Chapter 3

The sites of palmitoylation in the *Plasmodium falciparum* palmitome

The *Plasmodium falciparum* palmitome

Recent advances in proteome-based technologies have allowed the characterisation of palmitoylated proteins to shift from studies of individual palmitoyl-proteins, to the global characterisation of whole palmitomes in particular organisms [1-3]. The two main techniques which have been used to successfully purify whole palmitomes are acyl-biotinyl exchange (ABE) [2] and metabolic labelling with click chemistry [3](described in detail in Section 1.6). Both these techniques were recently used to purify the palmitome of the asexual intraerythrocytic schizont stages of *Plasmodium falciparum*, and coupled with quantitative mass spectrometry, these purification methods revealed that there are at least 494 putative palmitoyl-proteins in blood-stage *P. falciparum* parasites [4].

However, these palmitome purifications techniques do not provide site-identifying information, and as noted in Section 1.5, there is no consensus sequence for palmitoylation. Therefore, although our understanding of palmitoylation in *P. falciparum* has increased with the purification of the *P. falciparum* palmitome, it is still unknown which cysteine residue(s) within these 494 palmitoyl-proteins are actually palmitoylated. Identification of the actual palmitoylated cysteine residues is important, as a global study of palmitoylation sites could potentially assist in our understanding of how individual PATs recognise their substrates. Furthermore, information on the sites of palmitoylation could be used to experimentally investigate the function of palmitoylation for a particular protein. Thus, the first aim of this project was to adapt current methods of palmitome purification to incorporate the identification of specific sites of palmitoylation on individual palmitoyl-proteins, in order to allow a global study of palmitoylation sites within the *P. falciparum* schizont palmitome.

3.1. Development of the site-identification (ID) palmitome purification method

3.1.1. Method used for the purification of the P. falciparum total schizont palmitome

In an attempt to increase the number of known palmitoylation sites in *P. falciparum*, a series of trials were performed in order to develop a site-identification (ID) palmitome purification method. The site-ID palmitome purification method was adapted from the ABE method of palmitoyl-protein purification (described in detail in Section 1.6.1), which was one of the techniques previously used to purify the total palmitome from *P. falciparum* schizonts [4]. ABE was the method of choice to adapt the site-ID purification method as all the chemical steps of ABE take place post-extraction of the proteome and does not rely on *in vitro* metabolic labelling. This would then allow the site-ID

purification method to be used to identify palmitoylation sites in other stages of the *Plasmodium* life cycle (such as the mosquito stages) or in other *Plasmodium* species (such as *P. berghei*), which are not amenable to the metabolic labelling required in the metabolic labelling and click chemistry method of palmitoyl-protein purification (described in Section 1.6.2).

Two key differences were present between the site-ID purification method and the original ABE purification method used in the purification of the total schizont palmitome. Firstly, the ABE purification of the total palmitome was coupled with stable isotope labelling with amino acids in cell culture (SILAC), where parasites were metabolically labelled with 'light' or 'heavy' versions of isoleucine in culture, in order to allow accurate protein quantification by mass spectrometry [4-6]. However, as mentioned above, it was intended that the site-ID purification method should be compatible for use in other stages of the *Plasmodium* life cycle, or in other Apicomplexan species, which would not be amenable to *in vitro* metabolic labelling. Thus, it was decided that SILAC labelling would not be used in the site-ID purifications and quantification by mass spectrometry would be performed using stable isotope dimethyl labelling [7], which is performed after the ABE purification is complete, or using label-free methods [8] (as will be described in detail below).

Secondly, for the purification of the total schizont palmitome, the proteome was extracted from SILAC labelled schizont stages and subjected to the different chemical steps of ABE, followed by streptavidin affinity purification and elution (as described in Section 1.6.1). The eluted proteins were then separated by SDS-PAGE and digested in-gel with trypsin, before analysis by high-resolution tandem mass spectrometry (LC-MS/MS) [4]. In order to identify palmitoylation sites, digestion with trypsin would have to take place before the elution of the samples from the streptavidin resin (instead of in-gel digestion after elution) so that only palmitoylated peptides are purified, and this was what was performed in the site-ID purification trials (as will be described in detail below).

