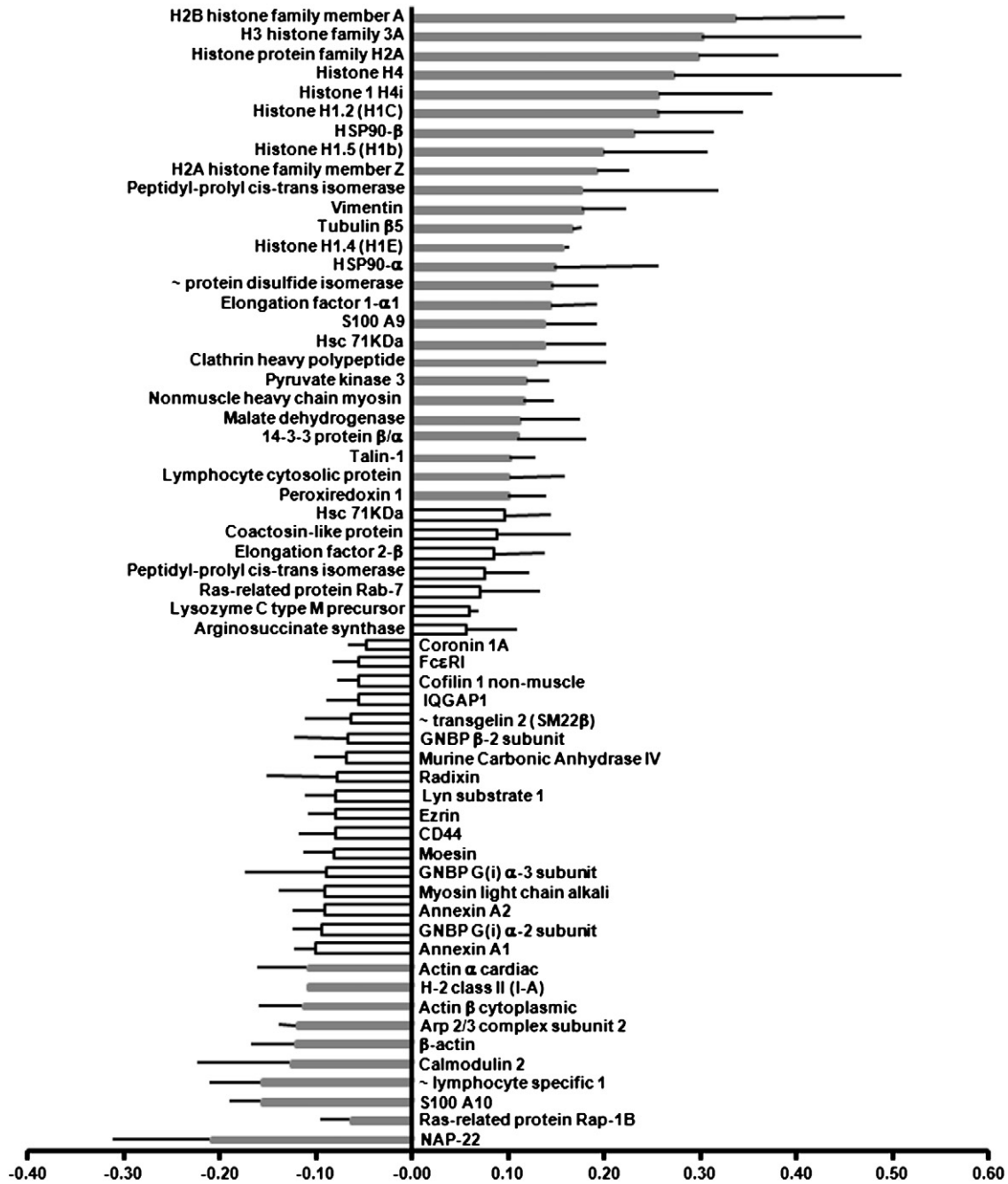


Fig. 3 – Changes on plasma membrane enriched-proteome of SEA-DCs versus MED-DCs. Up- and down-regulation of DC proteins as average  $\log(\text{ratio}) \pm \text{standard error -95\%}$ , obtained from three iTRAQ experiments with criteria for acceptance clearly stated in the [Results](#) section.

compare the protein expression changes between pro-Th2 and pro-Th1 DCs (Table 3). Amongst the proteins identified both in SEA-DCs versus MED-DCs and LPS-DCs versus MED-DCs, moesin, CD18, CD44 and Rab-7 were increased in SEA-DCs compared to LPS-DCs. Conversely, transgelin 2, NAP-22, MHC class II, annexin A1 and A4, β2-microglobulin, S100A10 protein, two actin isoforms and Rap-1b were down-regulated. The only protein similarly expressed by both types of differentially matured DCs was GAPDH.

### 3.5. Validation of protein expression by flow cytometry

Several proteins identified by our iTRAQ studies, for which commercially available antibodies are available, were validated for changes in expression by flow cytometry (Fig. 5). The MHC class II molecules IA-b and IA/IE were both significantly up-regulated on LPS-DCs compared to SEA-DCs and MED-DCs, confirming the changes in expression recorded using iTRAQ. CD29 and S100A10 were also up-regulated on

**Table 1 – Experimental design for iTRAQ. Summary of the six experiments comparing plasma membrane proteins from separate cultures of DCs grown in medium only (MED #1–6), or stimulated with SEA (SEA #1–3) or LPS (LPS # 1–3).**

Expt #	iTRAQ Label		
	114	115	116
1	MED-1	SEA-1	MED-1
2	MED-2	SEA-2	MED-2
3	MED-3	SEA-3	MED-3
4	MED-4	LPS-1	MED-4
5	MED-5	LPS-2	MED-5
6	MED-6	LPS-3	MED-6

LPS-DCs but down-regulated in SEA-DCs, mirroring proteomic analysis. Whilst CD98 was up-regulated in both LPS-DCs and SEA-DCs compared to MED-DCs, CD18 was only up-regulated in SEA-DCs and was more abundant than in LPS-DCs. CD44 was expressed at higher levels in SEA-DCs than LPS-DCs. Several molecules were expressed at similar levels in all three types of DCs (i.e. galectin-3 and CD301b) corroborating proteomic analysis, although the expression of Rab-7 which was elevated at the proteomic level in SEA-DCs compared to LPS-DCs, was not differentially expressed as judged by flow cytometry. Finally, the secretory molecule YM-1, which was significantly down-regulated in LPS-DC versus MED-DC as determined by iTRAQ, was also down-regulated as judged by flow cytometry, although it was detected at slightly lower levels in SEA-DCs compared to MED-DCs.

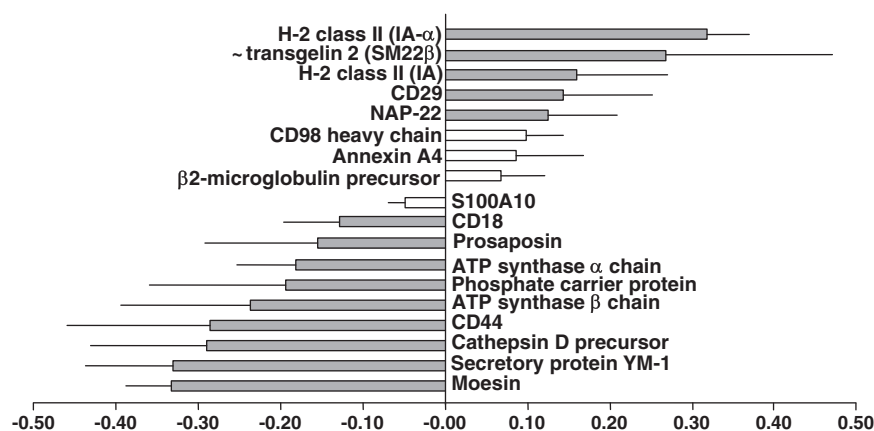
#### 4. Discussion

Over the past decade, proteomics has been utilised to determine differences between various immune cells [23] and in particular to study differences between whole cell extracts from immature DCs and differentially matured DCs with various PAMPs [24]. Mainly, these studies have employed 2-DE-based proteomics

and highlighted changes in the expression of cytoskeletal and cytoplasmic molecules needed for basic cellular functions [12–14,25]. Such an approach favours soluble and abundant cytosolic proteins but compromises the detection of scarcer and detergent-soluble proteins, such as those present in plasma membranes. In the present study, we focussed upon the proteomes of plasma membrane-enriched fractions of differentially matured DCs. Although the fractions isolated were relatively crude, they were clearly enriched in plasma membrane components (~30–50%).

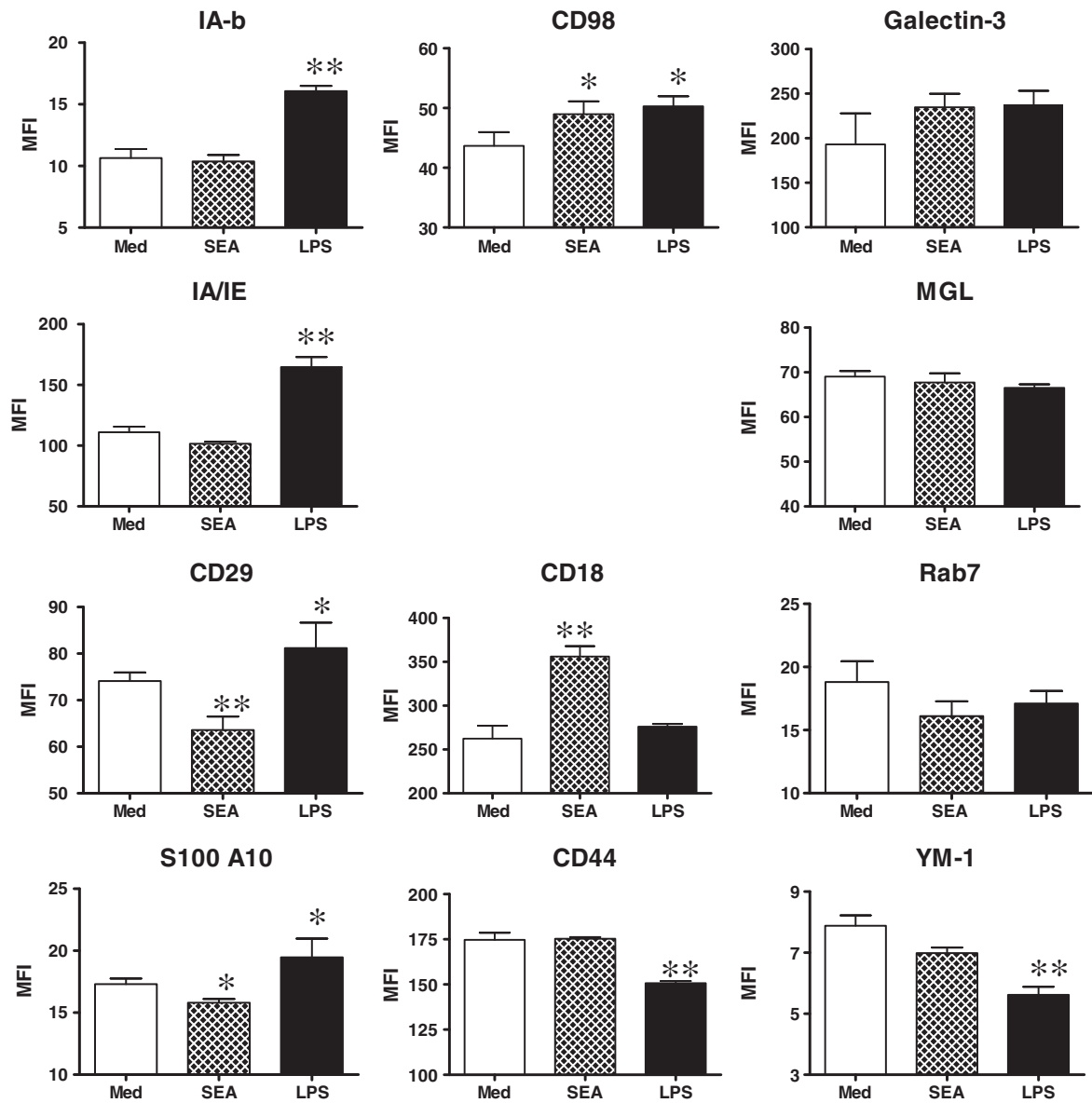
Indeed, LC-MS/MS has several advantages in the analysis of plasma membrane components. First, it requires only low amounts of any given protein for MS analyses, so is suitable for scarcer proteins. Second, digestion of enriched membranes can be performed in the presence of trace amounts of ionic detergents which, along with the use of reducing and alkylating reagents, allows for improved trypsin cleavage leading to higher sequence coverage for a given protein. Furthermore, labelling peptides with iTRAQ tags prior to LC-MS/MS allows the relative amounts of a protein to be determined in different cell groups, which is difficult to be accurately demonstrated by 'label-free' LC-MS/MS.

As revealed by our 'label-free' shotgun proteomics and found in all three types of DCs, the most abundant molecules associated with the plasma membrane including those with a CD prefix, ras-related proteins, MHC molecules and GNBPs. Others included proteins associated with membrane and vesicle trafficking such as the calcium-sensitive annexins [26] and phagocyte-specific S100 calcium binding proteins which are major damage associated molecular patterns [27]. The three differentially matured DCs also all expressed lymphocyte cytosolic protein 1 (also called 65 kDa macrophage protein), LSP-1 (F-actin binding protein) and Fc $\epsilon$ RI, although various immune associated proteins were only found in one or two of the DC types. For instance, immature MED-DCs and pro-Th2 SEA-DCs both expressed CD11b, CD44, CD45 and CD48, while both pro-Th2 DCs and pro-Th1 DCs expressed the heavy chain of CD98 which is involved in activation of naive and memory CD4 and CD8 T cells [28], thereby arguing against the 'default' hypothesis of pro-Th2 DCs.



**Fig. 4 – Changes on plasma membrane enriched-proteome of LPS-DCs versus MED-DCs. Up- and down-regulation of DC proteins as average log (ratio) ± standard error-95%, obtained from three iTRAQ experiments with criteria for acceptance clearly stated in the Results section.**





**Fig. 5 – Validation of protein expression on DCs judged by flow cytometry of antibody labelled cells. Various molecules expressed by DCs were quantified by flow cytometry of antibody labelled cells. Bars are means of MFI values of three separate cultures for each type of differentially matured DC. Significant differences of SEA-DCs and LPS-DCs compared to MED-DCs as \*\* $p < 0.01$  and \* $p < 0.05$ .**

The association of LC-MS/MS with iTRAQ labelling enabled comparison and quantification of changes in the plasma membrane proteome, many of which being confirmed by flow cytometry. As our study compared three biological replicates of each type of DCs, we were able to determine whether changes in the level of protein expression were reproducible. Our study revealed significantly increased levels of MHC class II molecules in pro-Th1 DCs, which were unchanged or decreased in pro-Th2 DCs. Surprisingly, CD29 which mediates the interaction of immature DCs with extracellular matrix components, and thus with a likely role in the retention of DCs in the periphery [29,30], was also highly expressed by LPS-DCs versus immature DCs. Conversely, pro-Th2 SEA-DCs expressed increased levels of proteins like CD18 and CD44.

The greatest decrease of any protein expressed in pro-Th2 SEA-DCs was for NAP-22 (also called BASP-1 or CAP-23) whereas it was greatly up-regulated on pro-Th1 DCs. This molecule was originally identified as a membrane/cytoskeletal protein from rat brain and is a member of a family of motility associated proteins linked to actin reorganisation and neurite development [31,32]. Recently, it has been defined in the context of blocking oncogenic Myc protein-induced cell transformation [33]. Although this molecule was not expected to be expressed by DCs, and has not previously been identified as having demonstrable immune function, its differential distribution between pro-Th1 versus pro-Th2 is intriguing and warrants further investigation. One of the most down-regulated molecules in pro-Th1 LPS-DCs versus MED-DCs was YM-1 as

**Table 2 – Analysis of differentially regulated proteins according to Gene Ontology term. Proteins were analysed for GO term enrichment and presented as up-regulated (top panel), or down-regulated (bottom panel) for SEA-DCs and LPS-DCs. Significance is given as p-value, whilst the number of different proteins identified within a given classification is presented as k.**

SEA-DCs versus MED-DCs				LPS-DCs versus MED-DCs			
Up-regulated				Up-regulated			
GO term		p-value	k	GO term		p-value	k
Cell projection	GO:0042995	1.49E-11	13	Antigen processing and presentation of exogenous peptide antigen	GO:0002478	1.27E-08	3
Purine ribonucleoside triphosphate binding	GO:0035639	1.34E-06	10	MHC protein complex	GO:0042611	1.67E-08	3
Metabolic process	GO:0008152	5.52E-05	17	Peptide antigen binding	GO:0042605	5.42E-06	2
Nucleoside phosphate metabolic process	GO:0006753	6.53E-05	5	Cellular developmental process	GO:0048869	9.77E-04	4
Down-regulated				Down-regulated			
GO term		p-value	k	GO term		p-value	k
Response to chemical stimulus	GO:0042221	8.38E-07	10	Active transmembrane transporter activity	GO:0022804	6.49E-05	3
GTPase activity	GO:0003924	2.02E-07	5	Catabolic process	GO:0009056	2.92E-04	4
Antigen processing and presentation of exogenous peptide antigen via MHC class II	GO:0019886	5.49E-05	2	Response to wounding	GO:0009611	2.12E-04	3
Regulation of adaptive immune response	GO:0002819	6.44E-05	3				

determined from the proteomic data and verified by flow cytometry. This molecule has been associated with the development of ‘alternatively-activated’ macrophages and the recruitment of eosinophils [34], particularly after multiple exposure to schistosome cercariae [35] and other helminth infections which induce Th2-type immune responses [36,37].

A number of other changes in protein expression may help explain the differential DC phenotype/function. For example, SEA-DCs may be more efficient at antigen uptake and processing than immature MED-DCs as several chaperone proteins and enzymes involved in protein folding were up-regulated in pro-Th2 DCs compared to immature DCs (i.e. HSP-90 $\alpha$ , two isoforms of hsc71, clathrin and Rab-7) although flow cytometry analysis revealed intracellular Rab-7 expression was similar between pro-Th2 DCs and pro-Th1 DCs. Another molecule up-regulated in SEA-DCs was protein disulfide isomerase which not only participates in the editing of MHC class I peptide repertoire [38] but also inhibits the transcriptional activity of NF- $\kappa$ B as a downstream signal of IL-10 [39]. In contrast, most of the proteins reported above were not identified in mature pro-Th1 DCs that have lost their capacity to take up and process antigen, however they are poised to present antigens, as suggested by our GO term enrichment analysis. Moreover, many enzymes and transporters related to general cell metabolism (i.e. malate dehydrogenase, pyruvate kinase 3, arginosuccinate synthase, fructose-bisphosphate aldolase A, enolase 1 and two chains of transporting ATPase Na<sup>+</sup>/K<sup>+</sup>) were up-regulated, or expressed at similar levels, between SEA-DCs versus MED-DCs, while their expression was mostly decreased in LPS-DCs versus MED-DCs. This suggests that MED-DCs and pro-Th2 DCs are more metabolically active compared to pro-Th1 LPS-DCs that have matured into an ‘end state’. GO term enrichment analysis confirmed that pro-Th2 DCs up-regulated proteins associated

with nucleotide metabolism, whilst at the same time had down-regulated responses to chemical stimulus and regulation of the adaptive immune response.

Many of the proteomic changes in pro-Th2 SEA-DCs affected cytoskeletal proteins confirming previous studies using 2-DE-based proteomics [14]. For example, vimentin, tubulin  $\beta$ 5, non-muscle myosin, talin and coactosin-like protein were all up-regulated in SEA-DCs, whereas proteins associated to actin such as Arp2/3 with a role in membrane trafficking [40] and ezrin-radixin-moesin complex that links cytoskeletal components with plasma membrane proteins [41] were down-regulated versus MED-DCs. Moreover, in LPS-DCs few cytoskeletal proteins were identified, alternatively they were down-regulated or expressed at similar levels compared to MED-DCs (i.e. actin isoforms, destrin and moesin). Finally, our study also reveals that SEA-DCs contain twice as many nuclear proteins than LPS-DCs or immature DCs. These proteins include initiation and elongation factors, several histones, signalling proteins (MAPK3, Erk-1) indicating that SEA may be a potent stimulus for DC proliferation.

Our study using iTRAQ labelling combined with LC-MS-MS has provided a valuable appraisal of the repertoire of proteins expressed by differentially matured DCs particularly within the plasma membrane. This offered new insights into how differentially matured DCs may function to promote Th1 or Th2 polarisation of the adaptive immune response. A caveat to studies using bone-marrow derived DCs is that whilst the cells have a high level of purity and are of immature/neutral status, they may not be representative of DCs from tissues *in vivo* which are subject to many local factors, such as the local cytokine environment [35], which influence their maturation as pro-Th1 versus pro-Th2 DCs. However, we believe that our combined approach has the potential to reveal the identity of further molecules involved in DC differentiation,

**Table 3 – Extrapolated proteins changes between SEA-DCs and LPS-DCs. List of proteins present both in Table 1 (SEA-DCs versus MED-DCs) and Table 2 (LPS-DCs versus MED-DCs) with their respective average log(ratio): (1) and (2). The extrapolated average log(ratio) for pro-Th2 SEA-DCs compared to pro-Th1 LPS-DCs equals (1)–(2). Shaded rows highlighted major changes at the most stringent threshold: log(ratio) values of  $\pm 0.1$ . The table presents the proteins from the most up-regulated at the top to the most down-regulated at the bottom. The accession numbers were the ones given by the Mascot search against the CDS mouse database.**

Accession no.	SEA versus MED	LPS versus MED	SEA versus LPS	Protein hits
<i>Up-regulation</i>				
/:spt P26041	–0.079	–0.334	0.255	Moesin (Membrane-organising extension spike protein)
/:trm Q80X37	–0.078	–0.285	0.207	CD44
/:pir S04847	–0.021	–0.129	0.108	CD18, Integrin $\beta$ -2 precursor (LFA-1/CR3/P150,95 $\beta$ -subunit)
/:dbj BAB23738.1	0.072	–0.037	0.108	Ras-related protein Rab-7
<i>No change</i>				
/:spt P16858	0.017	–0.023	0.040	Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)
<i>Down-regulation</i>				
/:rf NP_077777.1	–0.062	0.021	–0.082	Ras-related protein Rap-1B
/:trm Q9CCK3	–0.107	–0.017	–0.090	Actin, $\alpha$ , cardiac
/:rf NP_031419.1	–0.112	–0.014	–0.098	Actin, $\beta$ , cytoplasmic
/:spt P08207	–0.155	–0.048	–0.107	S100 A10, calpactin I light chain
/:pdb 1BZ9_B	–0.040	0.067	–0.107	$\beta$ -2-microglobulin precursor
/:trm Q7TMN7	–0.038	0.086	–0.124	Annexin A4
/:spt P10107	–0.098	0.031	–0.128	Annexin A1, calpactin II
/:trm Q9TP17	–0.107	0.159	–0.266	H-2 class II histocompatibility antigen, A $\beta$ chain precursor (I-A)
/:trm Q91XV3	–0.208	0.078	–0.286	22 kDa neuronal tissue-enriched acidic protein, NAP-22, BASP-1
/:trm Q91VU2	–0.062	0.269	–0.331	~ transgelin 2 (SM22 $\beta$ )

especially if increased peptide sample fractionation coupled to higher sensitivity mass spectrometry instrumentation is employed. Accordingly, Segura et al. [42] recently reported an extensive comparative list of membrane proteins from immunopurified DCs using 2D liquid peptide fractionation followed by MS/MS in an Orbitrap instrument. However, their analysis was performed using only a single biological sample and protein expression was inferred by spectral counting. Their study also did not examine differential maturation after stimulation of DCs with different PAMPs.

In conclusion, our study and resulting analysis of the proteomic data support the view of a 'limited maturation' of pro-Th2 DCs compared to pro-Th1 DCs. We also report evidence of the expression of a restricted number of membrane proteins which may be unexpected 'signatures' of pro-Th2 DCs and consequently warrant further investigation.

Supplementary materials related to this article can be found online at [doi:10.1016/j.jprot.2011.10.010](https://doi.org/10.1016/j.jprot.2011.10.010).

## Acknowledgements

S.F.B and W.C.B were supported by the UK Biotechnology and Biological Sciences Research Council (grants # BBS/B/08531 and # BB/C516328 respectively); A.P.M and J.D.T received funding from the Wellcome Trust (grants # 071762 and # 072255) and The European Union (INCO-CT-2006-032405). D.E.S. was funded by COLFUTURO and the Departamento Administrativo de Ciencia, Tecnología e Innovación de la República de Colombia (COLCIENCIAS). P.C.C. was in receipt of a BBSRC PhD studentship. We would like to thank Ann Bamford for assistance during the course of this project.

## REFERENCES

- [1] Heath WR, Carbone FR. Dendritic cell subsets in primary and secondary T cell responses at body surfaces. *Nat Immunol* 2009;10:1237–44.
- [2] Colonna M, Pulendran B, Iwasaki A. Dendritic cells at the host–pathogen interface. *Nat Immunol* 2006;7:117–20.
- [3] MacDonald AS, Straw AD, Bauman B, Pearce EJ. CD8- dendritic cell activation status plays an integral role in influencing Th2 response development. *J Immunol* 2001;167:1982–8.
- [4] Balic A, Harcus Y, Holland MJ, Maizels RM. Selective maturation of dendritic cells by *Nippostrongylus brasiliensis*-secreted proteins drives Th2 immune responses. *Eur J Immunol* 2004;34:3047–59.
- [5] Cervi L, MacDonald AS, Kane C, Dzierszynski F, Pearce EJ. Cutting edge: dendritic cells copulsed with microbial and helminth antigens undergo modified maturation, segregate the antigens to distinct intracellular compartments, and concurrently induce microbe-specific Th1 and helminth-specific Th2 responses. *J Immunol* 2004;172:2016–20.
- [6] Perona-Wright G, Jenkins SJ, MacDonald AS. Dendritic cell activation and function in response to *Schistosoma mansoni*. *Int J Parasitol* 2006;36:711–21.
- [7] Jenkins SJ, Mountford AP. Dendritic cells activated with products released by schistosome larvae drive Th2-type immune responses, which can be inhibited by manipulation of CD40 costimulation. *Infect Immun* 2005;73:395–402.
- [8] Kelsall BL, Biron CA, Sharma O, Kaye PM. Dendritic cells at the host–pathogen interface. *Nat Immunol* 2002;3:699–702.
- [9] Mac Donald AS, Maizels RM. Alarming dendritic cells for Th2 induction. *J Exp Med* 2008;205:13–7.
- [10] Le Naour F, Hohenkirk L, Grolleau A, Misek DE, Lescure P, Geiger JD, et al. Profiling changes in gene expression during differentiation and maturation of monocyte-derived dendritic cells using both oligonucleotide microarrays and proteomics. *J Biol Chem* 2001;276:17920–31.

- [11] Richards J, Le Naour F, Hanash S, Beretta L. Integrated genomic and proteomic analysis of signaling pathways in dendritic cell differentiation and maturation. *Ann N Y Acad Sci* 2002;975:91–100.
- [12] Pereira SR, Faça VM, Gomes GG, Chammas R, Fontes AM, Covas DT, et al. Changes in the proteomic profile during differentiation and maturation of human monocyte-derived dendritic cells stimulated with granulocyte macrophage colony stimulating factor/interleukin-4 and lipopolysaccharide. *Proteomics* 2005;5:1186–98.
- [13] Rivollier A, Perrin-Cocon L, Luche S, Diemer H, Strub JM, Hanau D, et al. High expression of antioxidant proteins in dendritic cells: possible implications in atherosclerosis. *Mol Cell Proteomics* 2006;5:726–36.
- [14] Ferret-Bernard S, Curwen RS, Mountford AP. Proteomic profiling reveals that Th2-inducing dendritic cells stimulated with helminth antigens have a 'limited maturation' phenotype. *Proteomics* 2008;8:980–93.
- [15] Gundacker NC, Haudek VJ, Wimmer H, Slany A, Griss J, Bochlov V, et al. Cytoplasmic proteome and secretome profiles of differently stimulated human dendritic cells. *J Proteome Res* 2009;8:2799–811.
- [16] Jenkins SJ, Hewitson JP, Ferret-Bernard S, Mountford AP. Schistosome larvae stimulate macrophage cytokine production through TLR4-dependent and -independent pathways. *Int Immunol* 2005;17:1409–18.
- [17] Paveley RA, Aynsley SA, Cook PC, Turner JD, Mountford AP. Fluorescent imaging of antigen released by a skin-invading helminth reveals differential uptake and activation profiles by antigen presenting cells. *PLoS Negl Trop Dis* 2009;3:e528.
- [18] Zieske LR. A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J Exp Bot* 2006;57:1501–8.
- [19] Gan CS, Chong PK, Pham TK, Wright PC. Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J Proteome Res* 2007;6:821–7.
- [20] Song X, Bandow J, Sherman J, Baker JD, Brown PW, McDowell MT, et al. iTRAQ experimental design for plasma biomarker discovery. *J Proteome Res* 2008;7:2952–8.
- [21] Spooncer E, Brouard N, Nilsson SK, Williams B, Liu MC, Unwin RD, et al. Developmental fate determination and marker discovery in hematopoietic stem cell biology using proteomic fingerprinting. *Mol Cell Proteomics* 2008;7:573–81.
- [22] Pierce A, Unwin RD, Evans CA, Griffiths S, Carney L, Zhang L, et al. Eight-channel iTRAQ enables comparison of the activity of six leukemogenic tyrosine kinases. *Mol Cell Proteomics* 2008;7:853–63.
- [23] Wang X, Zhao H, Andersson R. Proteomics and leukocytes: an approach to understanding potential molecular mechanisms of inflammatory responses. *J Proteome Res* 2004;3:921–9.
- [24] Ferreira GB, Mathieu C, Oberbergh L. Understanding dendritic cell biology and its role in immunological disorders through proteomic profiling. *Proteomics Clin Appl* 2010;4:190–203.
- [25] Slany A, Haudek VJ, Gundacker NC, Griss J, Mohr T, Wimmer H, et al. Introducing a new parameter for quality control of proteome profiles: consideration of commonly expressed proteins. *Electrophoresis* 2009;30:1306–28.
- [26] Gerke V, Creutz CE, Moss SE. Annexins: linking Ca<sup>2+</sup> signalling to membrane dynamics. *Nat Rev Mol Cell Biol* 2005;6:449–61.
- [27] Foell D, Wittkowski H, Vogl T, Roth J. S100 proteins expressed in phagocytes: a novel group of damage-associated molecular pattern molecules. *J Leukoc Biol* 2007;81:28–37.
- [28] Woodhead VE, Stonehouse TJ, Binks MH, Speidel K, Fox DA, Gaya A, et al. Novel molecular mechanisms of dendritic cell-induced T cell activation. *Int Immunol* 2000;12:1051–61.
- [29] Randolph GJ, Angeli V, Swartz MA. Dendritic-cell trafficking to lymph nodes through lymphatic vessels. *Nat Rev Immunol* 2005;5:617–28.
- [30] Lehner M, Stilper A, Morhart P, Holter W. Plasticity of dendritic cell function in response to prostaglandin E<sub>2</sub> (PGE<sub>2</sub>) and interferon-gamma (IFN-gamma). *J Leukoc Biol* 2008;83:883–93.
- [31] Epand RF, Sayer BG, Epand RM. Induction of raft-like domains by a myristoylated NAP-22 peptide and its Tyr mutant. *FEBS J* 2005;272:1792–803.
- [32] Korshunova I, Caroni P, Kolkova K, Berezin V, Bock E, Walmod PS. Characterization of BASP1-mediated neurite outgrowth. *J Neurosci Res* 2008;86:2201–13.
- [33] Hartl M, Nist A, Khan MI, Valovka T, Bister K. Inhibition of Myc-induced cell transformation by brain acid-soluble protein 1 (BASP1). *Proc Natl Acad Sci U S A* 2009;106:5604–9.
- [34] Martinez FO, Helming L, Gordon S. Alternative activation of macrophages: an immunologic functional perspective. *Annu Rev Immunol* 2009;27:451–83.
- [35] Cook PC, Aynsley SA, Turner JD, Jenkins GR, Van Rooijen N, Leeto M, et al. Multiple helminth infection of the skin causes lymphocyte hypo-responsiveness mediated by Th2 conditioning of dermal myeloid cells. *PLoS Pathog* 2011;7:e1001323.
- [36] Prieto-Lafuente L, Gregory WF, Allen JE, Maizels RM. MIF homologues from a filarial nematode parasite synergize with IL-4 to induce alternative activation of host macrophages. *J Leukoc Biol* 2009;85:844–54.
- [37] Reyes JL, Terrazas CA, Vera-Arias L, Terrazas LI. Differential response of antigen presenting cells from susceptible and resistant strains of mice to *Taenia crassiceps* infection. *Infect Genet Evol* 2009;9:1115–27.
- [38] Park B, Lee S, Kim E, Cho K, Riddell SR, Cho S, et al. Redox regulation facilitates optimal peptide selection by MHC class I during antigen processing. *Cell* 2006;127:369–82.
- [39] Higuchi T, Watanabe Y, Waga I. Protein disulfide isomerase suppresses the transcriptional activity of NF-kappaB. *Biochem Biophys Res Commun* 2004;318:46–52.
- [40] Takenawa T, Suetsugu S. The WASP-WAVE protein network: connecting the membrane to the cytoskeleton. *Nat Rev Mol Cell Biol* 2007;8:37–48.
- [41] Tsukita S, Yonemura S. ERM (ezrin/radixin/moesin) family: from cytoskeleton to signal transduction. *Curr Opin Cell Biol* 1997;9:70–5.
- [42] Segura E, Kapp E, Gupta N, Wong J, Lim J, Ji H, et al. Differential expression of pathogen-recognition molecules between dendritic cell subsets revealed by plasma membrane proteomic analysis. *Mol Immunol* 2010;47:1765–73.

## Appendix 5B

Declaration from Stephanie Ferret-Bernard attesting to my contribution to paper 5.





UNIVERSITY  
*of York*

### Declaration of reviewee's contribution

#### **Paper**

Ferret-Bernard, S., Castro-Borges, W., **DOWLE, A.A.**, Cook, P.C., Turner, J.D., MacDonald, A.S., Thomas, J.R. & Mountford, A.P. (2012) Plasma membrane proteomes of differentially matured dendritic cells identified by LC-MS/MS combined with iTRAQ labelling. *J. Proteomics* **75**, 938-948.

#### **Candidate's contribution**


The submitting candidate performed quantitative mass spectrometry-based proteomic analyses for membrane enriched protein mixtures, which were prepared by Stéphanie Ferret-Bernard. The candidate was responsible for all aspects of LC-MALDI-MS/MS analysis including: proteomic method design, data acquisition, database searching and iTRAQ quantification. The work was provided as fee for service through the Bioscience Technology Facility, Department of Biology, University of York. Contributions to data analysis, production of figures and input into writing the paper were provided on a collaborative basis.

#### **Declaration**

I attest that the summation above is a true reflection of the candidate's (Adam A. Dowle) contribution to the paper.

#### **Candidate**

Name: Adam Dowle

Signature: 

Date: 7<sup>th</sup> June 2016

#### **First author**

Name: Stéphanie Ferret-Bernard

Signature: 

Date: June 7<sup>th</sup>, 2016



## Appendix 5C

Declaration from Jerry Thomas attesting my contribution to paper 5.



UNIVERSITY  
*of York*

## **Declaration of reviewee's contribution**

### **Paper**

Ferret-Bernard, S., Castro-Borges, W., **DOWLE, A.A.**, Cook, P.C., Turner, J.D., MacDonald, A.S., Thomas, J.R. & Mountford, A.P. (2012) Plasma membrane proteomes of differentially matured dendritic cells identified by LC-MS/MS combined with iTRAQ labelling. *J. Proteomics* **75**, 938-948.

### **Candidate's contribution**


The submitting candidate performed quantitative mass spectrometry-based proteomic analyses for membrane enriched protein mixtures, which were prepared Stéphanie Ferret-Bernard. The candidate was responsible for all aspects of LC-MALDI-MS/MS analysis including: protein digestion, iTRAQ labelling, method design, data acquisition, database searching and iTRAQ quantification. The work was provided as fee for service through the Bioscience Technology Facility, Department of Biology, University of York. Contributions to data analysis, production of figures and input into writing the paper were provided on a collaborative basis.

### **Declaration**

I attest that the summation above is a true reflection of the candidate's (Adam A. Dowle) contribution to the paper.

### **Candidate**

**Name:** Adam Dowle

**Signature:** 

**Date:** 17<sup>th</sup> May 2016

### **Co-author**

**Name:** Jerry Thomas

**Signature:** 

**Date:** 17-5-16

## Appendix 6A

Submitted paper 6 – *S*-Adenosyl-*S*-carboxymethyl-L-homocysteine: a novel cofactor found in the putative tRNA-modifying enzyme CmoA.

# S-Adenosyl-S-carboxymethyl-L-homocysteine: a novel cofactor found in the putative tRNA-modifying enzyme CmoA

Robert T. Byrne,<sup>a,‡</sup> Fiona Whelan,<sup>a,§</sup> Pierre Aller,<sup>b</sup> Louise E. Bird,<sup>c,d</sup> Adam Dowle,<sup>e</sup> Carina M. C. Lobley,<sup>b</sup> Yamini Reddivari,<sup>c,d</sup> Joanne E. Nettleship,<sup>c,d</sup> Raymond J. Owens,<sup>c,d</sup> Alfred A. Antson<sup>a</sup> and David G. Waterman<sup>f,\*</sup>

<sup>a</sup>York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington YO10 5DD, England, <sup>b</sup>Diamond Light Source Ltd, Diamond House, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, England, <sup>c</sup>OPPF-UK, Research Complex at Harwell, R92 Rutherford Appleton Laboratory, Didcot, Oxfordshire OX11 0FA, England, <sup>d</sup>Division of Structural Biology, Oxford University, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, England, <sup>e</sup>Bioscience Technology Facility, Department of Biology, University of York, Heslington YO10 5DD, England, and <sup>f</sup>STFC, Rutherford Appleton Laboratory, Didcot, Oxfordshire OX11 0FA, England

<sup>‡</sup> Current address: Section of Structural Biology, The Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, England.

<sup>§</sup> Current address: Department of Biology, University of York, Heslington YO10 5DD, England.

Correspondence e-mail:  
david.waterman@stfc.ac.uk

Uridine at position 34 of bacterial transfer RNAs is commonly modified to uridine-5-oxyacetic acid (cmo<sup>5</sup>U) to increase the decoding capacity. The protein CmoA is involved in the formation of cmo<sup>5</sup>U and was annotated as an S-adenosyl-L-methionine-dependent (SAM-dependent) methyltransferase on the basis of its sequence homology to other SAM-containing enzymes. However, both the crystal structure of *Escherichia coli* CmoA at 1.73 Å resolution and mass spectrometry demonstrate that it contains a novel cofactor, S-adenosyl-S-carboxymethyl-L-homocysteine (SCM-SAH), in which the donor methyl group is substituted by a carboxymethyl group. The carboxyl moiety forms a salt-bridge interaction with Arg199 that is conserved in a large group of CmoA-related proteins but is not conserved in other SAM-containing enzymes. This raises the possibility that a number of enzymes that have previously been annotated as SAM-dependent are in fact SCM-SAH-dependent. Indeed, inspection of electron density for one such enzyme with known X-ray structure, PDB entry 1im8, suggests that the active site contains SCM-SAH and not SAM.