During the development of the site-ID palmitome purification method, several different conditions were trialled in order to determine the optimum conditions required, and each of these trial purifications are described in the sections below. All trials of the site-ID purification method were performed on *P. falciparum* schizont stages, both because schizonts contain multiple individual merozoites, and therefore allow the generation of larger amounts of material than earlier stages of development where only a single parasite is present, and also to allow comparison of the site-ID datasets with the previously published schizont total palmitome dataset [4]. Although it is unlikely that the site-ID purification method would produce a dataset which completely overlapped with the total palmitome, comparison between the site-ID datasets and the total palmitome could still act as a rough guide for the efficiency and accuracy of the trial purifications.

The development and optimisation of the site-ID palmitome purification method described below was performed together with Mark Collins, Choudhary lab, and all quantitative mass spectrometry analysis was kindly performed by Mark Collins, Choudhary lab.

3.1.2. Development of the site-ID palmitome purification method – Trial 1

Briefly, ABE consists of three main sequential chemical steps which take place post-extraction of the proteome: (1) treatment of samples with NEM, which irreversibly blocks any free cysteine thiols present, (2) cleavage of the thioester bonds between palmitoyl groups and cysteine residues using hydroxylamine and (3) biotinylation of newly-exposed free thiols using the thiol-reactive biotinylation reagent, HPDP-biotin, which allows the streptavidin affinity purification of biotinylated proteins, that can be then eluted from streptavidin-agarose beads by reduction with a reducing agent. Importantly, the control treatment of the proteome, where an equal amount of the proteome is processed in parallel under identical conditions, but in the absence of hydroxylamine cleavage (during step 2), must always be present, as the quantitative measurement of the enrichment of a protein in the hydroxylamine-treated 'palmitome' samples over the non-hydroxylamine-treated 'control' samples determines whether the protein can be regarded as palmitoylated.

For the first trial of the site-ID palmitome purification, approximately 10^9 *P. falciparum* strain 3D7 schizont stage parasites (equivalent to 200 mL of parasite *in vitro* blood-stage culture at 5% haematocrit and 8-10% parasitemia) were extracted from infected erythrocytes by saponin lysis, as described in Section 2.2.5 of the Materials and Methods. The proteome was extracted from saponin-treated parasite pellets by lysis in a Triton X-100-based buffer and subjected to the three chemical steps of ABE, followed by streptavidin affinity purification, as described above and in Section 2.6.1 of the Materials and Methods. As before, an equal amount of the proteome was subjected to identical steps of ABE, but without hydroxylamine treatment (instead of hydroxylamine, an equal volume of 100 mM Tris-Cl pH 7.4 was added into the control samples). Before elution of the biotinylated palmitome and control samples from streptavidin-agarose beads, both samples were digested with trypsin while bound to the beads in order to digest the proteins into peptide fragments. The samples were then stringently washed so that unbound peptide fragments were washed away and only the biotinylated peptide fragments remained bound to streptavidin-agarose. The peptide fragments were then eluted from the beads using the reducing agent Tris(2-carboxyethyl)phosphine (TCEP). The trypsin digestion of the samples before elution ensured that only biotinylated peptide fragments

(which by definition contained the palmitoylated cysteine) were purified. The steps of Trial 1 of the site-ID palmitome purification are as shown in the schematic in Figure 3.1A.