## 1. Introduction

Following transcription by RNA polymerase, a transfer RNA transcript is converted into a mature tRNA through processing and nucleoside modification. Processing in bacteria entails the removal of the 5' leader and 3' tail sequences and, if necessary, the re-synthesis of the 3' CCA extension required for aminoacylation. Nucleoside modification is part of the maturation process that extends the physicochemical properties of tRNA by providing a wider complement of nucleosides than the canonical four introduced by RNA polymerase during transcription. Approximately 10% of nucleosides in a typical tRNA are modified, which corresponds to around seven modified nucleosides per tRNA (Jühling *et al.*, 2009). Of the 109 modified nucleosides of RNA, 93 are found in tRNA, making it the most diversely modified RNA (Cantara *et al.*, 2011). Modified nucleosides confer diverse properties upon tRNA, but they generally fine-tune the structure and stability such that it is optimized for functioning in processes such as aminoacylation and translation (Motorin & Helm, 2010).

The anticodon stem loop (ASL) is one of the most heavily modified regions of tRNA and is frequently modified at positions 34 (the wobble position) and 37 (immediately after the anticodon) (Jühling *et al.*, 2009). Modification of the ASL reduces its flexibility and shifts its structure towards the U-turn conformation; this has been seen in crystal structures of mature tRNA, solution structures of isolated modified ASL domains and structures of the 70S ribosome in complex with tRNA and mRNA (Selmer *et al.*, 2006; Shi & Moore, 2000;

Received 26 January 2013  
Accepted 20 February 2013

PDB Reference: CmoA, 4iwn

Vendeix *et al.*, 2008). Because modifications impose this conformation upon the ASL, the entropic penalty otherwise associated with remodelling during translation is avoided (Agris, 2008). A number of nucleoside modifications at positions 34 and 37 also expand the decoding capacity beyond that predicted by the 'Wobble hypothesis' (Agris *et al.*, 2007). In bacteria, the 5-oxyuridine derivatives commonly found at position 34 fall into this category because they allow the tRNA to decode not only codons ending in A and G (as predicted by the Wobble hypothesis), but also U and C, so that a single tRNA can consequently decode all four codons within a single box of the genetic code (Nasvall *et al.*, 2004). The derivative uridine-5-oxyacetic acid ( $\text{cmo}^5\text{U}$ ) has so far been found in  $\text{tRNA}^{\text{Ala}}$ ,  $\text{tRNA}^{\text{Pro}}$ ,  $\text{tRNA}^{\text{Ser}}$  and  $\text{tRNA}^{\text{Val}}$  (Jühling *et al.*, 2009). These modifications are therefore beneficial because they not only reduce the entropic penalty associated with translation but also reduce the number of different tRNAs that a cell must produce for decoding.

While the function of  $\text{cmo}^5\text{U}$  within the context of a tRNA is well characterized, its synthesis has received relatively little attention. The proposed biosynthetic pathway involves the stepwise modification of U to  $\text{cmo}^5\text{U}$  through the intermediates 5-hydroxyuridine ( $\text{ho}^5\text{U}$ ) and 5-methoxyuridine ( $\text{mo}^5\text{U}$ ) (Fig. 1). In keeping with the majority of other nucleoside-modification pathways, it is likely that the substrate base is modified while remaining part of the tRNA and is not simply exchanged for a pre-modified base in a transglycosylation reaction. Mutations in the *cmoB* and *cmoA* genes result in accumulation of  $\text{ho}^5\text{U}$  and  $\text{mo}^5\text{U}$ , respectively, indicating that CmoB is involved in the modification of  $\text{ho}^5\text{U}$  to  $\text{mo}^5\text{U}$  and that CmoA is involved in the modification of  $\text{mo}^5\text{U}$  to  $\text{cmo}^5\text{U}$  (Nasvall *et al.*, 2004). Both CmoA and CmoB contain S-adenosylmethionine (SAM) binding motifs, hinting that they are methyltransferases, but only one of the two C atoms in the side chain of  $\text{cmo}^5\text{U}$  is derived from SAM (Hagervall *et al.*, 1990). Furthermore, the synthesis of 5-oxyuridine derivatives is also dependent upon chorismic acid, although the nature of this dependency has not yet been determined (Hagervall *et al.*, 1990; Nasvall *et al.*, 2004). This suggests that the modification pathway has not been fully elucidated or that parts of the  $\text{cmo}^5\text{U}$  side chain may be derived from other metabolites.

We decided to investigate the functions of CmoA and CmoB in detail by X-ray crystallography to help fill the gaps in our understanding of  $\text{cmo}^5\text{U}$  biosynthesis. Although CmoA was expected to be similar to the known structure of its homologue *Haemophilus influenzae* YecO, there is no sufficiently high-quality model of CmoB in the Protein Data Bank to form structure-based hypotheses about the function of the system. We hope to rectify this by

providing high-quality structures of both proteins from the same target organism, allowing detailed models of the enzymatic pathway and RNA interactions to be constructed and further tested. Here, we report the structure of *Escherichia coli* CmoA, which unexpectedly reveals a cofactor that, to the best of our knowledge, has not been observed before.

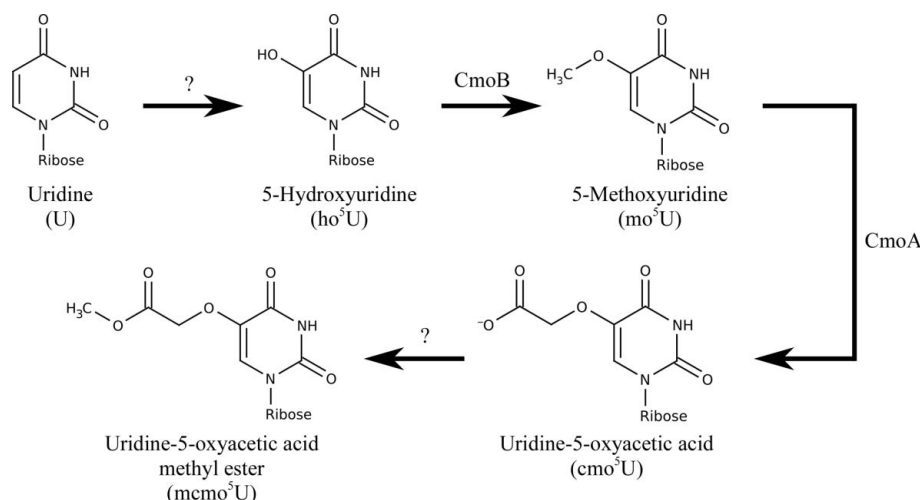
## 2. Materials and methods

### 2.1. Molecular biology and protein production

The coding sequence *cmoA* was amplified from OmniMax II cells (Invitrogen) and cloned into the vector pOPINF using the In-Fusion method to generate the construct OPPF7299 (Berrow *et al.*, 2007). The final construct has an R100H point mutation with respect to the deposited sequence of CmoA from *E. coli* K-12 strain MG1655 (UniProt P76290), which may be either a PCR mutation or a genuine difference in this strain. Sequence analysis shows that amino acids with diverse properties are found at this position in other UniProt CmoA-family members. Once the X-ray structure had been determined, it became clear that this residue is located on the surface of the protein on the side opposite to the dimer interface and at a distance of  $\sim 15$  Å from the nearest atom of the SCM-SAH cofactor. *E. coli* Rosetta pLysS (DE3) cells were transformed with the resulting vector and grown in Overnight Express Instant TB medium (Merck). The cells were incubated at 310 K until an  $\text{OD}_{600\text{ nm}}$  of 0.6 was attained, at which point the temperature was reduced to 298 K and the cells were grown for a further 20 h. The cells were then harvested by centrifugation and stored at 193 K.

### 2.2. Protein purification

The cells were resuspended in lysis buffer [500 mM NaCl, 50 mM Tris pH 7.5, 30 mM imidazole, 0.2% (v/v) Tween], lysed



**Figure 1**

The proposed modification pathway of 5-oxyuridine derivatives. CmoA has been implicated in the modification of  $\text{mo}^5\text{U}$  to  $\text{cmo}^5\text{U}$  (Nasvall *et al.*, 2004), although this reaction involves more than the addition of a single methyl group, indicating that either additional enzymes and/or cofactors are involved. No enzymes involved in the conversion of U to  $\text{ho}^5\text{U}$  or  $\text{cmo}^5\text{U}$  to  $\text{mcmo}^5\text{U}$  have been identified.

using a Basic Z cell disruptor (Constant Systems) and clarified by centrifugation. The supernatant was loaded onto a 1 ml HisTrap FF column (GE Healthcare) equilibrated with wash buffer (500 mM NaCl, 50 mM Tris pH 7.5, 30 mM imidazole) and bound protein was eluted with elution buffer (500 mM NaCl, 50 mM Tris pH 7.5, 500 mM imidazole). Fractions containing CmoA were concentrated and loaded onto a Superdex 200 HiLoad 16/60 column (GE Healthcare) equilibrated with gel-filtration buffer (200 mM NaCl, 20 mM Tris pH 7.5). Fractions containing CmoA were pooled and the N-terminal hexahistidine tag was removed by digesting the protein with rhinovirus 3C protease. The mixture was then reverse-purified by performing an additional round of Ni<sup>2+</sup>-affinity chromatography as described above and collecting the flowthrough. This protein was then buffer-exchanged into gel-filtration buffer and concentrated to 20 mg ml<sup>-1</sup> for crystallization.

### 2.3. Size-exclusion chromatography coupled with static light scattering (SEC-SLS) analysis

The oligomeric state of CmoA in solution was analysed using size-exclusion chromatography with a Superdex 200 column followed by light scattering using a Viscotek Tetra Array Detector measuring refractive index, right-angle light scattering and absorbance at 280 nm. A 100 µl sample at 0.77 mg ml<sup>-1</sup> was applied onto the size-exclusion column and was observed to correspond to dimeric CmoA.

### 2.4. Crystallization

Sitting-drop experiments were performed in a CrystalQuick crystallization plate (Greiner Bio-One) at 294 K. 100 nl CmoA solution was mixed with 100 nl crystallization solution and equilibrated against a reservoir of 200 µl crystallization solution. Crystals were grown in condition E8 of the Morpheus crystallization screen (Molecular Dimensions): 0.3 M diethylene glycol, 0.3 M triethylene glycol, 0.3 M tetraethylene glycol, 0.3 M pentaethylene glycol, 0.1 M MOPS/HEPES-Na pH 7.5, 12.5% (w/v) PEG 1000, 12.5% (w/v) PEG 3350, 12.5% (w/v) MPD (Gorrec, 2009). Crystals grew after 5 h and were flash-cooled in liquid nitrogen without any additional cryoprotection.

### 2.5. Crystallography

Data were collected on beamline I04 of Diamond Light Source, Didcot, England and were processed with *xia2* (Winter, 2010, Evans, 2011; Leslie, 2006; Sauter *et al.*, 2004; Zhang *et al.*, 2006). The structure was determined by molecular replacement using the structure of *H. influenzae* YecO (PDB entry 1im8; chain B; Lim *et al.*, 2001) with both the SAM cofactor and solvent molecules removed and the programs *CHAINSAW* (Stein, 2008) and *MOLREP* (Vagin & Teplyakov, 2010) as implemented in the *MrBUMP* pipeline (Keegan & Winn, 2007). The molecular-replacement solution contained two molecules of CmoA and had initial  $R_{\text{work}}/R_{\text{free}}$  values of 45.9/45.5%. The model was then improved through alternate cycles of manual rebuilding using *Coot* (Emsley *et al.*,

**Table 1**

Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data collection	
Wavelength (Å)	0.9795
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2
Unit-cell parameters (Å)	<i>a</i> = 77.12, <i>b</i> = 91.38, <i>c</i> = 70.64
Resolution (Å)	55.9–1.73 (1.78–1.73)
No. of reflections	
Total	261435 (19274)
Unique	52750 (3856)
Completeness (%)	99.9 (100.0)
Multiplicity	5.0 (5.0)
$\langle I/\sigma(I) \rangle$	15.1 (2.0)
$R_{\text{merge}}^{\dagger}$	0.056 (0.648)
$R_{\text{p.i.m.}}^{\ddagger}$	0.034 (0.366)
Wilson <i>B</i> factor (Å <sup>2</sup> )	21.6
Refinement	
Resolution (Å)	55.9–1.73 (1.78–1.73)
No. of reflections	
Working	50019 (3640)
Free	2687 (212)
No. of atoms	
Total	3942
Protein	3593
SCM-SAH	60
Solvent	289
$R_{\text{work}}^{\S}$ (%)	19.6 (29.6)
$R_{\text{free}}^{\S}$ (%)	23.1 (31.9)
Mean <i>B</i> factor (Å <sup>2</sup> )	
Overall	31.1
Protein	30.7
SCM-SAH	29.6
Solvent	36.0
Geometry	
R.m.s.d., bond lengths (Å)	0.014
R.m.s.d., bond angles (°)	1.7
Ramachandran plot (%)	
Favoured	98.2
Allowed	1.8

<sup>†</sup>  $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ . <sup>‡</sup>  $R_{\text{p.i.m.}} = \frac{\sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ . <sup>§</sup>  $R_{\text{work}}$  and  $R_{\text{free}}$  =  $\frac{\sum_{hkl} |F_{\text{obs}} - F_{\text{calc}}|}{\sum_{hkl} |F_{\text{obs}}|}$ .  $R_{\text{free}}$  was calculated from a randomly chosen set of reflections (5% of the total) excluded from the  $R_{\text{work}}$  set used for refinement.

2010) and restrained refinement with *REFMAC5* (Murshudov *et al.*, 2011) using an isotropic *B* factor for each atom and one TLS group per chain (Winn *et al.*, 2001). A restraints file for *S*-adenosyl-*S*-carboxymethyl-L-homocysteine was created using the *PRODRG2* server (Schüttelkopf & van Aalten, 2004).

The final model contains two molecules of CmoA (residues 19–247 in chain *A* and residues 20–244 in chain *B*), two molecules of *S*-adenosyl-*S*-carboxymethyl-L-homocysteine, two molecules of MPD and 273 water molecules. Model statistics are provided in Table 1. The Ramachandran plot of the model was calculated with *RAMPAGE* (Lovell *et al.*, 2003) and the figures were created with *CCP4mg* (McNicholas *et al.*, 2011) and the *PoseView* server (Stierand *et al.*, 2006). The coordinates and structure factors have been deposited in the Protein Data Bank with accession code 4iwn.

### 2.6. Mass spectrometry

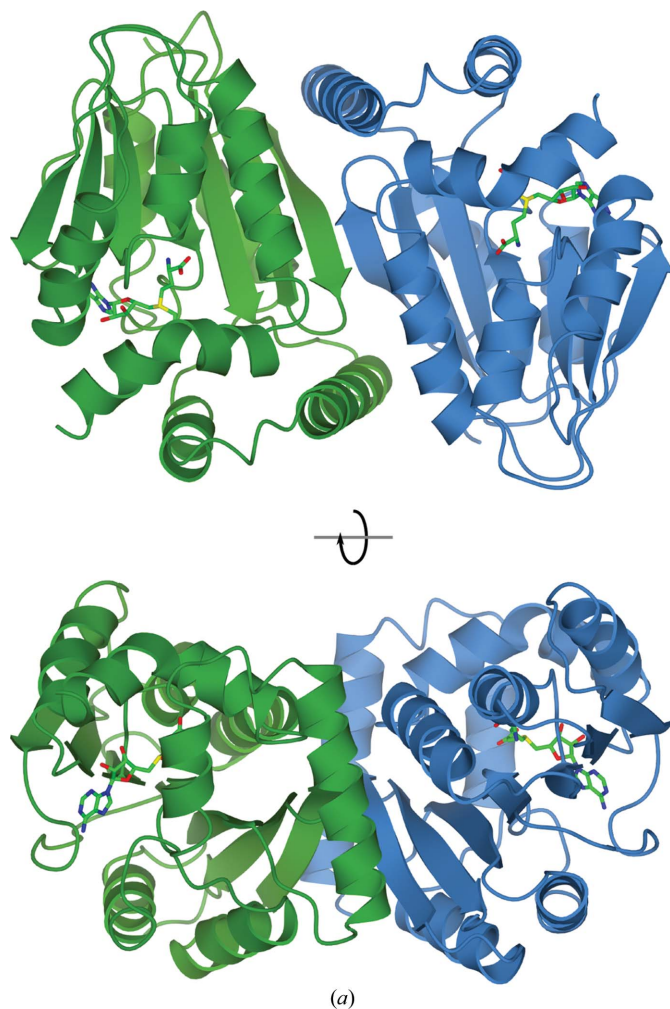
Samples of CmoA were purified as described above and diluted to a concentration of 5 µM in 50% (v/v) aqueous



acetonitrile containing 1% formic acid. These samples were introduced into the mass spectrometer using a TriVersa NanoMate ion source (Advion BioSciences) in positive-ion mode. Mass spectra were acquired using a solarix FT-MS (Bruker Daltonics) with a 9.4 T superconducting magnet. Tandem MS of the released ligand was performed by collision-induced dissociation in the hexapole (Q-CID) with argon collision gas. Spectra were processed using *DataAnalysis* v.4.0 (Bruker Daltonics). Protein mass deconvolution was performed using v.2.0 of the *SNAP* algorithm and mass measurements from released ligand and fragmentation spectra were calculated from centroided data.

### 2.7. Systematic name of the cofactor

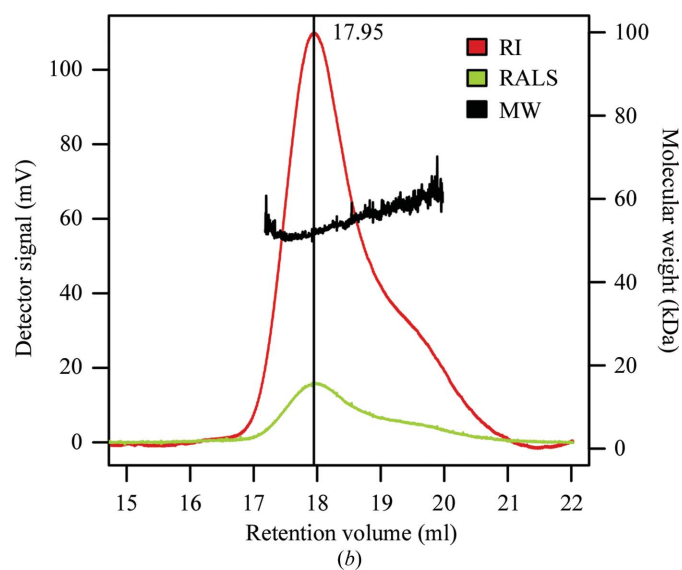
The IUPAC name for the *S*-adenosyl-*S*-carboxymethyl-L-homocysteine (SCM-SA)H cofactor is [(3*S*)-3-amino-3-carboxypropyl][[(2*S*,3*S*,4*R*,5*R*)-5-(6-aminopurin-9-yl)-3,4-dihydroxyoxolan-2-yl]methyl](carboxymethyl)sulfanium.