The eluted peptide fragments were then subjected to stable isotope dimethyl labelling in order to allow quantitative mass spectrometry analysis of the peptide fragments. Stable isotope dimethyl labelling uses formaldehyde and cyanoborohydride to modify all primary amines (N-termini and the side chain of lysine residues). By using different isopotomers of formaldehyde and cyanoborohydride, the mass added to each sample can be varied. The labelled samples are then pooled and analysed by LC-MS/MS, where peptides from the different isotope-labelled samples can be identified due to their mass difference. Quantification is then performed by measuring the ratio of the peak intensities of the isotope-labelled peptide pairs [7, 9]. For Trial 1 of the site-ID purification method, non-hydroxylamine-treated control samples were treated with 'medium' formaldehyde (deuterated formaldehyde) and sodium cyanoborohydride, and hydroxylamine-treated palmitome samples were treated with 'heavy' formaldehyde (^{13}C-deuterated formaldehyde) and sodium cyanoborodeuteride. The samples were analysed by LC-MS/MS and MaxQuant [10] was used to generate specific enrichment ratios for each of the peptide identifications, where the enrichment ratio was the intensity of the peptide in the palmitome (hydroxylamine-treated) sample compared to the intensity of the peptide in the control (non-hydroxylamine-treated) sample.

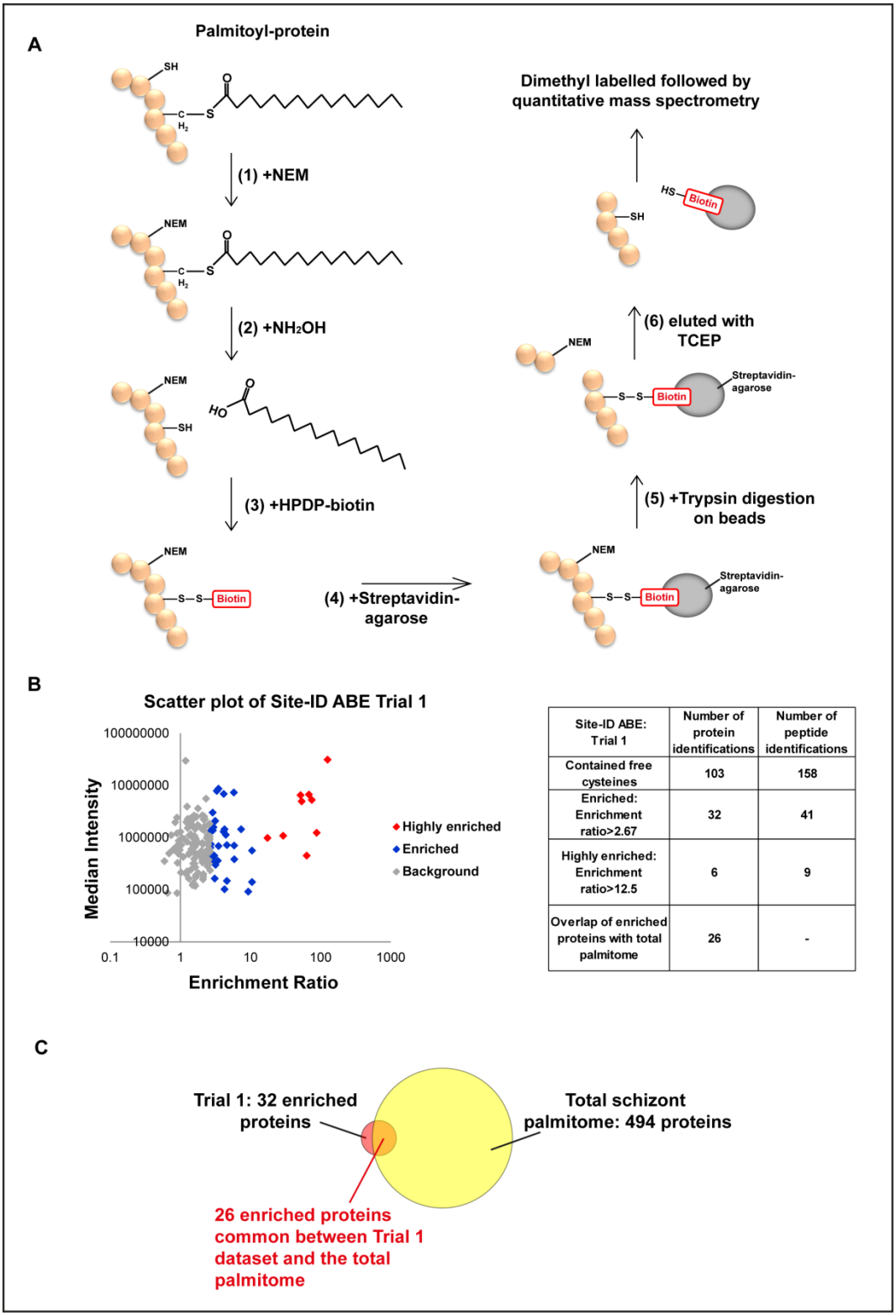
In total, 158 peptides, which had enrichment ratios and contained a free cysteine, were identified. This corresponded to a total of 103 protein identifications, indicating that some proteins had multiple peptides which contained a palmitoylated cysteine, suggesting that some proteins may be multiply-palmitoylated. Alternatively, the same peptide fragment could also be identified more than once, perhaps due to the abundance of the protein, and this also contributed to the higher number of peptide identifications compared to protein identifications. The 158 peptides identified were then grouped into enriched or highly enriched classes based on the same cut-off criteria which was used previously in the analysis of the total schizont palmitome [4]:

- Enriched peptides: All peptides with an enrichment ratio above the median enrichment ratio plus the median of the absolute deviation of each individual enrichment ratio from the median enrichment ratio.
- Highly enriched peptides: All peptides with an enrichment ratio above the mean enrichment ratio plus the mean of the absolute deviation of each individual enrichment ratio from the mean enrichment ratio.

Using these cut-off criteria, 41 peptides (corresponding to 32 protein identifications) were defined as enriched, and out of these 41 peptides, 9 peptides (corresponding to 6 protein identifications) could be further defined as highly enriched. A scatter plot showing the median peptide intensities against the MaxQuant generated enrichment ratios for all 158 identified peptides are shown in Figure 3.1B, with peptides classified as enriched and highly enriched highlighted in blue and red respectively.

Of the 32 proteins which had peptide identifications classified as enriched, 26 proteins were also present in the total schizont palmitome [4]. This indicated that 81% of the proteins classified as enriched in this dataset (26 out of 32 proteins) were also present in the total palmitome (Figure 3.1C). As mentioned above, a complete overlap between the site-ID purifications and the total palmitome is not expected. This is firstly due to the fact that the site-ID purification was based solely on ABE, whereas the total palmitome was compiled from multiple replicates of both ABE and metabolic labelling with click chemistry purifications, both of which purify overlapping as well as unique sets of proteins. Secondly, in the site-ID purifications, individual peptide identifications were analysed, whereas in the total palmitome purification, protein identifications were analysed, and each protein was identified by contributions from multiple peptide fragments. Nevertheless, the presence of 81% of the enriched proteins of this site-ID dataset in the total palmitome was encouraging.

Included in the overlap list were highly studied palmitoylated proteins such as glideosome-associated protein 45 (GAP45), chloroquine resistance transporter (CRT) and myosin A tail domain interacting protein (MTIP) [4]. This, along with the good overlap between enriched proteins in this dataset and the total palmitome, indicated that in this first trial of the site-ID palmitome purification, palmitoylated proteins were indeed being purified and the use of this method resulted in the identification of specific palmitoylated peptide fragments, containing the putative palmitoylated cysteines, for each of the enriched proteins purified.