### 3. Results and discussion

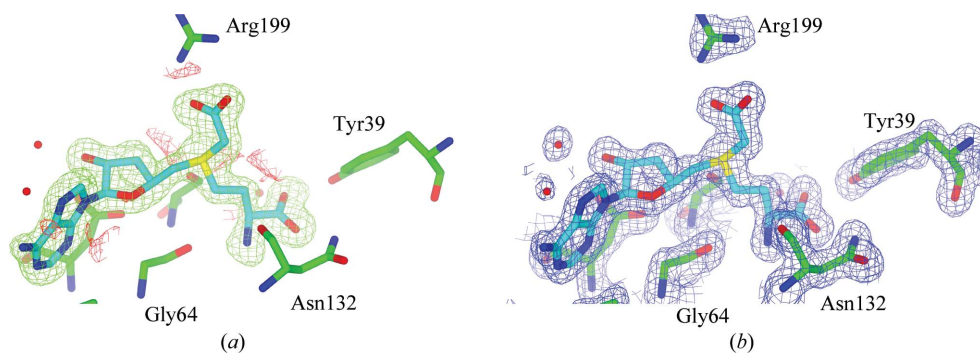
The crystal structure of *E. coli* CmoA was determined by molecular replacement and was refined to  $R_{\text{work}}$  and  $R_{\text{free}}$  values of 19.6 and 23.1%, respectively, using data to a resolution of 1.73 Å (Table 1). The protein copurified with a cofactor from the *E. coli* cells that we anticipated would be either *S*-adenosylmethionine (SAM) or *S*-adenosylhomocysteine (SAH) on the basis of the *S*-adenosylmethionine-binding motifs present in the sequence of CmoA. Unexpectedly, both molecules of CmoA contain the novel derivative *S*-adenosyl-*S*-carboxymethyl-L-homocysteine (SCM-SA)H, which differs from SAM by the substitution of the methyl donor group ( $R = -\text{CH}_3$ ) by a carboxymethyl group ( $R = -\text{CH}_2\text{COOH}$ ).

There are two molecules of CmoA present in the asymmetric unit that are related to each other by a noncrystallographic twofold rotational axis. Apart from minor differences at the N- and C-termini, the two molecules adopt the same conformation and superpose with an r.m.s.d. of 0.3 Å (225 aligned  $C^\alpha$  atoms). Analysis of the structure with *PISA* (Krissinel & Henrick, 2007) reveals that the interface between the two molecules is extensive, with 1274 Å<sup>2</sup> of buried surface area per monomer (Fig. 2). In addition to an antiparallel  $\beta$ -sheet formed by the  $\beta_6$  strands of both molecules, there are additional interactions between helix  $\alpha_6$  of one molecule and helix  $\alpha_2$  and strand  $\beta_6$  of the other molecule. Together, these interactions comprise 15 hydrogen bonds and a number of hydrophobic interactions, suggesting that the CmoA dimer present in the asymmetric unit may also represent the oligomeric state of CmoA in solution. To confirm the existence of this dimer in solution, purified CmoA was analysed by size-exclusion chromatography and static light scattering (SEC-SLS). A single species was visible on the chromatogram and

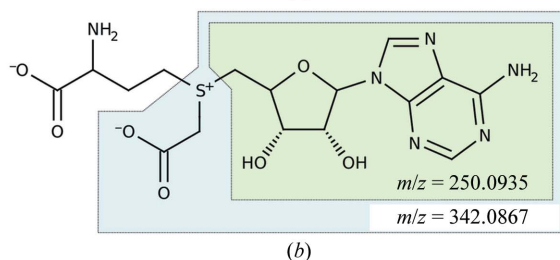
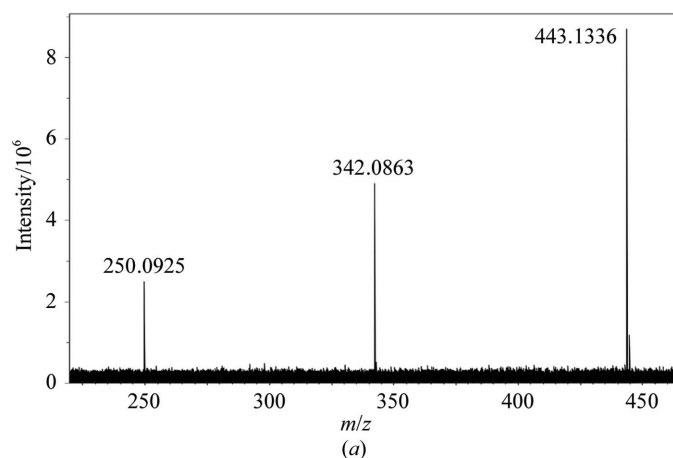


**Figure 2**

The overall structure of the CmoA dimer. (a) Two orthogonal views of the two monomers forming the dimer (green and blue ribbons). The *S*-adenosyl-*S*-carboxymethyl-L-homocysteine (SCM-SA)H cofactor is shown with O atoms in red, N atoms in blue and C atoms in green. (b) The SEC-SLS chromatogram that confirms that CmoA is a dimer in solution. The refractive index (RI; red line) and right-angle light scattering (RALS; green line) traces are displayed. The molecular weight calculated by the *OmniSEC* software (MW; black line) is shown above the elution peak. The dispersity,  $M_w/M_n$ , defined as the ratio of the weight average to number average molecular weights was reported to be 1.001, indicating a highly uniform sample.



**Figure 3**  
Structure of *S*-adenosyl-*S*-carboxymethyl-*L*-homocysteine (SCM-SAH). The final coordinates are displayed with (a) the likelihood-weighted  $mF_o - DF_c$  difference electron-density maps contoured at  $3\sigma$  calculated prior to the modelling of SCM-SAH and (b) the  $2mF_o - DF_c$  electron-density maps contoured at  $1.5\sigma$ . The SCM-SAH model is shown with C atoms in cyan, O atoms in red and N atoms in blue. C atoms of protein residues are shown in green.



**Figure 4**  
Mass spectrum of SCM-SAH. (a) The fragmentation spectrum and (b) the chemical structure of SCM-SAH. Signals consistent with the entire SCM-SAH cofactor as well as two fragments (in the blue and green boxes) were detected.

the molecular weight was calculated to be 52.5 kDa, which is consistent with the theoretical molecular weight of 55.6 kDa for the dimer (Fig. 2). Interestingly, *PISA* identifies an equivalent dimer in the crystal structure of *H. influenzae* YecO (PDB entry 1im8) between chain *A* and its symmetry mate ( $x - y, -y, -z$ ) with an interface area of  $1194 \text{ \AA}^2$  per monomer (Lim *et al.*, 2001). The conserved nature of this interface indicates that the oligomeric state may be important for the structure and function of CmoA.

### 3.1. CmoA contains a novel *S*-adenosylmethionine derivative

During the refinement of the structure, it became apparent from inspection of both the  $2mF_o - DF_c$  and the  $mF_o - DF_c$

electron-density maps that the putative active site of CmoA contained a cofactor that was neither *S*-adenosylmethionine (SAM) nor *S*-adenosylhomocysteine (SAH). Prior to the modelling of the ligand, unambiguous positive density was visible in the  $mF_o - DF_c$  electron-density map for all of the features expected for SAM: *L*-methionine and both the adenine and ribose rings were visible at a contour level of  $5\sigma$ , while the S atom was visible at a contour level of  $18\sigma$  (Fig. 3a). However, additional positive density was present at the end of

the methyl group, indicating that the cofactor was a covalently modified derivative of SAM. This positive density was bifurcated, planar in shape and visible at a contour level of  $7\sigma$ . Because these maps were generated prior to the modelling of any cofactor and no cofactor was present in the search model that was used during molecular replacement, the presence of this additional density was not a consequence of model bias. The shape of the density was most consistent with an *S*-adenosylmethionine derivative in which the methyl group had been derivatized with a functional group with trigonal planar geometry.

To confirm the presence of the cofactor and investigate its identity, samples of CmoA were analysed by Fourier transform mass spectrometry (FT-MS). To exclude the possibility that the modification occurred during crystallization or data collection (as a result of the chemicals present in the crystallization solution or of exposure to X-rays), the sample used for analysis was not crystallized but was from the same preparation as that used for crystallization. A signal at  $m/z = 27\,764.2$  was assigned as a protonated molecular-ion peak for CmoA, in agreement with the value of  $m/z = 27\,763.9$  estimated from the sequence alone (a difference of 0.3 Da after accounting for the proton). An additional signal at  $m/z = 443.1333$  was detected and this was isolated and further analysed by collision-induced dissociation. The resulting fragmentation spectrum contained three signals: a parent ion at  $m/z = 443.1336$  and two fragments at  $m/z = 342.0863$  and  $250.0925$  (Fig. 4a). These are not consistent with the theoretical values for SAM ( $m/z = 399.1445$ ) or SAH ( $m/z = 385.1289$ ), and the fragmentation spectrum featured no signals at these values. A search of the PubChem database for compounds structurally similar to SAH or SAM with a molecular weight of between 442.6 and 443.6 Da resulted in single hit: a SAM derivative in which the methyl group is substituted by a carboxymethyl group (CID 11212932; Fig. 4b). We refer to this derivative as *S*-adenosyl-*S*-carboxymethyl-*L*-homocysteine (or *S*-carboxymethylated SAH), which could be further abbreviated as SCM-SAH; the full IUPAC systematic name is given in §2.7. The  $m/z$  values calculated for this compound are in close agreement with those determined experimentally: the parent

ion has a calculated value of  $m/z = 443.1336$  ( $-0.70$  mDa difference) and two potential fragments may be generated with calculated values of  $m/z = 342.0867$  and  $250.0935$  ( $-0.33$  and  $-0.93$  mDa difference, respectively; Figs. 4*a* and 4*b*).

Restraints for the modelling and refinement of SCM-SAH were generated with *PRODRG2* and the ligand was then modelled into the active site of each monomer (Fig. 3*b*). Following refinement, there was no obvious distortion of the ligand geometry with respect to the ideal geometry and there were no significant positive or negative peaks in the  $mF_o - DF_c$  difference electron-density map in the immediate vicinity of the ligand. Taken together, the crystallographic and mass-spectrometric data suggest that the active site of CmoA contains an *S*-adenosylmethionine derivative in which the methyl group is substituted by a carboxymethyl group.

### 3.2. Comparison with *H. influenzae* YecO

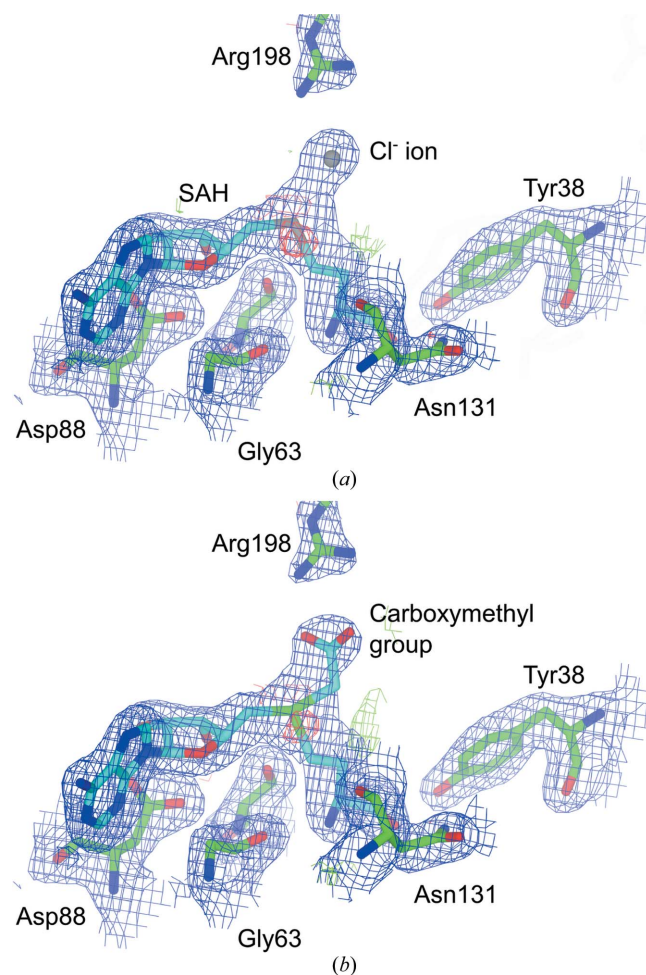
*H. influenzae* YecO (PDB entry 1im8) was identified as the most similar deposited structure to CmoA by both sequence-based (68% sequence identity) and structure-based (r.m.s.d. =  $0.58$  Å for 222 aligned C $^\alpha$  atoms) search methods (Krissinel & Henrick, 2004). The structure of YecO was originally determined by multiple-wavelength anomalous diffraction (MAD) using a selenomethionine derivative produced in *E. coli* B834 (DE3) cells grown in minimal medium in the presence of L-selenomethionine (Lim *et al.*, 2001). The authors noted that in addition to the number of selenium sites expected on the basis of the protein sequence, one additional selenium site per monomer was detected during structure determination. This was incorporated into the cofactor that copurified with YecO, suggesting that the cofactor was derived from L-methionine (in the case of cells grown in non-labelled medium) or L-selenomethionine (in the case of cells grown in minimal medium with L-selenomethionine). In the deposited structure of YecO the cofactor was modelled as Se-substituted SAH with a Cl $^-$  ion  $2.9$  Å away from the Se atom.

Given the high degree of similarity between CmoA and YecO, we re-examined both the coordinates and the structure factors for YecO deposited in the PDB. Re-refinement of the deposited structure resulted in  $R_{\text{work}}$  and  $R_{\text{free}}$  values of 19.2 and 24.8%, respectively, which are comparable with the values of 18.6 and 25.5% originally reported (Fig. 5*a*). Refinement of the YecO structure without any cofactor modelled results in electron density in both the  $2mF_o - DF_c$  and the  $mF_o - DF_c$  electron-density maps into which the Se-substituted form of the SCM-SAH cofactor found in CmoA can be modelled, and the electron-density maps after refinement are also compatible with the presence of this cofactor (Fig. 5*b*). However, the electron density is not defined well enough to allow a distinction between the possibilities of SAH and a Cl $^-$  ion (as modelled originally) or of SCM-SAH (as modelled in CmoA). We note, however, that we were unable to find evidence of SAH and an equivalently positioned Cl $^-$  ion in a manual inspection of PDB entries that are (i) annotated as methyltransferases (EC 2.1.1) and (ii) contain at least one Cl $^-$  ion. Furthermore, in the case of CmoA the electron-density maps

are better defined and the mass-spectrometric data argue against the cofactor modelled in YecO.

### 3.3. Overall structure of CmoA

CmoA has a Rossmann fold that comprises seven  $\beta$ -strands and eight  $\alpha$ -helices. The  $\beta$ -strands form a single sheet in which all strands are parallel except  $\beta 7$ . The majority of the  $\alpha$ -helices pack against both faces of the  $\beta$ -sheet, although helices  $\alpha 2$ ,  $\alpha 6$  and  $\alpha 7$  form a compact lid-like structure that sits over the region containing the SCM-SAH and renders it almost inaccessible to solvent. Superposition of CmoA with the structures of the other RNA methyltransferases TrmA (PDB entry 3bt7) and RumA (RlmD; PDB entry 2bh2) reveals that while the Rossmann-fold core is conserved between these enzymes, the lid-like structure of CmoA obstructs the region used for the binding of the RNA substrate in these methyltransferases (Lee *et al.*, 2005; Alian *et al.*, 2008). The conserved location of the substrate nucleoside with respect to the SAM cofactor in TrmA, RumA and other DNA and RNA methyltransferases



**Figure 5** Modelling of SCM-SAH into YecO. The re-refined coordinates with either (a) the originally modelled Se-substituted SAH and Cl $^-$  ion or (b) remodelled SCM-SAH are displayed along with the  $2mF_o - DF_c$  electron-density maps contoured at  $1\sigma$  and the  $mF_o - DF_c$  difference electron-density maps contoured at  $3\sigma$ . SCM-SAH is depicted with C atoms in cyan, while C atoms of protein residues are shown in green.



suggests that the binding mode is relatively fixed. In order to place the substrate nucleoside in the corresponding position with respect to the SCM-SAH cofactor in CmoA, the lid would need to undergo a large conformational change to allow access to the cofactor and to prevent significant clashes with the neighbouring nucleotides of the tRNA substrate. An alternative possibility is that CmoA acts using additional factors which assist in the modification of the substrate uridine and does not bind tRNA directly.

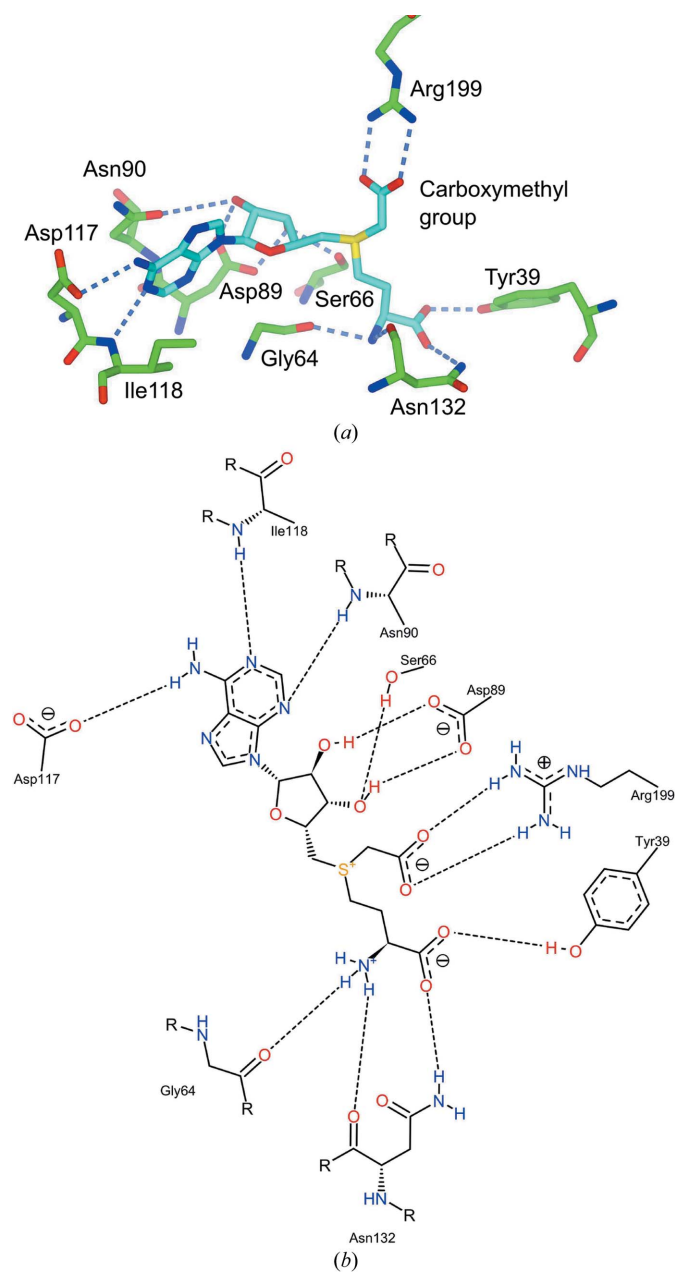
In common with many enzymes possessing a Rossmann fold, the majority of the conserved residues are located at the C-terminal ends of the  $\beta$ -strands or in the loops which

immediately follow. In CmoA these residues are directly involved in contacting the SAM derivative: the adenine ring is hydrogen-bonded by the side chain of Asp117 and the main chains of Asn90, Ile118 and Phe137 (*via* a water molecule), the ribose is hydrogen bonded by the side chains of Ser66, Asp89 and Asn90 (*via* a water molecule), and the L-methionine is hydrogen bonded by the side chains of Tyr39, Asp62 (*via* a water molecule) and Asn132 and the main chains of Gly64, Ala70 (*via* a water molecule) and Asn132 (Figs. 6*a* and 6*b*). The negatively charged carboxylate of the carboxymethyl group interacts with the positively charged guanidinium group of Arg199 through a salt bridge with a length of 2.7 Å. The high conservation of these residues within members of the UniProt CmoA family indicates that they are important for binding the SAM derivative and may also play a role in its biochemistry.