Site-ID ABE: Trial 1	Number of protein identifications	Number of peptide identifications
Contained free cysteines	103	158
Enriched: Enrichment ratio>2.67	32	41
Highly enriched: Enrichment ratio>12.5	6	9
Overlap of enriched proteins with total palmitome	26	-

Figure 3.1: Trial 1 site-ID palmitome purification by ABE. (A) The site-ID palmitome purification was based on the ABE method of palmitoyl-protein purification. In the Trial 1 site-ID purification, the proteome extracted from saponin-treated parasite pellets was (1) first treated with NEM to irreversibly block free thiols. This was followed by (2) cleavage of palmitate groups using hydroxylamine (NH_2OH) and the (3) biotinylation of newly-exposed free thiols with HPDP-biotin. As always, a 'control' sample, which was subjected to identical conditions as the 'palmitome' sample, but without hydroxylamine treatment, was also prepared. Biotinylated proteins were purified by (4) streptavidin affinity purification, and (5) trypsin digestion was performed while the

samples were bound to streptavidin-agarose beads, in order to digest the proteins into peptide fragments. After stringent washes, only biotinylated peptide fragments remained bound to the beads and were (6) eluted by reduction with TCEP, followed by dimethyl labelling and quantitative mass spectrometry analysis. **(B)** Scatter plot displaying the median protein intensities against the MaxQuant generated enrichment ratios for all the peptides purified by the Trial 1 site-ID purification. Peptides classified as enriched and highly enriched (according to the cut-off criteria described in the text) are highlighted in blue and red respectively. The numbers of protein and peptide identifications found are listed in the table on the right. **(C)** Venn diagram representing the overlap between the enriched proteins in the Trial 1 dataset and the total schizont palmitome.

However, as described in Section 1.6.3, a number of false positive identifications are unavoidable, and were probably present in this dataset, even in the proteins defined as enriched. This is due firstly to the ABE method itself, which enriches other proteins that also use thioester linkages, and is highly dependent on the complete and irreversible blockage of free thiols before hydroxylamine cleavage (detailed in Section 1.6.3). Secondly, false positive identifications can also occur due to the inherent background that occurs in large-scale protein purifications and in streptavidin bead enrichments.

Additionally, cysteine-containing peptides were also directly identified by MaxQuant in the control (non-hydroxylamine-treated) samples. A low level of cysteine-containing peptides are expected to be present in control samples, and in fact, relative proteomic quantitation and the generation of the enrichment ratios is based on the presence of these cysteine-containing-peptides in the control samples. Typically, these peptides should usually be at a very low abundance in control samples, and can be regarded as 'noise'. The identification of a particular peptide ion by MaxQuant in the palmitome sample will then result in a search by MaxQuant for the corresponding peptide ion in the noise of the control sample, based on matching the mass and retention time of the peptide ion, in order to generate an enrichment ratio. The peptide intensities in the noise of the control samples are usually too low to be directly identified by MaxQuant. The fact that some cysteine-containing-peptides were being directly sequenced and identified by MaxQuant in the control samples of this purification however indicates that there is a background of cysteine-containing peptides present in the control samples, which is higher than the normal noise usually found in control samples. This suggests that the initial treatment with NEM to block free thiols was not efficient or may not have gone to completion. The presence of this high background of cysteine-containing peptides in the control samples could have an effect on the enrichment ratios generated, and this could subsequently have an effect on the definition of the cut-off criteria and the classification of proteins as enriched, potentially increasing the number of false-positive identifications.

In summary, although palmitoylated proteins with corresponding cysteine-containing peptides were identified in this first trial, the number of enriched protein identifications was relatively small compared to the total number of proteins in the *P. falciparum* proteome (there are over 5000 proteins coded for in the *P. falciparum* genome [11] with over 1000 proteins up-regulated in schizont stages [12]). Additionally, although 81% of the enriched proteins (26 out of 32 enriched proteins) were also present in the total schizont palmitome, this was equivalent to only 5% of the total schizont palmitome (26 out of 494 proteins in the total palmitome) (Figure 3.1C), indicating that many palmitoylated proteins were potentially not purified, or did not pass the cut-off criteria in this trial purification. Taken together, this suggested that the total complement of palmitoylated proteins was not being purified by this first trial purification and further optimisation was still required.