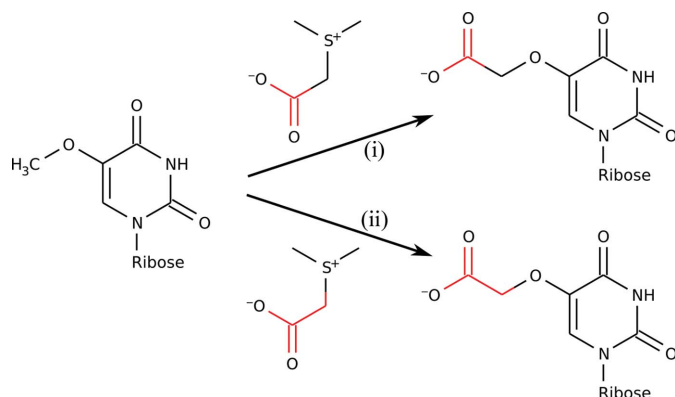
#### 4. Conclusions

We have determined the structure of *E. coli* CmoA, a putative methyltransferase that is involved in the post-transcriptional modification of U34 in a number of bacterial tRNAs. While the sequence motifs and Rossmann fold of the enzyme suggest that it is a typical SAM-dependent methyltransferase, analysis of the electron-density maps and mass-spectrometric data revealed that the protein contains an atypical SAM derivative in which the donor methyl group is replaced by a carboxymethyl group. We name this previously unobserved derivative *S*-adenosyl-*S*-carboxymethyl-L-homocysteine (SCM-SAH). According to the UniProt database, the CmoA family currently contains 1566 proteins that are currently annotated as putative SAM-dependent methyltransferases. However, conservation of Arg199, the key residue of CmoA that stabilizes the negative charge of the carboxyl group of the SCM-SAH cofactor, suggests that these proteins contain the SCM-SAH cofactor instead of SAM and are currently annotated incorrectly. The equivalent residue in known SAM-dependent methyltransferases is not conserved. Although CmoA homologues are only found in bacteria, it is possible that such SAM derivatives are widespread in nature, being present in other enzymes currently annotated as methyltransferases.

Previous genetic studies have indicated that CmoA is involved in the modification of  $mo^5U$  to  $cmo^5U$ , a reaction that involves the addition of a carboxyl group onto the methoxy group of  $mo^5U$  but that cannot be catalysed solely by a methyltransferase (Nasvall *et al.*, 2004). Assuming that the cofactor in CmoA is directly involved in modification of the tRNA, we speculate that it may participate in the formation of  $cmo^5U$  by either (i) the transfer of just the carboxyl group of SCM-SAH onto the methoxy group of  $mo^5U$  or (ii) the substitution of the methyl group of the side chain in  $mo^5U$  by the entire carboxymethyl group from SCM-SAH (Fig. 7). A third possibility is that the carboxymethyl group is transferred directly onto the hydroxyl group of  $ho^5U$ . Although this proposal is not supported by the observation that mutations in *cmoA* result in accumulation of  $mo^5U$  and not  $ho^5U$ , this has



**Figure 6**  
Binding of SCM-SAH by CmoA. (a) SCM-SAH (cyan cylinders) interacts with multiple main-chain and side-chain atoms of CmoA (green cylinders) through hydrogen bonds (blue dashes). (b) A two-dimensional schematic of the active site.



**Figure 7**

Speculative roles for SCM-SAH in the modification of m<sup>5</sup>U. The modification might involve either (i) the addition of the carboxyl group (red) from SCM-SAH onto the methoxy group of m<sup>5</sup>U or (ii) the substitution of the methyl group of m<sup>5</sup>U with the entire carboxymethyl group.

been suggested previously (Murao *et al.*, 1978) and a precedent for the chemistry of this reaction can be found in the O-methyltransferases. All three possibilities, however, would be compatible with the observation that only one of the C atoms in the side chain of cmo<sup>5</sup>U is derived from L-methionine. The proposed mechanisms for SAM-dependent methyltransferases often involve a general base. Superpositions of CmoA with the 5-methyluridine methyltransferases TrmA and RumA show that the C5 atom of the substrate uridine is neighbored by Glu164. This residue is highly conserved in members of the CmoA family and the only other amino acid found at this position is aspartic acid. As the side chain of cmo<sup>5</sup>U derivatives is attached to the C5 atom of the pyrimidine ring, this residue would potentially be able to act as a general base during the reaction.

Although DNA and RNA methyltransferases are able to use synthetic S-adenosylmethionine analogues with extended carbon chains both *in vitro* and *in vivo* (Schlenk & Dainko, 1975; Klimasauskas & Weinhold, 2007; Motorin *et al.*, 2011), there do not appear to be any reports indicating that DNA or RNA methyltransferases actually use such analogues *in vivo* for the modification of nucleic acids. We hope that our findings will lead to further characterization of the function and mechanism of CmoA and its SCM-SAH cofactor.

During the final stages of preparation of our manuscript, we became aware of PDB deposition 4gek by the New York Structural Genomics Research Consortium, in which *E. coli* CmoA is also observed in complex with SCM-SAH. A comparison of these independently determined structures adds support to the conclusions presented in this paper.

The authors would like to thank Gideon Grogan (York Structural Biology Laboratory, University of York, England) for useful suggestions. We would also like to thank Osnat Herzberg (Institute for Bioscience and Biotechnology Research, University of Maryland, USA) for a discussion regarding the structure of YecO. The OPPF-UK is funded by the Medical Research Council and the Biotechnology and

Biological Sciences Research Council. Wellcome Trust Centre for Human Genetics is supported by the Wellcome Trust (grant No. 075491). Mass-spectrometric experiments were funded by the Wellcome Trust grant 098230 to AAA. These experiments were performed at York Centre of Excellence in Mass Spectrometry funded by the Yorkshire Forward/Northern Way Initiative. We thank Matthew Jennions from the Membrane Protein Laboratory, Imperial College for the SEC-SLS experiment and for assistance with interpreting the data.

## References

- Agris, P. F. (2008). *EMBO Rep.* **9**, 629–635.
- Agris, P. F., Vendeix, F. A. & Graham, W. D. (2007). *J. Mol. Biol.* **366**, 1–13.
- Alian, A., Lee, T. T., Griner, S. L., Stroud, R. M. & Finer-Moore, J. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 6876–6881.
- Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I. & Owens, R. J. (2007). *Nucleic Acids Res.* **35**, e45.
- Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., Vendeix, F. A., Fabris, D. & Agris, P. F. (2011). *Nucleic Acids Res.* **39**, D195–D201.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. R. (2011). *Acta Cryst.* **D67**, 282–292.
- Gorrec, F. (2009). *J. Appl. Cryst.* **42**, 1035–1042.
- Hagervall, T. G., Jönsson, Y. H., Edmonds, C. G., McCloskey, J. A. & Björk, G. R. (1990). *J. Bacteriol.* **172**, 252–259.
- Jühling, F., Mörl, M., Hartmann, R. K., Sprinzl, M., Stadler, P. F. & Pütz, J. (2009). *Nucleic Acids Res.* **37**, D159–D162.
- Keegan, R. M. & Winn, M. D. (2007). *Acta Cryst.* **D63**, 447–457.
- Klimasauskas, S. & Weinhold, E. (2007). *Trends Biotechnol.* **25**, 99–104.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Lee, T. T., Agarwalla, S. & Stroud, R. M. (2005). *Cell*, **120**, 599–611.
- Leslie, A. G. W. (2006). *Acta Cryst.* **D62**, 48–57.
- Lim, K., Zhang, H., Tempczyk, A., Bonander, N., Toedt, J., Howard, A., Eisenstein, E. & Herzberg, O. (2001). *Proteins*, **45**, 397–407.
- Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
- McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst.* **D67**, 386–394.
- Motorin, Y., Burhenne, J., Teimer, R., Koynov, K., Willnow, S., Weinhold, E. & Helm, M. (2011). *Nucleic Acids Res.* **39**, 1943–1952.
- Motorin, Y. & Helm, M. (2010). *Biochemistry*, **49**, 4934–4944.
- Murao, K., Ishikura, H., Albani, M. & Kersten, H. (1978). *Nucleic Acids Res.* **5**, 1273–1281.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Nasvall, S. J., Chen, P. & Björk, G. R. (2004). *RNA*, **10**, 1662–1673.
- Sauter, N. K., Grosse-Kunstleve, R. W. & Adams, P. D. (2004). *J. Appl. Cryst.* **37**, 399–409.
- Schlenk, F. & Dainko, J. L. (1975). *Biochim. Biophys. Acta*, **385**, 312–323.
- Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Cryst.* **D60**, 1355–1363.
- Selmer, M., Dunham, C. M., Murphy, F. V. IV, Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R. & Ramakrishnan, V. (2006). *Science*, **313**, 1935–1942.
- Shi, H. & Moore, P. B. (2000). *RNA*, **6**, 1091–1105.

- Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.
- Stierand, K., Maass, P. C. & Rarey, M. (2006). *Bioinformatics*, **22**, 1710–1716.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst. D* **66**, 22–25.
- Vendeix, F. A., Dziergowska, A., Gustilo, E. M., Graham, W. D., Sproat, B., Malkiewicz, A. & Agris, P. F. (2008). *Biochemistry*, **47**, 6117–6129.
- Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst. D* **57**, 122–133.
- Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
- Zhang, Z., Sauter, N. K., van den Bedem, H., Snell, G. & Deacon, A. M. (2006). *J. Appl. Cryst.* **39**, 112–119.



## Appendix 6B

Declaration from Robert Byrne attesting to my contribution to paper 6.



UNIVERSITY  
*of York*

## Declaration of reviewee's contribution

### Paper

Byrne, R.T., Whelan, F., Aller, P., Bird, L.E., **DOWLE, A.A.**, Lobley, C. Reddivari, Y., Nettleship, J.E., Owens, R.J., Antson, A.A. & Waterman, D.G. (2013) *S*-adenosyl-*S*-carboxymethyl-*L*-homocysteine: a novel co-factor found in the putative tRNA modifying enzyme CmoA *Acta Crystallographica Section D* **69**, 1090-1098

### Candidate's contribution

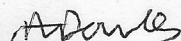
The submitting candidate performed mass spectrometric analysis of the purified protein CmoA with the aim of identifying its novel co-factor. The candidate was responsible for all aspects FT-ICR-MS analysis including: method development, acquisition, and data analysis. Purified protein was produced by the Oxford Protein Production Facility. The work was provided as fee for service through the Bioscience Technology Facility, Department of Biology, University of York. Contributions to data analysis, production of figures and input into writing the paper were provided on a collaborative basis.

### Declaration

I attest that the summation above is a true reflection of the candidate's (Adam A. Dowle) contribution to the paper.

### Candidate


Name: Adam Dowle

Signature: 

Date: 1st June 2016

### Co-author

Name: Robert Byrne

Signature: 

Date: 1 JUNE 2016

## Appendix 7A

Submitted paper 7 - Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein.

## PLANT SCIENCE

# Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein

Thilo Winzer,<sup>1</sup> Marcelo Kern,<sup>1</sup> Andrew J. King,<sup>1</sup> Tony R. Larson,<sup>1</sup> Roxana I. Teodor,<sup>1</sup> Samantha L. Donninger,<sup>1</sup> Yi Li,<sup>1</sup> Adam A. Dowle,<sup>2</sup> Jared Cartwright,<sup>2</sup> Rachel Bates,<sup>2</sup> David Ashford,<sup>2</sup> Jerry Thomas,<sup>2</sup> Carol Walker,<sup>3</sup> Tim A. Bowser,<sup>3</sup> Ian A. Graham<sup>1\*</sup>

Morphinan alkaloids from the opium poppy are used for pain relief. The direction of metabolites to morphinan biosynthesis requires isomerization of (*S*)- to (*R*)-reticuline. Characterization of high-reticuline poppy mutants revealed a genetic locus, designated *STORR* [(*S*)- to (*R*)-reticuline] that encodes both cytochrome P450 and oxidoreductase modules, the latter belonging to the aldo-keto reductase family. Metabolite analysis of mutant alleles and heterologous expression demonstrate that the P450 module is responsible for the conversion of (*S*)-reticuline to 1,2-dehydroreticuline, whereas the oxidoreductase module converts 1,2-dehydroreticuline to (*R*)-reticuline rather than functioning as a P450 redox partner. Proteomic analysis confirmed that these two modules are contained on a single polypeptide in vivo. This modular assembly implies a selection pressure favoring substrate channeling. The fusion protein *STORR* may enable microbial-based morphinan production.

The naturally occurring opiates of the morphinan subclass of benzylisoquinoline alkaloids (BIAs) include morphine, codeine, and thebaine. Morphine and codeine can be directly used as analgesic painkillers, and thebaine is widely used as a feedstock for the synthesis of a number of semisynthetic opiates, including hydrocodone, hydromorphone, oxycodone, and oxymorphone, as well as the opioid antagonist naloxone. The discovery and isolation of morphine from the opium poppy (*Papaver somniferum* L.) by Friedrich Sertürner in 1806 (1) are a milestone in the history of pharmacy. More than 200 years later, opiate alkaloid-based pharmaceutical formulations remain the most potent treatment for severe pain, with sales totaling \$US1.6 billion in 2013 (2).

BIAs are found in a number of species in the Papaveraceae family, but morphine production has only been reported in the opium poppy and the closely related *P. setigerum* (3). The morphinan backbone contains five asymmetric carbon centers. Total chemical synthesis, although possible (4), is not an economically viable means of production. Consequently, morphinan alkaloids are still exclusively sourced from the opium poppy plant. Much effort has gone into the elucidation of the morphinan branch of BIA metabolism over the past 25 years, resulting in the identi-

fication of genes for all but the gateway step involving the epimerization of (*S*)- to (*R*)-reticuline (3, 5–14). Thus, although it has been possible to produce both (*S*)- and racemic mixtures of (*R,S*)-reticuline and morphinans in metabolically engineered microbial systems (15–18), the clean enzymatic conversion of (*S*)- to (*R*)-reticuline remains the goal.

(*S*)-reticuline is the central intermediate of BIA metabolism (fig. S1), and conversion to its *R* epimer is believed to be a two-step process (19, 20). The *S* epimer is first oxidized to the quaternary positively charged amine 1,2-dehydroreticuline, followed by reduction to (*R*)-reticuline (Fig. 1A). Activities for each step have been reported, but the identity of the corresponding proteins has not been established (19, 20). We have combined a candidate gene approach with genetic analyses of F<sub>2</sub> populations of *P. somniferum* segregating for mutations that are deficient in (*S*)- to (*R*)-reticuline conversion and discovered that a fusion protein is responsible for sequentially catalyzing both steps of the epimerization.

RNA interference (RNAi) knockdown of codeinone reductase in the opium poppy was reported to cause accumulation of (*S*)-reticuline, which is eight steps upstream of the codeinone reductase substrate (21), a result the authors attributed to metabolite channeling. We considered an alternative hypothesis to be the off-target cosilencing of a closely related oxidoreductase involved in the conversion of (*S*)- to (*R*)-reticuline. Using the sequence of the RNAi silencing construct to query an in-house expressed sequence tag (EST) library from stem and capsule tissue of opium poppies (22), we identified a contiguous assembly comprising a cytochrome P450 mono-

oxigenase that is 3'-linked to an oxidoreductase. Sequencing cDNA clones from opium poppy stems confirmed the in-frame fusion transcript (Fig. 1A and fig. S2). The P450 module was designated CYP82Y2.

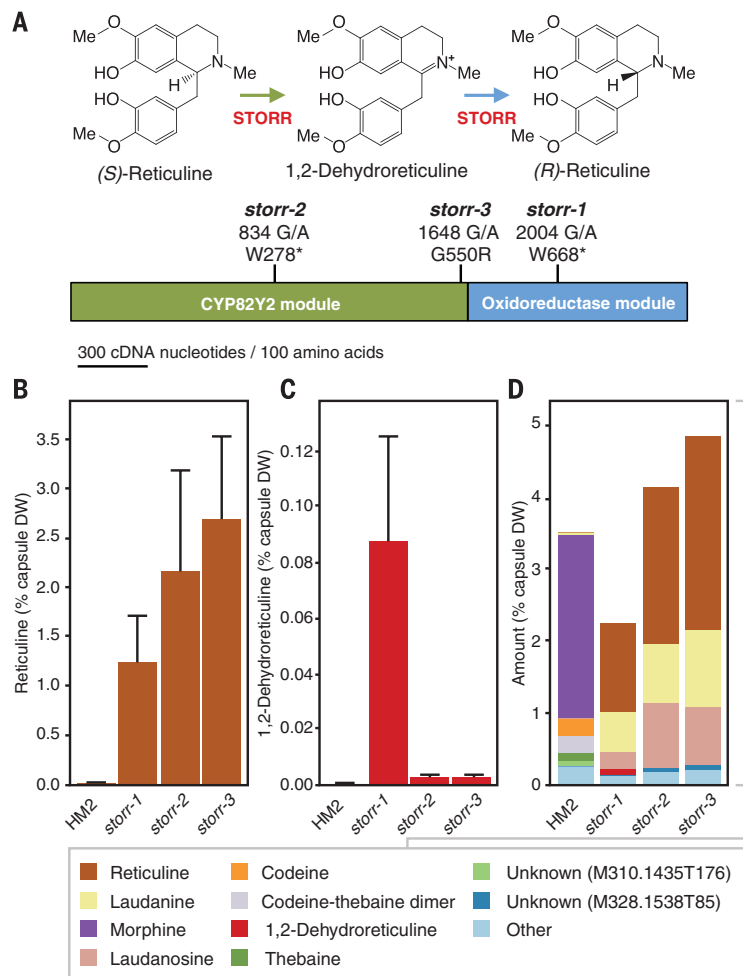
To investigate whether the corresponding gene is a candidate for one or both steps in the epimerization of (*S*)- to (*R*)-reticuline, we sequenced corresponding cDNA clones from three independent mutants identified from an ethyl methanesulfonate-mutagenized population of a high-morphine cultivar, HM2 (23). All three mutants have lost the ability to produce morphinan alkaloids and instead accumulate high levels of (*S*)-reticuline as well as the (*S*)-reticuline-derived alkaloids laudanine and laudanosine (Fig. 1, B and D). We found that all three mutant lines carry mutations in the corresponding gene locus (Fig. 1A), which we name *STORR* [(*S*)- to (*R*)-reticuline]. The *storr-1* allele carries a premature stop codon corresponding to amino acid position W668 in the oxidoreductase module of the predicted fusion protein. *storr-1* plants also contain low but significant levels of 1,2-dehydroreticuline (Fig. 1C), which suggests that the oxidoreductase module catalyzes the second step of the epimerization, the reduction of 1,2-dehydroreticuline to (*R*)-reticuline. *storr-2* and *storr-3* are both disrupted in the CYP82Y2 module: *storr-2* contains a premature stop at codon position W278, and *storr-3* contains a missense mutation causing a glycine-to-arginine substitution at position 550 (Fig. 1A). Dried capsules of *storr-2* and *storr-3* accumulate (*S*)-reticuline but not 1,2-dehydroreticuline, suggesting that the CYP82Y2 module is responsible for the first epimerization step, the oxidation of (*S*)-reticuline to 1,2-dehydroreticuline. Complementation tests and F<sub>2</sub> segregation analysis confirmed that a single genetic locus is responsible for the high-reticuline phenotype, association of the three recessive *storr* alleles with the high-reticuline phenotype, and the consecutive roles of the CYP82Y2 and oxidoreductase modules in the epimerization of (*S*)- to (*R*)-reticuline (tables S2 and S3).

To establish whether the *STORR* locus is not only transcribed but also translated as a fusion protein, we used a quantitative mass spectrometry approach after gel fractionation of crude protein extract from HM2. Peptides from across the entire *STORR* protein were found to be most abundant in the gel regions covering the 100.65-kD predicted size of the fusion protein, confirming this as the in vivo form (Fig. 2). For direct functional characterization, the *STORR* fusion protein and the separate modules were expressed in *Saccharomyces cerevisiae*, and enzyme assays were performed on soluble extracts and microsomal preparations (Fig. 3). We found that 1,2-dehydroreticuline is converted to (*R*)-reticuline with 100% conversion efficiency by both the *STORR* fusion protein and the oxidoreductase module, but not by the CYP82Y2 module plus its redox partner (Fig. 3A). In contrast, the CYP82Y2 module plus its redox partner catalyzed 97% conversion of (*S*)-reticuline to 1,2-dehydroreticuline,

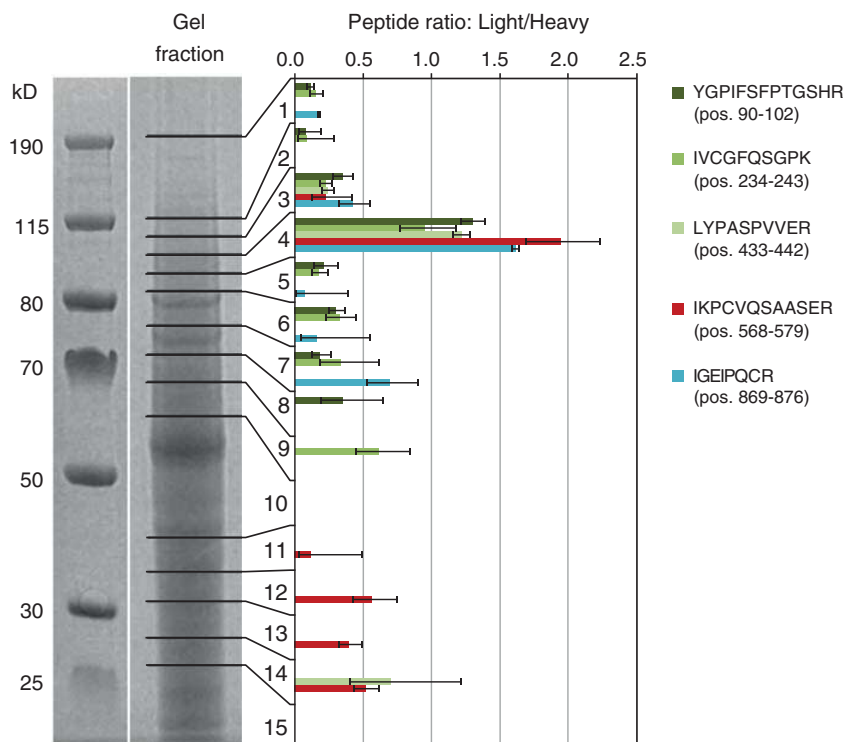
<sup>1</sup>Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, UK. <sup>2</sup>Bioscience Technology Facility, Department of Biology, University of York, York YO10 5DD, UK. <sup>3</sup>GlaxoSmithKline, 1061 Mountain Highway, Post Office Box 168, Boronia, Victoria 3155, Australia.

\*Corresponding author. E-mail: ian.graham@york.ac.uk

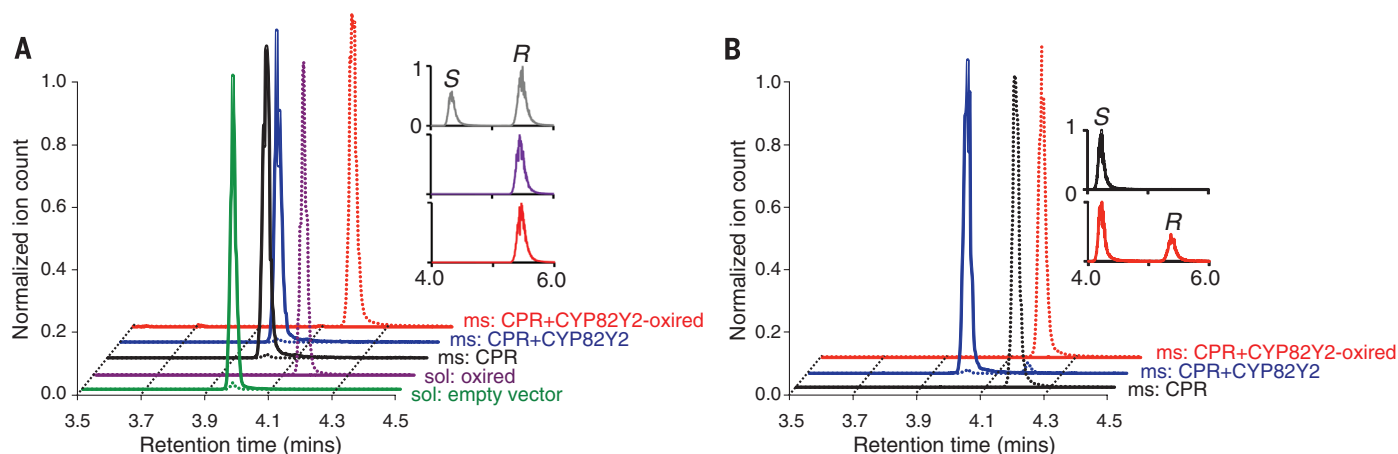
**Fig. 1. Characterization of opium poppy mutants disrupted in the conversion of (S)- to (R)-reticuline.** (A) Schematic showing epimerization of (S)- to (R)-reticuline and position of the *storr-1*, *storr-2*, and *storr-3* mutations in the predicted fusion protein. (B) Mean  $\pm$  SD capsule reticuline content in the HM2 wild-type cultivar and *storr* mutants (HM2,  $n = 5$ ; *storr-1*,  $n = 12$ ; *storr-2*,  $n = 17$ ; *storr-3*,  $n = 15$ ). DW, dry weight. Reticuline content was verified as  $>99.2\%$  (S)-reticuline in all mutants by chiral high-performance liquid chromatography (HPLC) (table S1). (C) 1,2-dehydroreticuline. (D) All compounds  $>1\%$  total alkaloids ( $n = 10$ ) are individually identified, with minor peaks ( $n = 379$ ), grouped as "Other."



**Fig. 2. Size determination of STORR protein in the opium poppy.** Protein extracts from three stem samples of HM2 wild type were fractionated, together with size markers, by SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and the three lanes were each cut into 15 fractions (see horizontal lines on one representative lane) to resolve the predicted fusion protein and putative individual CYP450 and oxidoreductase modules (101, 65, and 36 kD, respectively). For relative quantification, an equal amount of a tryptic digest of  $^{15}\text{N}$ -labeled recombinant STORR protein was spiked into the in-gel digest of each SDS-PAGE fraction before HPLC mass spectrometry (HPLC-MS). Ratios of peak areas from extracted-ion chromatograms of light (endogenous) to heavy (labeled) versions of five peptides from across the STORR protein sequence were compared. Ratios of normalized peak areas from extracted-ion chromatograms were converted to binary logarithms for the calculation of means and standard errors of the mean. Only measurements where the respective peptides were found in all three biological replicates are shown.







**Fig. 3. Functional characterization of the STORR fusion protein by heterologous expression in *S. cerevisiae*.** (A) HPLC-MS analysis of the in vitro conversion of 1,2-dehydroreticuline to (*R*)-reticuline. Crude soluble (sol) or microsomal (ms) preparations harboring the empty pESC-TRP vector (green), vector containing the oxidoreductase module (purple), an opium poppy cytochrome P450 reductase (CPR) redox partner (black), CPR + CYP82Y2 (blue), or CPR + the CYP82Y2-oxidoreductase fusion (red) were assayed (21). The solid lines of the HPLC-MS chromatograms show the normalized total ion count at a mass/charge ratio (*m/z*) of 328, corresponding to 1,2-dehydroreticuline (substrate), whereas the dotted lines show the normalized total ion count at *m/z* 330, corresponding to reticuline (product). The inset panel shows the chiral analysis of reticuline: The gray

trace is for an (*S*)- and (*R*)-reticuline standard, and the purple and red traces correspond to reticuline derived by activity of the oxidoreductase and CYP82Y2-oxidoreductase fusion, respectively. (B) HPLC-MS analysis of the conversion of (*S*)-reticuline into 1,2-dehydroreticuline and (*R*)-reticuline. Crude microsomal preparations obtained from *S. cerevisiae* harboring expression vector pESC-TRP containing CPR only (black), CPR + a CYP82Y2 module (blue), or CPR + CYP82Y2-oxidoreductase fusion (red) were assayed (21). The solid lines of the HPLC-MS chromatograms show the normalized total ion count at *m/z* 328, corresponding to 1,2-dehydroreticuline, and the dotted lines show the normalized total ion count at *m/z* 330, corresponding to reticuline. The inset panel shows the chiral analysis of reticuline, with the same line colors as in the main panel.

demonstrating that it acts as a 1,2-dehydroreticuline synthase (Fig. 3B). Microsomal preparations harboring the entire STORR fusion protein converted about 20% of the added (*S*)-reticuline to (*R*)-reticuline, confirming the bifunctional role of the protein in performing sequential reactions in the epimerization of reticuline. Kinetic analysis revealed that the microsomal CYP82Y2 module alone and the CYP82Y2-oxidoreductase fusion had similar Michaelis constant values of 13 and 14  $\mu\text{M}$  for (*S*)-reticuline and 1,2-dehydroreticuline, respectively (fig. S3). Consistent with the plant mutant phenotypes, we found that microsomally expressed STORR carrying the *storr-2* mutation lacks both the P450 and oxidoreductase activities, the *storr-3* mutation lacks the P450 activity but does still exhibit the oxidoreductase activity, and the *storr-1* mutation has lost the oxidoreductase activity but maintains very low levels of P450 activity (fig. S4).

P450-redox systems where the P450 enzyme is covalently linked to redox partner reductase components are well known in both prokaryotes and lower eukaryotes (24). In the STORR fusion protein, the P450 module is linked to a reductase, but rather than functioning as a redox partner for the P450, this reductase catalyzes the product of the P450 to complete a two-step epimerization of (*S*)- to (*R*)-reticuline. Other forms of bifunctional P450 fusions with oxygenase/peroxidase, hydrolase, and dioxygenase modules have been reported to occur in ascomycetes, and all of these also appear to catalyze sequential

reactions (25–27). A possible explanation as to why such fusion proteins evolve is that they facilitate efficient channeling of highly unstable or reactive intermediates. Evidence for efficient substrate channeling in the case of the STORR fusion protein comes from the observation that microsomal fractions harboring the fusion protein directly convert (*S*)- to (*R*)-reticuline, with no detectable accumulation of 1,2-dehydroreticuline (Fig. 3B).

Phylogenetic analysis suggests that the STORR fusion occurred after the split of the CYP82X and the CYP82Y subfamilies and that the oxidoreductase module falls into subfamily 4 of the aldo-keto reductases, with closest homology to the codeinone reductase family from *P. somniferum* (fig. S5). A query of the 1K plant transcriptome resource (28) identified similar predicted module arrangements in EST collections from two other morphinan-producing *Papaver* species, *P. bracteatum* and *P. setigerum*, but not in *P. rhoeas*, which does not make morphinans (table S4). We hypothesize that the STORR fusion was the key step in the evolution of the morphinan branch of BIA metabolism. After this step, other enzymes were recruited and adapted, including dioxygenases and reductases, ultimately giving rise to codeine and morphine in *P. somniferum* and *P. setigerum* (29). Thus, this morphinan biosynthetic pathway, and probably other plant secondary metabolic pathways, depends on the organization of both individual gene structure and genome rearrangement (22).

## REFERENCES AND NOTES

- F. Sertürner, *J. Pharmacie* **14**, 47–93 (1806).
- IMS Health Database, Formulation sales by opiate molecule, www.imshealth.com (2013).
- J. Ziegler et al., *Plant J.* **48**, 177–192 (2006).
- M. Gates, G. Tschudi, *J. Am. Chem. Soc.* **78**, 1380–1393 (1956).
- N. Samanani, D. K. Liscombe, P. J. Facchini, *Plant J.* **40**, 302–313 (2004).
- A. Ounaron, G. Decker, J. Schmidt, F. Lottspeich, T. M. Kutchan, *Plant J.* **36**, 808–819 (2003).
- K. B. Choi, T. Morishige, F. Sato, *Phytochemistry* **56**, 649–655 (2001).
- H. H. Pauli, T. M. Kutchan, *Plant J.* **13**, 793–801 (1998).
- T. Morishige, T. Tsujita, Y. Yamada, F. Sato, *J. Biol. Chem.* **275**, 23398–23405 (2000).
- A. Gesell et al., *J. Biol. Chem.* **284**, 24432–24442 (2009).
- R. Lenz, M. H. Zenk, *J. Biol. Chem.* **270**, 31091–31096 (1995).
- T. Grothe, R. Lenz, T. M. Kutchan, *J. Biol. Chem.* **276**, 30717–30723 (2001).
- J. M. Hagel, P. J. Facchini, *Nat. Chem. Biol.* **6**, 273–275 (2010).
- B. Unterlinner, R. Lenz, T. M. Kutchan, *Plant J.* **18**, 465–475 (1999).
- A. Nakagawa et al., *Nat. Commun.* **2**, 326 (2011).
- A. Nakagawa et al., *Sci. Rep.* **4**, 6695 (2014).
- K. M. Hawkins, C. D. Smolke, *Nat. Chem. Biol.* **4**, 564–573 (2008).
- K. Thodey, S. Galanie, C. D. Smolke, *Nat. Chem. Biol.* **10**, 837–844 (2014).
- W. De-Eknankul, M. H. Zenk, *Phytochemistry* **31**, 813–821 (1992).
- K. Hirata, C. Poeaknapo, J. Schmidt, M. H. Zenk, *Phytochemistry* **65**, 1039–1046 (2004).
- R. S. Allen et al., *Nat. Biotechnol.* **22**, 1559–1566 (2004).
- T. Winzer et al., *Science* **336**, 1704–1708 (2012).
- Materials and methods are available as supplementary materials on Science Online.
- F. P. Guengerich, A. W. Munro, *J. Biol. Chem.* **288**, 17065–17073 (2013).



## Appendix 7B

Declaration from Thilo Winzer attesting my contribution to paper 7.



UNIVERSITY  
*of York*

## **Declaration of reviewee's contribution**

### **Paper**

Winzer, T., Kern, M., King, A. J., Larson, T. R., Teodor, R., Donninger, S., Li, Y., **DOWLE, A. A.**, Cartwright, J., Bates, R., Ashford, D., Thomas, J., Walker, C., Bowser, T. A., Graham, I. A.\* (2015) Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein *Science* **349(6245)**, 309-12

### **Candidate's contribution**

The candidate's role was to carry out mass spectrometry-based proteomics analysis to confirm that the protein STORR is endogenously expressed as a fusion of two gene products. Proteomic analysis included relative quantification against a  $^{15}\text{N}$  standard produced by Jared Cartwright. Thilo Winzer and Marcelo Kern were responsible for the provision of all endogenous protein samples. The submitting candidate played a highly significant role in all aspects of the mass spectrometric analysis including, experimental design, data acquisition and analysis. The work was provided as fee for service through the Bioscience Technology Facility, Department of Biology, University of York. Contributions to data analysis, production of figures and input into writing the paper were provided on a collaborative basis.

### **Declaration**

I attest that the summation above is a true reflection of the candidate's (Adam A. Dowle) contribution to the paper.

### **Candidate**

Name: Adam Dowle

Signature: 

Date: 20<sup>th</sup> May 2016

### **Co-author**

Name: Thilo Winzer

Signature: 

Date: 20/5/2016

## Appendix 7C

Declaration from Jerry Thomas attesting to my contribution to paper 7.



UNIVERSITY  
*of York*

## **Declaration of reviewee's contribution**

### **Paper**

Winzer, T., Kern, M., King, A. J., Larson, T. R., Teodor, R., Donninger, S., Li, Y., **DOWLE, A. A.**, Cartwright, J., Bates, R., Ashford, D., Thomas, J., Walker, C., Bowser, T. A., Graham, I. A.\* (2015) Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein *Science* **349(6245)**, 309-12

### **Candidate's contribution**


The candidate's role was to carry out mass spectrometry-based proteomics analysis to confirm that the protein STORR is endogenously expressed as a fusion of two gene products. Proteomic analysis included relative quantification against a  $^{15}\text{N}$  standard produced by Jared Cartwright. Thilo Winzer and Marcelo Kern were responsible for the provision of all endogenous protein samples. The submitting candidate played a highly significant role in all aspects of the mass spectrometric analysis including, experimental design, data acquisition and analysis. The work was provided as fee for service through the Bioscience Technology Facility, Department of Biology, University of York. Contributions to data analysis, production of figures and input into writing the paper were provided on a collaborative basis.

### **Declaration**

I attest that the summation above is a true reflection of the candidate's (Adam A. Dowle) contribution to the paper.