3.1.3. Development of the site-ID palmitome purification method – Trial 2

For Trial 2 of the site-ID palmitome purification, approximately 10^9 *P. falciparum* strain 3D7 schizont stage parasites (200 mL *in vitro* culture at 5% haematocrit and 8-10% parasitemia) was extracted from infected erythrocytes by saponin-lysis, as was done for Trial 1. However, in this trial purification, the extraction of the proteome from saponin-treated parasite pellets was performed more extensively (as described in Section 2.6.2 of the Materials and Methods), in order to ensure that the maximum amount of the parasite proteome was extracted. Briefly, lysis of the saponin-treated parasite pellets was performed using an SDS-based buffer instead of a Triton X-100-based buffer, and the lysate was also homogenised, and then sheered by passing through a fine gauge needle. Additionally, the insoluble material remaining from the initial lysis with SDS buffer was further treated with buffer containing 8 M urea in order to solubilise and denature the more hydrophobic proteins.

The initial irreversible blockage of free thiol groups in the proteome sample was performed using iodoacetamide (IAA), another sulphydryl-reactive alkylating agent, instead of NEM, in order to determine whether the blockage of free cysteine thiols occurred more efficiently and completely using IAA. The IAA-blocked proteome sample was then subjected to the remaining two chemical steps of ABE – cleavage of palmitoyl groups using hydroxylamine and biotinylation of newly-exposed free thiols with HPDP-biotin. In this trial purification however, the hydroxylamine-cleavage and biotinylation steps, as well as all the wash steps in between, were performed using 30 kDa molecular weight cut-off (MWCO) spin columns. For the purification of the total palmitome [4] and in Trial 1 of the site-ID purification, buffer exchange during washes and when changing to the different ABE