### **Candidate**

**Name:** Adam Dowle

**Signature:** 

**Date:** 17<sup>th</sup> May 2016

### **Co-author**

**Name:** Jerry Thomas

**Signature:** 

**Date:** 17-5-16

## Abbreviations

1D-PAGE	One-dimensional polyacrylamide gel electrophoresis
2D-FT-ICR-MS/MS	Two-dimensional Fourier transform ion cyclotron resonance tandem mass spectrometry
2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
A	Alanine
ac	Acetylated
BES	<i>Brugia malayi</i> excretory-secretory proteome
BmA	<i>Brugia malayi</i> soluble somatic protein extract
C	Cysteine
cDNA	Complementary deoxyribonucleic acid
CI	Confidence interval
CID	Collision induced dissociation
C-S	Carbon-sulfur
D	Aspartic acid
Da	Daltons
DC	Dendritic cells
DNA	Deoxyribonucleic acid
E	Glutamic acid
<i>E. coli</i>	<i>Escherichia coli</i>
emPAI	Exponentially modified protein abundance index
ERM	Egg-released material
ES	Excretory-secretory
EST	Expressed sequences tag
ETD	Electron transfer dissociation
F	Phenylalanine
FA	Fred Antson
FDR	False discovery rate
FTICR-MS	Fourier transform ion cyclotron resonance mass spectrometry
FW	Fiona Whelan
G	Glycine
GD	Georgina Drury
GO	Gene ontology
Gray	Gray = one joule of radiation per kilogram
H	Histidine
HES	<i>Heligmosomoides polygyrus</i> excretory-secretory proteome
Hex	<i>Heligmosomoides polygyrus</i> soluble somatic protein extract

HILIC	Hydrophilic interaction liquid chromatography
I	Isoleucine
IS	Internal standard
iBAQ	Intensity-based absolute quantification
iTRAQ	Isobaric tags for relative and absolute quantification
JC	Jared Cartwright
JH	James Hewitson
K	Lysine
kDa	Kilodaltons
L	Leucine
L4	Larval stage four secreted <i>Heligmosomoides polygyrus</i> excretory-secretory proteome
LC	Liquid chromatography
LC-ESI-MS/MS	Liquid chromatography-electrospray ionisation-tandem mass spectrometry
LC-MALDI-MS/MS	Liquid chromatography-matrix-assisted laser desorption/ionisation-tandem mass spectrometry
LC-MS	Liquid chromatography-mass spectrometry
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LOPIT	Localisation of organelle proteins using isotope tagging
LPS	Lipopolysaccharide
LPS-DCs	Dendritic cells exposed to bacterial lipopolysaccharide
M	Methionine
MALDI	Matrix-assisted laser desorption/ionisation
MALDI-TOF/TOF	Matrix-assisted laser desorption/ionisation-tandem time of flight mass spectrometry
mM	Millimolar
mDa	Millidaltons
me	Methylated
me <sub>2</sub>	Dimethylated
MED	Culture medium
MED-DCs	Dendritic cells exposed to culture medium
[M+H] <sup>+</sup>	Singly protonated
MK	Marcello Kern
MS	Mass spectrometry
ms	Milliseconds
MS/MS	Tandem mass spectrometry



<i>m/z</i>	Mass-to-charge ratio
N	Asparagine
NCBI	National Center for Biotechnology Information
N	Number (count)
$N_{obs}$	Number of observed peptides
$N_{theo}$	Number of theoretically observable peptides
P	Proline
pI	Isoelectric point
PLOS	Public library of science
PMF	Peptide mass fingerprinting
pos	Position
Pr	Propionylated
PTMs	Post translational modifications
Q	Glutamine
qCID	Collision induced dissociation in the hexapole
QconCAT	Concatamer of quantitative peptides
qTOF	Quadrupole time of flight
R	Arginine
<i>R</i>	Rank
RB	Robert Byrne
RNA	Ribonucleic acid
S	Serine
s	Seconds
<i>s</i>	Variance
SCX	Strong cation exchange
SDS	Sodium dodecyl sulfate
SE	Standard error
SEA	Schistosome egg antigen
SEA-DCs	Dendritic cells exposed to schistosome egg antigen
SF-B	Stéphanie Ferret-Bernard
sin	Sine
SILAC	Stable isotope labeling with amino acids in cell culture
SNAP	Sophisticated numerical annotation procedure
STORR	( <i>S</i> )- to ( <i>R</i> )-reticuline
T	Threonine
<i>t</i>	<i>t</i> -test statistic
Th	T helper lymphocyte cells

Th1	Type 1 immune response
Th2	Type 2 immune response
TOF	Time of flight
TMT	Tandem mass tags
TW	Thilo Winzer
tRNA	Transfer ribonucleic acid
V	Valine
VAL	Venom allergen/ancylostoma secreted proteins
v/v	Volume-to-volume ratio
W	Tryptophan
$W$	Wilcoxon signed-rank test statistic
$\bar{X}$	Mean
Y	Tyrosine
$\Sigma$	Sum

## References

1. Drury GE, Dowle AA, Ashford DA, Waterworth WM, Thomas J, West CE. Dynamics of plant histone modifications in response to DNA damage. *Biochem J*. **2012** Aug 1; 445(3): 393-401
2. Hewitson JP, Harcus YM, Curwen RS, Dowle AA, Atmadja AK, Ashton PD, Wilson A, Maizels RM. The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory-secretory products. *Mol Biochem Parasitol*. **2008** Jul; 160(1): 8-21.
3. Hewitson JP, Harcus Y, Murray J, van Agtmaal M, Filbey KJ, Grainger JR, Bridgett S, Blaxter ML, Ashton PD, Ashford DA, Curwen RS, Wilson RA, Dowle AA, Maizels RM. Proteomic analysis of secretory products from the model gastrointestinal nematode *Heligmosomoides polygyrus* reveals dominance of venom allergen-like (VAL) proteins. *J Proteomics*. **2011** Aug 24; 74(9): 1573-94.
4. Hewitson JP, Ivens AC, Harcus Y, Filbey KJ, McSorley HJ, Murray J, Bridgett S, Ashford D, Dowle AA, Maizels RM. Secretion of protective antigens by tissue-stage nematode larvae revealed by proteomic analysis and vaccination-induced sterile immunity. *PLoS Pathog*. **2013** Aug; 9(8): e1003492
5. Ferret-Bernard S, Castro-Borges W, Dowle AA, Sanin DE, Cook PC, Turner JD, MacDonald AS, Thomas JR, Mountford AP. Plasma membrane proteomes of differentially matured dendritic cells identified by LC-MS/MS combined with iTRAQ labelling. *J Proteomics*. **2012** Jan 4; 75(3): 938-48.
6. Byrne RT, Whelan F, Aller P, Bird LE, Dowle A, Lobley CM, Reddivari Y, Nettleship JE, Owens RJ, Antson AA, Waterman DG. S-Adenosyl-S-carboxymethyl-L-homocysteine: a novel cofactor found in the putative tRNA-modifying enzyme CmoA. *Acta Crystallogr D Biol Crystallogr*. **2013** Jun; 69(Pt6): 1090-8.
7. Winzer T, Kern M, King AJ, Larson TR, Teodor RI, Donninger SL, Li Y, Dowle AA, Cartwright J, Bates R, Ashford D, Thomas J, Walker C, Bowser TA, Graham IA. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science*. **2015** Jul 17; 349(6245): 309-12.
8. Wright GL Jr. High resolution two-dimensional polyacrylamide electrophoresis of human serum proteins. *Am J Clin Pathol*. **1972** Feb; 57(2): 173-85.
9. O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem*. **1975** May 25; 250(10): 4007-21.
10. Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL, Hochstrasser DF. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)*. **1996** Jan; 14(1): 61-5.
11. Weber K, Osborn M. The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J Biol Chem*. **1969** Aug 25; 244(16): 4406-12.
12. Blakesley RW, Boezi JA. A new staining technique for proteins in polyacrylamide gels using coomassie brilliant blue G250. *Anal Biochem*. **1977** Oct; 82(2): 580-2.
13. Porro M, Viti S, Antoni G, Saletti M. Ultrasensitive silver-stain method for the detection of proteins in polyacrylamide gels and immunoprecipitates on agarose gels. *Anal Biochem*. **1982** Dec; 127(2): 316-21.
14. Steinberg TH, Jones LJ, Haugland RP, Singer VL. SYPRO orange and SYPRO red protein gel stains: one-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. *Anal Biochem*. **1996** Aug 1; 239(2): 223-37.
15. Burnette WN. "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal Biochem*. **1981** Apr; 112(2): 195-203.

16. Larsson T, Norbeck J, Karlsson H, Karlsson KA, Blomberg A. Identification of two-dimensional gel electrophoresis resolved yeast proteins by matrix-assisted laser desorption ionization mass spectrometry. *Electrophoresis*. **1997** Mar-Apr; 18(3-4): 418-23.
17. Mørtz E, Vorm O, Mann M, Roepstorff P. Identification of proteins in polyacrylamide gels by mass spectrometric peptide mapping combined with database search. *Biol Mass Spectrom*. **1994** May; 23(5): 249-61.
18. Olsen JV, Ong SE, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics*. **2004** Jun; 3(6): 608-14.
19. Mann M, Højrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom*. **1993** Jun; 22(6): 338-45.
20. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. **1999** Dec; 20(18): 3551-67.
21. Clauser KR, Baker P, Burlingame AL. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem*. **1999** Jul 15; 71(14): 2871-82.
22. Parker KC. Scoring methods in MALDI peptide mass fingerprinting: ChemScore, and the ChemApplex program. *J Am Soc Mass Spectrom*. **2002** Jan; 13(1): 22-39.
23. McLafferty FW. Tandem mass spectrometry. *Science*. **1981** Oct 16; 214(4518): 280-7.
24. Thorne GC, Gaskell SJ. Elucidation of some fragmentations of small peptides using sequential mass spectrometry on a hybrid instrument. *Rapid Commun Mass Spectrom*. **1989** Jul; 3(7): 217-21.
25. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods*. **2007** Sep; 4(9): 709-12.
26. Suckau D, Resemann A, Schuerenberg M, Hufnagel P, Franzen J, Holle A. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal Bioanal Chem*. **2003** Aug; 376(7): 952-65.
27. Tang XJ, Thibault P, Boyd RK. Fragmentation reactions of multiply-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision-induced dissociation. *Anal Chem*. **1993** Oct 15; 65(20):2 824-34.
28. Lau KW, Hart SR, Lynch JA, Wong SC, Hubbard SJ, Gaskell SJ. Observations on the detection of b- and y-type ions in the collisionally activated decomposition spectra of protonated peptides. *Rapid Commun Mass Spectrom*. **2009** May; 23(10): 1508-14.
29. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom*. **1984** Nov; 11(11): 601.
30. Biemann K. Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom*. **1988** Oct; 16(1-12): 99-111.
31. Tang XJ, Boyd RK. An investigation of fragmentation mechanisms of doubly protonated tryptic peptides. *Rapid Commun Mass Spectrom*. **1992** Nov; 6(11) :651-7.
32. McCormack AL, Somogyi A, Dongré AR, Wysocki VH. Fragmentation of protonated peptides: surface-induced dissociation in conjunction with a quantum mechanical approach. *Anal Chem*. **1993** Oct 15; 65(20): 2859-72.
33. Tsaprailis G, Nair GH, Somogyi A, Wysocki VH, Zhong WQ, Futrell JH, Summerfield SG, Gaskell SJ. Influence of Secondary Structure on the Fragmentation of Protonated Peptides *J. Am. Soc. Mass Spectrom*, **1999**; 121: 5142–5154
34. Nesvizhskii AI. Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol Biol*. **2007**; 367: 87-119.

35. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol.* **2001** Mar; 19(3): 242-7.
36. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC. The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis.* **2000** Apr; 21(6): 1104-15.
37. Karas M, Bachmann D, Bahr U, Hillenkamp F. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int. J. Mass Spectrom.* **1987**; (78): 53-68.
38. Huang EC, Henion JD. LC/MS and LC/MS/MS determination of protein tryptic digests. *J Am Soc Mass Spectrom.* **1990** Apr; 1(2): 158-65.
39. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* **1989** Oct 6; 246(4926): 64-71.
40. Hsieh EJ, Bereman MS, Durand S, Valaskovic GA, MacCoss MJ. Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples. *J Am Soc Mass Spectrom.* **2013** Jan; 24(1): 148-53.
41. Wolters DA, Washburn MP, Yates JR 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem.* **2001** Dec 1; 73(23): 5683-90.
42. Boersema PJ, Divecha N, Heck AJ, Mohammed S. Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J Proteome Res.* **2007** Mar; 6(3): 937-46.
43. Yang F, Shen Y, Camp DG, Smith RD. High pH reversed-phase chromatography with fraction concatenation as an alternative to strong-cation exchange chromatography for two-dimensional proteomic analysis. *Expert Review of Proteomics.* **2012**; 9(2): 129-134.
44. Apffel A, Fischer S, Goldberg G, Goodley PC, Kuhlmann FE. Enhanced sensitivity for peptide mapping with electrospray liquid chromatography-mass spectrometry in the presence of signal suppression due to trifluoroacetic acid-containing mobile phases. *J Chromatogr A.* **1995** Sep 29; 712(1): 177-90.
45. Beynon RJ, Pratt JM. Metabolic labeling of proteins for proteomics. *Mol Cell Proteomics.* **2005** Jul; 4(7): 857-72.
46. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics.* **2002** May; 1(5): 376-86.
47. Oda Y, Huang K, Cross FR, Cowburn D, Chait BT. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci USA.* **1999** Jun 8; 96(12): 6591-6.
48. Russell MR, Lilley KS. Pipeline to assess the greatest source of technical variance in quantitative proteomics using metabolic labelling. *J Proteomics.* **2012** Dec 21; 77: 441-54.
49. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci USA.* **2003** Jun 10; 100(12): 6940-5.
50. Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat Methods.* **2005** Aug; 2(8): 587-9.
51. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem.* **2007** Oct; 389(4): 1017-31.
52. Hsu JL, Huang SY, Chow NH, Chen SH. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem.* **2003** Dec 15; 75(24): 6843-52.
53. Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C. Proteolytic <sup>18</sup>O labelling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem.* **2001** Jul 1; 73(13): 2836-42.

54. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol.* **1999** Oct; 17(10): 994-9.
55. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics.* **2004** Dec; 3(12): 1154-69.
56. Choe L, D'Ascenzo M, Relkin NR, Pappin D, Ross P, Williamson B, Guertin S, Pribil P, Lee KH. 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics.* **2007** Oct; 7(20): 3651-60.
57. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Johnstone R, Mohammed AK, Hamon C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem.* **2003** Apr 15; 75(8): 1895-904.
58. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, Burkhard PR, Sanchez JC. Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal Chem.* **2008** Apr 15; 80(8): 2921-31.
59. Werner T, Becher I, Sweetman G, Doce C, Savitski MM, Bantscheff M. High-resolution enabled TMT 8-plexing. *Anal Chem.* **2012** Aug 21; 84(16): 7188-94.
60. McAlister GC, Huttlin EL, Haas W, Ting L, Jedrychowski MP, Rogers JC, Kuhn K, Pike I, Grothe RA, Blethrow JD, Gygi SP. Increasing the multiplexing capacity of TMTs using reporter ion isotopologues with isobaric masses. *Anal Chem.* **2012** Sep 4; 84(17): 7469-78.
61. Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, Lee A, van Sluyter SC, Haynes PA. Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics.* **2011** Feb; 11(4): 535-53.
62. Higgs RE, Knierman MD, Gelfanova V, Butler JP, Hale JE. Comprehensive label-free method for the relative quantification of proteins from biological samples. *J Proteome Res.* **2005** Jul-Aug; 4(4): 1442-50.
63. Cappadona S, Baker PR, Cutillas PR, Heck AJ, van Breukelen B. Current challenges in software solutions for mass spectrometry-based quantitative proteomics. *Amino Acids.* **2012** Sep; 43(3): 1087-108.
64. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem.* **2004** Jul 15; 76(14): 4193-201.
65. Hoehenwarter W, Wienkoop S. Spectral counting robust on high mass accuracy mass spectrometers. *Rapid Commun Mass Spectrom.* **2010** Dec 30; 24(24): 3609-14.
66. Cooper B, Feng J, Garrett WM. Relative, label-free protein quantitation: spectral counting error statistics from nine replicate MudPIT samples. *J Am Soc Mass Spectrom.* **2010** Sep; 21(9): 1534-46.
67. Lundgren DH, Hwang SI, Wu L, Han DK. Role of spectral counting in quantitative proteomics. *Expert Rev Proteomics.* **2010** Feb; 7(1): 39-53.
68. Manning G, Plowman GD, Hunter T, Sudarsanam S. Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci.* **2002** Oct; 27(10): 514-20.
69. Choudhary C, Mann M. Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol.* **2010**; 11: 427-439.
70. Dix MM, Simon GM, Wang C, Okerberg E, et al. Functional interplay between caspase cleavage and phosphorylation sculpts the apoptotic proteome. *Cell.* **2012**; 150: 426-440.
71. Doll S, Burlingame AL. Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem Biol.* **2015** Jan 16; 10(1): 63-71.



72. Gnad F, de Godoy LM, Cox J, Neuhauser N, Ren S, Olsen JV, Mann M. High-accuracy identification and bioinformatic analysis of in vivo protein phosphorylation sites in yeast. *Proteomics*. **2009** Oct; 9(20): 4642-52.
73. Jenuwein T, Allis CD. Translating the histone code. *Science*. **2001** Aug 10; 293(5532): 1074-80.
74. Kornberg RD, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*. **1999** Aug 6; 98(3): 285-94.
75. Vempati RK, Jayani RS, Notani D, Sengupta A, Galande S, Haldar D. p300-mediated acetylation of histone H3 lysine 56 functions in DNA damage response in mammals. *J Biol Chem*. **2010** Sep 10; 285(37): 28553-64.
76. Kurdistani SK, Grunstein M. Histone acetylation and deacetylation in yeast. *Nat Rev Mol Cell Biol*. **2003** Apr; 4(4): 276-84.
77. Syka JE, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, Ueberheide B, Garcia B, Busby S, Muratore T, Shabanowitz J, Hunt DF. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J Proteome Res*. **2004** May; 3(3): 621-6.
78. Beausoleil SA, Villén J, Gerber SA, Rush J, Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol*. **2006** Oct; 24(10): 1285-92.
79. Savitski MM, Lemeer S, Boesche M, Lang M, Mathieson T, Bantscheff M, Kuster B. Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics*. **2011** Feb; 10(2): M110.003830.
80. Chalkley RJ, Clauser KR. Modification Site Localization Scoring: Strategies and Performance. *Mol Cell Proteomics*. **2012**; 11(5): 3-14.
81. Martin DM, Nett IR, Vandermoere F, Barber JD, Morrice NA, Ferguson MA. Prophossi: automating expert validation of phosphopeptide-spectrum matches from tandem mass spectrometry. *Bioinformatics*. **2010** Sep 1; 26(17): 2153-9.
82. Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, Jacobsen SE. Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res*. **2004** Dec 14; 32(22): 6511-8.
83. Zhang K, Sridhar VV, Zhu J, Kapoor A, Zhu JK. Distinctive core histone post-translational modification patterns in Arabidopsis thaliana. *PLoS One*. **2007** Nov 21; 2(11): e1210.
84. Earley KW, Shook MS, Brower-Toland B, Hicks L, Pikaard CS. In vitro specificities of Arabidopsis co-activator histone acetyltransferases: implications for histone hyperacetylation in gene activation. *Plant J*. **2007** Nov; 52(4): 615-26.
85. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*. **2011** Oct 30; 480(7376): 254-8.
86. Catherman AD, Durbin KR, Ahlf DR, Early BP, Fellers RT, Tran JC, Thomas PM, Kelleher NL. Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Mol Cell Proteomics*. **2013** Dec; 12(12): 3465-73.
87. Durbin KR, Fornelli L, Fellers RT, Doubleday PF, Narita M, Kelleher NL. Quantitation and Identification of Thousands of Human Proteoforms below 30 kDa. *J Proteome Res*. **2016** Mar 4; 15(3): 976-82.
88. Smith KT, Workman JL. Chromatin proteins: key responders to stress. *PLoS Biol*. **2012**; 10(7): e1001371.