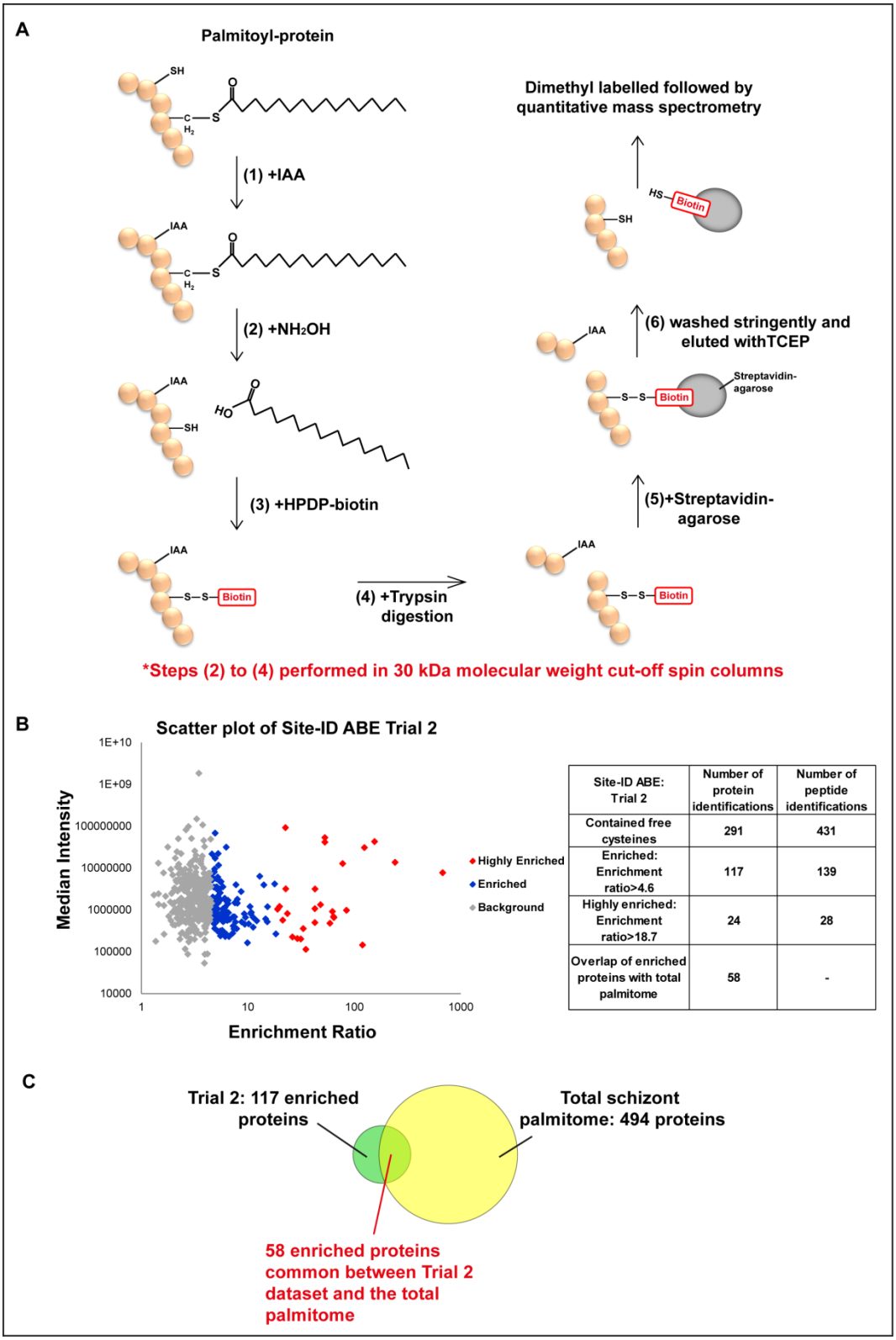
chemical steps, occurred by performing multiple chloroform-methanol (C/M) precipitations, which was time-consuming and increased the chances of sample loss between each C/M precipitation. The use of the 30 kDa MWCO spin columns here allowed buffer exchange to occur more easily and efficiently, and reduced the loss of samples between each wash and each chemical step. As before, the non-hydroxylamine-treated control was also prepared from an equal amount of the proteome, and this control sample was also treated in the 30 kDa MWCO spin columns. Additionally, all wash steps were performed in buffer containing 8 M urea, in order to ensure that the proteins were fully denatured and unfolded so that all palmitate groups would be exposed for hydroxylamine cleavage.

After biotinylation of the free thiols produced from hydroxylamine cleavage of palmitate groups, the samples were treated with trypsin in order to digest the proteins into peptide fragments. Trypsin digestion was also performed in the 30 kDa MWCO spin columns. Biotinylated peptide fragments were then purified by streptavidin affinity purification. The trypsin digestion of the protein samples before streptavidin affinity purification ensured that only biotinylated peptide fragments (which by definition contained the palmitoylated cysteine) would be bound by the streptavidin-agarose beads. Extensive and stringent wash steps were then performed in order to remove any unbound peptide fragments and the remaining bound palmitoylated peptide fragments were then eluted from the streptavidin-agarose beads using TCEP. The steps of Trial 2 of the site-ID palmitome purification are as shown in the schematic in Figure 3.2A.

The eluted peptide fragments were then subjected to stable isotope dimethyl labelling as before (described in Section 3.1.2) and analysed by LC-MS/MS. MaxQuant was used to generate specific enrichment ratios (palmitome over control) for each of the peptide identifications.