89. Gupta D, Eldakak M, Rohila JS, Basu C. Biochemical analysis of 'kerosene tree' *Hymenaea courbaril* L. under heat stress. *Plant Signal Behav.* **2014**; 9(10): e972851.
90. Sidler C, Li D, Kovalchuk O, Kovalchuk I. Development-dependent expression of DNA repair genes and epigenetic regulators in *Arabidopsis* plants exposed to ionizing radiation. *Radiat Res.* **2015** Feb; 183(2): 219-32.
91. Donà M, Mittelsten Scheid O. DNA Damage Repair in the Context of Plant Chromatin. *Plant Physiol.* **2015** Aug; 168(4): 1206-18.
92. Waterworth WM, Drury GE, Blundell-Hunter G, West CE. *Arabidopsis* TAF1 is an MRE11-interacting protein required for resistance to genotoxic stress and viability of the male gametophyte. *Plant J.* **2015** Nov; 84(3): 545-57.
93. Guo P, Qi YP, Yang LT, Ye X, Huang JH, Chen LS. Long-Term Boron-Excess-Induced Alterations of Gene Profiles in Roots of Two Citrus Species Differing in Boron-Tolerance Revealed by cDNA-AFLP. *Front Plant Sci.* **2016** Jun; 7: 898.
94. Hotez PJ, Brindley PJ, Bethony JM, King CH, Pearce EJ, Jacobson J. Helminth infections: the great neglected tropical diseases. *J Clin Invest.* **2008** Apr; 118(4): 1311-21.
95. Maizels RM, Yazdanbakhsh M. Immune regulation by helminth parasites: cellular and molecular mechanisms. *Nat Rev Immunol.* **2003** Sep; 3(9): 733-44.
96. Adisakwattana P, Saunders SP, Nel HJ, Fallon PG. Helminth-derived immunomodulatory molecules. *Adv Exp Med Biol.* **2009**; 666: 95-107.
97. Hewitson JP, Filbey KJ, Grainger JR, Dowle AA, Pearson M, Murray J, Harcus Y, Maizels RM. *Heligmosomoides polygyrus* elicits a dominant nonprotective antibody response directed against restricted glycan and peptide epitopes. *J Immunol.* **2011** Nov 1; 187(9): 4764-77.
98. Monroy FG, Enriquez FJ. *Heligmosomoides polygyrus*: a model for chronic gastrointestinal helminthiasis. *Parasitol Today.* **1992** Feb; 8(2): 49-54.
99. Ghedin E, Wang S, Spiro D, Caler E, Zhao Q, Crabtree J, Allen JE, Delcher AL, Guiliano DB, Miranda-Saavedra D, Angiuoli SV, Creasy T, Amedeo P, Haas B, El-Sayed NM, Wortman JR, Feldblyum T, Tallon L, Schatz M, Shumway M, Koo H, Salzberg SL, Schobel S, Perlea M, Pop M, White O, Barton GJ, Carlow CK, Crawford MJ, Daub J, Dimmic MW, Estes CF, Foster JM, Ganatra M, Gregory WF, Johnson NM, Jin J, Komuniecki R, Korf I, Kumar S, Laney S, Li BW, Li W, Lindblom TH, Lustigman S, Ma D, Maina CV, Martin DM, McCarter JP, McReynolds L, Mitreva M, Nutman TB, Parkinson J, Peregrín-Alvarez JM, Poole C, Ren Q, Saunders L, Sluder AE, Smith K, Stanke M, Unnasch TR, Ware J, Wei AD, Weil G, Williams DJ, Zhang Y, Williams SA, Fraser-Liggett C, Slatko B, Blaxter ML, Scott AL. Draft genome of the filarial nematode parasite *Brugia malayi*. *Science.* **2007** Sep 21; 317(5845): 1756-60.
100. Chevallet M, Luche S, Rabilloud T. Silver staining of proteins in polyacrylamide gels. *Nat Protoc.* **2006**; 1(4): 1852-8.
101. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics.* **2005** Sep; 4(9): 1265-72.
102. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. Global quantification of mammalian gene expression control. *Nature.* **2011** May 19; 473(7347): 337-42.
103. Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics.* **2006** Jan; 5(1): 144-56.
104. Gan CS, Chong PK, Pham TK, Wright PC. Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ). *J Proteome Res.* **2007** Feb; 6(2): 821-7.
105. Russell MR, Lilley KS. Pipeline to assess the greatest source of technical variance in quantitative proteomics using metabolic labelling. *J Proteomics.* **2012** Dec 21; 77: 441-54.

106. Shen X, Hu Q, Li J, Wang J, Qu J. Experimental Null Method to Guide the Development of Technical Procedures and to Control False-Positive Discovery in Quantitative Proteomics. *J Proteome Res.* **2015** Oct 2; 14(10): 4147-57.
107. Hewitson JP, Rückerl D, Harcus Y, Murray J, Webb LM, Babayan SA, Allen JE, Kurniawan A, Maizels RM. The secreted triose phosphate isomerase of *Brugia malayi* is required to sustain microfilaria production in vivo. *PLoS Pathog.* **2014** Feb 27; 10(2): e1003930.
108. Kim JY, Cho MK, Choi SH, Lee KH, Ahn SC, Kim DH, Yu HS. Inhibition of dextran sulfate sodium (DSS)-induced intestinal inflammation via enhanced IL-10 and TGF-beta production by galectin-9 homologues isolated from intestinal parasites. *Mol Biochem Parasitol.* **2010** Nov; 174(1): 53-61.
109. Maldonado RF, Sá-Correia I, Valvano MA. Lipopolysaccharide modification in Gram-negative bacteria during chronic infection. *FEMS Microbiol Rev.* **2016** Apr 12
110. Elkins KL, Cowley SC, Bosio CM. Innate and adaptive immunity to *Francisella*. *Ann N Y Acad Sci.* **2007** Jun; 1105: 284-324.
111. Cervi L, MacDonald AS, Kane C, Dzierszinski F, Pearce EJ. Cutting edge: dendritic cells coupled with microbial and helminth antigens undergo modified maturation, segregate the antigens to distinct intracellular compartments, and concurrently induce microbe-specific Th1 and helminth-specific Th2 responses. *J Immunol.* **2004**; 172: 2016–20.
112. Ferret-Bernard S, Curwen RS, Mountford AP. Proteomic profiling reveals that Th2-inducing dendritic cells stimulated with helminth antigens have a 'limited maturation' phenotype. *Proteomics* **2008**; 8: 980–93.
113. Cohen SL, Chait BT. Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Anal Chem.* **1996** Jan 1; 68(1): 31-7.
114. Spooncer E, Brouard N, Nilsson SK, Williams B, Liu MC, Unwin RD, et al. Developmental fate determination and marker discovery in hematopoietic stem cell biology using proteomic fingerprinting. *Mol Cell Proteomics* **2008**; 7: 573–81.
115. Suckau D, Resemann A, Schuereberg M, Hufnagel P, Franzen J, Holle A. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal Bioanal Chem.* **2003** Aug; 376(7): 952-65.
116. Ting L, Rad R, Gygi SP, Haas W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods.* **2011** Oct 2; 8(11): 937-40.
117. McAlister GC, Nusinow DP, Jedrychowski MP, Wühr M, Huttlin EL, Erickson BK, Rad R, Haas W, Gygi SP. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal Chem.* **2014** Jul 15; 86(14): 7150-8.
118. Levin Y. The role of statistical power analysis in quantitative proteomics. *Proteomics.* **2011** Jun; 11(12): 2565-7.
119. Luo R, Colangelo CM, Sessa WC, Zhao H. Bayesian Analysis of iTRAQ Data with Nonrandom Missingness: Identification of Differentially Expressed Proteins. *Stat Biosci.* **2009** Nov; 1(2): 228-245
120. Goeminne LJ, Argentini A, Martens L, Clement L. Summarization vs Peptide-Based Models in Label-Free Quantitative Proteomics: Performance, Pitfalls, and Data Analysis Guidelines. *J Proteome Res.* **2015** Jun 5; 14(6): 2457-65.
121. Dowle AA, Wilson J, Thomas JR. Comparing the Diagnostic Classification Accuracy of iTRAQ, Peak-Area, Spectral-Counting, and empAI Methods for Relative Quantification in Expression Proteomics. *J Proteome Res.* **2016** Oct 7; 15(10): 3550-3562.
122. Christoforou A, Arias AM, Lilley KS. Determining protein subcellular localization in mammalian cell culture with biochemical fractionation and iTRAQ 8-plex quantification. *Methods Mol Biol.* **2014**; 1156: 157-74.

123. Maekawa S, Kobayashi Y, Odagaki S, Makino M, Kumanogoh H, Nakamura S, Morita M, Hayashi F. Interaction of NAP-22 with brain glutamic acid decarboxylase (GAD). *Neurosci Lett*. **2013** Mar 14; 537: 50-4.
124. Plekhanov AY, Antonova OS, Petrova EI, Reznik SY, Klyueva NZ. The changes in metabolism of the regulatory brain protein NAP-22 at the early stages of postnatal ontogeny in spontaneous hypertensive and WKY rats born to females fed with calcium-deficient diet. *Dokl Biol Sci*. **2013** Sep; 452: 261-5.
125. Srivastava L, Tundup S, Choi BS, Norberg T, Harn D. Immunomodulatory glycan lacto-N-fucopentaose III requires clathrin-mediated endocytosis to induce alternative activation of antigen-presenting cells. *Infect Immun*. **2014** May; 82(5): 1891-903.
126. Lu R, Drubin DG, Sun Y. Clathrin-mediated endocytosis in budding yeast at a glance. *J Cell Sci*. 2016 Apr 15; 129(8): 1531-6.
127. Kliková K, Pilchova I, Stefanikova A, Hatok J, Dobrota D, Racay P. The Role of Heat Shock Proteins in Leukemia. *Klin Onkol*. **2016**; 29(1): 29-38.
128. Wu Q, Yuan H, Zhang L, Zhang Y. Recent advances on multidimensional liquid chromatography-mass spectrometry for proteomics: from qualitative to quantitative analysis – a review. *Anal Chim Acta*. **2012** Jun 20; 731: 1-10.
129. Hopkins RA, Connolly JE. The specialized roles of immature and mature dendritic cells in antigen cross-presentation. *Immunol Res*. **2012** Sep; 53(1-3): 91-107.
130. Shen XL, Zhang Y, Xu W, Liang R, Zheng J, Luo Y, Wang Y, Huang K. An iTRAQ-based mitoproteomics approach for profiling the nephrotoxicity mechanisms of ochratoxin A in HEK 293 cells. *J Proteomics*. **2013** Jan 14; 78: 398-415.
131. Hansson J, Krijgsveld J. Proteomic analysis of cell fate decision. *Curr Opin Genet Dev*. **2013** Oct; 23(5): 540-7.
132. Srivastava L, Tundup S, Choi BS, Norberg T, Harn D. Immunomodulatory glycan lacto-N-fucopentaose III requires clathrin-mediated endocytosis to induce alternative activation of antigen-presenting cells. *Infect Immun*. **2014** May; 82(5): 1891-903.
133. Liang R, Shen XL, Zhang B, Li Y, Xu W, Zhao C, Luo Y, Huang K. Apoptosis signal-regulating kinase 1 promotes Ochratoxin A-induced renal cytotoxicity. *Sci Rep*. **2015** Jan 28; 5: 8078.
134. Sethi S, Chourasia D, Parhar IS. Approaches for targeted proteomics and its potential applications in neuroscience. *J Biosci*. **2015** Sep; 40(3): 607-27.
135. Sanin DE, Prendergast CT, Mountford AP. IL-10 Production in Macrophages Is Regulated by a TLR-Driven CREB-Mediated Mechanism That Is Linked to Genes Involved in Cell Metabolism. *J Immunol*. **2015** Aug 1; 195(3): 1218-32.
136. Mo Y, Hou H, Li D, Liang Y, Chen D, Zhou Y. Mitochondrial protein targets of radiosensitisation by 1,8-dihydroxy-3-acetyl-6-methyl-9,10-anthraquinone on nasopharyngeal carcinoma cells. *Eur J Pharmacol*. **2014** Sep 5; 738: 133-41.
137. Everts B, Pearce EJ. Metabolic control of dendritic cell activation and function: recent advances and clinical implications. *Front Immunol*. **2014** May 8; 5: 203.
138. Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J. tRNADB 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res*. **2009** Jan; 37: D159-62.
139. Agris PF. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep*. **2008** Jul; 9(7): 629-35.
140. Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K, McCann PP. S-adenosylmethionine and methylation. *FASEB J*. **1996** Mar; 10(4): 471-80.

141. Tehlivets O, Hasslacher M, Kohlwein SD. S-adenosyl-L-homocysteine hydrolase in yeast: key enzyme of methylation metabolism and coordinated regulation with phospholipid synthesis. *FEBS Lett.* **2004** Nov 19; 577(3): 501-6.
142. Tehlivets O, Malanovic N, Visram M, Pavkov-Keller T, Keller W. S-adenosyl-L-homocysteine hydrolase and methylation disorders: yeast as a model system. *Biochim Biophys Acta.* **2013** Jan; 1832(1): 204-15.
143. Patching SG. Surface plasmon resonance spectroscopy for characterisation of membrane protein-ligand interactions and its potential for drug discovery. *Biochim Biophys Acta.* **2014** Jan; 1838 (1 Pt A): 43-55.
144. Jecklin MC, Schauer S, Dumelin CE, Zenobi R. Label-free determination of protein-ligand binding constants using mass spectrometry and validation using surface plasmon resonance and isothermal titration calorimetry. *J Mol Recognit.* **2009** Jul-Aug; 22(4): 319-29.
145. Stafford WF 3rd. Protein-protein and ligand-protein interactions studied by analytical ultracentrifugation. *Methods Mol Biol.* **2009**; 490: 83-113.
146. Vander Meulen KA, Horowitz S, Trievel RC, Butcher SE. Measuring the Kinetics of Molecular Association by Isothermal Titration Calorimetry. *Methods Enzymol.* **2016**; 567: 181-213.
147. Chen M, Yu J, Tanaka Y, Tanaka M, Tanaka I, Yao M. Structure of dihydrouridine synthase C (DusC) from *Escherichia coli*. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* **2013** Aug; 69 (Pt 8): 834-8.
148. Sousa FL, Martin WF. Biochemical fossils of the ancient transition from geoeconomics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochim Biophys Acta.* **2014** Jul; 1837(7): 964-81.
149. Björk GR, Hagervall TG. Transfer RNA Modification: Presence, Synthesis, and Function. *EcoSal Plus.* **2014** May; 6(1).
150. Chen X, Wang X, Feng J, Chen Y, Fang Y, Zhao S, Zhao A, Zhang M, Liu L. Structural insights into the catalytic mechanism of *Synechocystis* magnesium protoporphyrin IX O-methyltransferase (ChIM). *J Biol Chem.* **2014** Sep 12; 289(37): 25690-8.
151. Sakai Y, Miyauchi K, Kimura S, Suzuki T. Biogenesis and growth phase-dependent alteration of 5-methoxycarbonylmethoxyuridine in tRNA anticodons. *Nucleic Acids Res.* **2016** Jan 29; 44(2): 509-23.
152. Hawkins KM, Smolke CD. Production of benzyloquinoline alkaloids in *Saccharomyces cerevisiae*. *Nat Chem Biol.* **2008** Sep; 4(9): 564-73.
153. Fossati E, Ekins A, Narcross L, Zhu Y, Falgout JP, Beaudoin GA, Facchini PJ, Martin VJ. Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydrosanguinarine in *Saccharomyces cerevisiae*. *Nat Commun.* **2014**; 5: 3283.
154. Fossati E, Narcross L, Ekins A, Falgout JP, Martin VJ. Synthesis of Morphinan Alkaloids in *Saccharomyces cerevisiae*. *PLoS One.* **2015** Apr 23; 10(4): e0124459.
155. Sertürner, F. Darstellung der reinen Mohnsäure (Opiumsäure) nebst einer Chemischen Untersuchung des Opiums mit vorzüglicher Hinsicht auf einer darin neu entdeckten Stoff und die dahin gehörigen Bemerkungen *J. Pharmacie* **1806**; 14: 47-93
156. ten Have S, Boulon S, Ahmad Y, Lamond AI. Mass spectrometry-based immuno-precipitation proteomics - the user's guide. *Proteomics.* **2011** Mar; 11(6): 1153-9.
157. Vikis HG, Guan KL. Glutathione-S-transferase-fusion based assays for studying protein-protein interactions. *Methods Mol Biol.* **2004**; 261: 175-86.
158. Floris F, van Agthoven M, Chiron L, Soulby AJ, Wootton CA, Lam YP, Barrow MP, Delsuc MA, O'Connor PB. 2D FT-ICR MS of Calmodulin: A Top-Down and Bottom-Up Approach. *J Am Soc Mass Spectrom.* **2016** Sep; 27(9): 1531-8.

159. Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD. Complete biosynthesis of opioids in yeast. *Science*. **2015** Sep 4; 349(6252): 1095-100.
160. Grogan G, Turner NJ. InspiRED by Nature: NADPH-Dependent Imine Reductases (IREDs) as Catalysts for the Preparation of Chiral Amines. *Chemistry*. **2016** Feb; 22(6): 1900-1907.
161. Yang L, Yang C, Li C, Zhao Q, Liu L, Fang X, Chen XY. Recent advances in biosynthesis of bioactive compounds in traditional Chinese medicinal plants. *Sci Bull (Beijing)*. **2016**; 61: 3-17.
162. Narcross L, Fossati E, Bourgeois L, Dueber JE, Martin VJ. Microbial Factories for the Production of Benzylisoquinoline Alkaloids. *Trends Biotechnol*. **2016** Mar; 34(3): 228-41.
163. Schläger S, Dräger B. Exploiting plant alkaloids. *Curr Opin Biotechnol*. **2016** Feb; 37: 155-64.
164. Kries H, O'Connor SE. Biocatalysts from alkaloid producing plants. *Curr Opin Chem Biol*. **2016** Apr; 31: 22-30.
165. Pyne ME, Liu X, Moo-Young M, Chung DA, Chou CP. Genome-directed analysis of prophage excision, host defence systems, and central fermentative metabolism in *Clostridium pasteurianum*. *Sci Rep*. **2016** Sep 19; 6: 26228.
166. Suástegui M, Shao Z. Yeast factories for the production of aromatic compounds: from building blocks to plant secondary metabolites. *J Ind Microbiol Biotechnol*. **2016** Nov; 43(11):1611-1624.
167. Nakagawa A, Matsumura E, Koyanagi T, Katayama T, Kawano N, Yoshimatsu K, Yamamoto K, Kumagai H, Sato F, Minami H. Total biosynthesis of opiates by stepwise fermentation using engineered *Escherichia coli*. *Nat Commun*. **2016** Feb 5; 7: 10390.
168. Tan GY, Deng Z, Liu T. Recent advances in the elucidation of enzymatic function in natural product biosynthesis. Version 2. *F1000Res*. **2015** Dec 4 [revised 2016 Jan 1]; 4. pii: F1000 Faculty Rev-1399.
169. Geetha S, Srinivasan D, Jayapriya V, Jeevidha B. Molecular and Protein characterization of Human cytochrome P450 expressed Recombinant E.Coli and its application in the synthesis of Tramadol metabolite. *IJIPBART*. **2015**; 2(3): 237-248.
170. Xu G, Cai W, Gao W, Liu C. A novel glucuronosyltransferase has an unprecedented ability to catalyse continuous two-step glucuronosylation of glycyrrhetic acid to yield glycyrrhizin. *New Phytol*. **2016** Oct; 212(1): 123-35.
171. Nützmann HW, Huang A, Osbourn A. Plant metabolic clusters - from genetics to genomics. *New Phytol*. **2016** Aug; 211(3): 771-89.
172. Morris JS, Dastmalchi M, Li J, Chang L, Chen X, Hagel JM, Facchini PJ. Plug-and-Play Benzylisoquinoline Alkaloid Biosynthetic Gene Discovery in Engineered Yeast. *Methods Enzymol*. **2016**; 575: 143-78.
173. Morris JS, Facchini PJ. Isolation and Characterization of Reticuline N-Methyltransferase Involved in Biosynthesis of the Aporphine Alkaloid Magnoflorine in Opium Poppy. *J Biol Chem*. **2016** Sep 15. pii: jbc.M116.750893.
174. Pyne ME, Narcross L, Fossati E, Bourgeois L, Burton E, Gold ND, Martin VJ. Reconstituting Plant Secondary Metabolism in *Saccharomyces cerevisiae* for Production of High-Value Benzylisoquinoline Alkaloids. *Methods Enzymol*. **2016**; 575: 195-224.
175. Whitelaw DA, Tonkin R, Meints CE, Wolthers KR. Kinetic analysis of electron flux in cytochrome P450 reductases reveals differences in rate-determining steps in plant and mammalian enzymes. *Arch Biochem Biophys*. **2015** Oct 15; 584: 107-15.
176. Murata J, Matsumoto E, Morimoto K, Koyama T, Satake H. Generation of Triple-Transgenic *Forsythia* Cell Cultures as a Platform for the Efficient, Stable, and Sustainable Production of Lignans. *PLoS One*. **2015** Dec 7; 10(12): e0144519.



177. Tatsis EC, O'Connor SE. New developments in engineering plant metabolic pathways. *Curr Opin Biotechnol.* **2016** Apr 28; 42: 126-132.
178. Han M, Yin H, Zou Y, Brock NL, Huang T, Deng Z, Chu Y, Lin S. An Acyl Transfer Reaction Catalyzed by an Epimerase MarH. *ACS Catalysis.* **2016** Jan 5; 6(2): 788-92.