In total, 431 peptides containing enrichment ratios and a free cysteine were identified, and this corresponded to a total of 291 proteins. This set of 431 peptide identifications were then grouped into enriched and highly enriched classes based on the same cut-off criteria described in Section 3.1.2 above. Using these cut-off criteria, 139 peptides (corresponding to 117 protein identifications) were defined as enriched, and out of these 139 peptides, 28 peptides (corresponding to 24 protein identifications) could be further defined as highly enriched. A scatter plot showing the median peptide intensities against the MaxQuant generated enrichment ratios for all 431 peptide identifications are shown in Figure 3.2B. Of the 117 proteins which had peptide identifications classified as enriched, 58 proteins were also present in the total palmitome [4]. This indicated that 50% of the enriched proteins in this dataset (58 out of 117 proteins) were present in the total palmitome (Figure 3.2C). Again, proteins previously shown to be palmitoylated, such as GAP45 and CRT, were also present and classified as enriched in this trial purification.

Although only 50% of enriched proteins in this dataset overlapped with the total schizont palmitome, this was equivalent to 12% of the total palmitome (58 proteins out of 494 proteins in the total palmitome) (Figure 3.2C). This implied that more of the total palmitome was being purified in this trial compared to Trial 1, suggesting that more palmitoylated proteins were being purified or managed to pass the cut-off criteria in this purification. However, it also appeared a larger portion of the Trial 2 enriched dataset did not overlap with the total palmitome (50% of the Trial 2 enriched dataset overlapped with the total palmitome compared to 81% of the Trial 1 enriched dataset). As explained in Section 3.1.2 above, a complete overlap between the site-ID purifications and the total palmitome is not expected. This is further exacerbated for the Trial 2 purification where there were significant differences in the methodology (for example, different lysis conditions and treatment with IAA instead of NEM), and this could be the reason why only half of the Trial 2 enriched dataset overlapped with the total palmitome.



Site-ID ABE: Trial 2	Number of protein identifications	Number of peptide identifications
Contained free cysteines	291	431
Enriched: Enrichment ratio>4.6	117	139
Highly enriched: Enrichment ratio>18.7	24	28
Overlap of enriched proteins with total palmitome	58	-

Figure 3.2: Trial 2 site-ID palmitome purification by ABE. (A) In the Trial 2 site-ID purification, the proteome was extensively extracted from saponin-treated parasite pellets (as described in the text) and was (1) treated with IAA to irreversibly block free thiols. This was followed by (2) cleavage of palmitate groups using hydroxylamine (NH_2OH) and the (3) biotinylation of newly-exposed free thiols with HPDP-biotin. As always, a non-hydroxylamine-treated control sample was processed in parallel. (4) Trypsin digestion was then performed in order to digest the proteins into peptide fragments. Steps (2) to (4) were performed in 30 kDa molecular weight cut off (MWCO) spin columns in order to ease buffer exchange during washes and changes to different

chemical steps. Biotinylated peptides were purified by (5) streptavidin affinity purification, and unbound peptides removed by stringent washes. The biotinylated peptide fragments which remained bound to the beads were (6) eluted by reduction with TCEP, followed by dimethyl labelling and quantitative mass spectrometry analysis. **(B)** Scatter plot displaying the median protein intensities against the MaxQuant generated enrichment ratios for all the peptides purified by the Trial 2 site-ID purification. Peptides classified as enriched and highly enriched (according to the cut-off criteria described in the text) are highlighted in blue and red respectively. The numbers of protein and peptide identifications found are listed in the table on the right. **(C)** Venn diagram representing the overlap between the enriched proteins in the Trial 2 dataset and the total schizont palmitome.

In summary, the site-ID palmitome purification method used in this Trial 2 purification appeared to be an improved purification compared to that of Trial 1, as evidenced by the increased total number of peptides identified, the increased number of peptides classified as enriched, as well as the greater spread of enrichment ratios, and the identification of more proteins which were also present in the total schizont palmitome. However, cysteine-containing peptides were still directly identified by MaxQuant in the control samples, indicating that the background of cysteine-containing peptides, with intensities above the normal 'noise' usually found in control samples, was still present. This means that the use of IAA instead of NEM did not improve the blockage of free thiols. Additionally, the dataset of protein identifications was again smaller than total palmitome dataset, indicating that further improvements were still required.

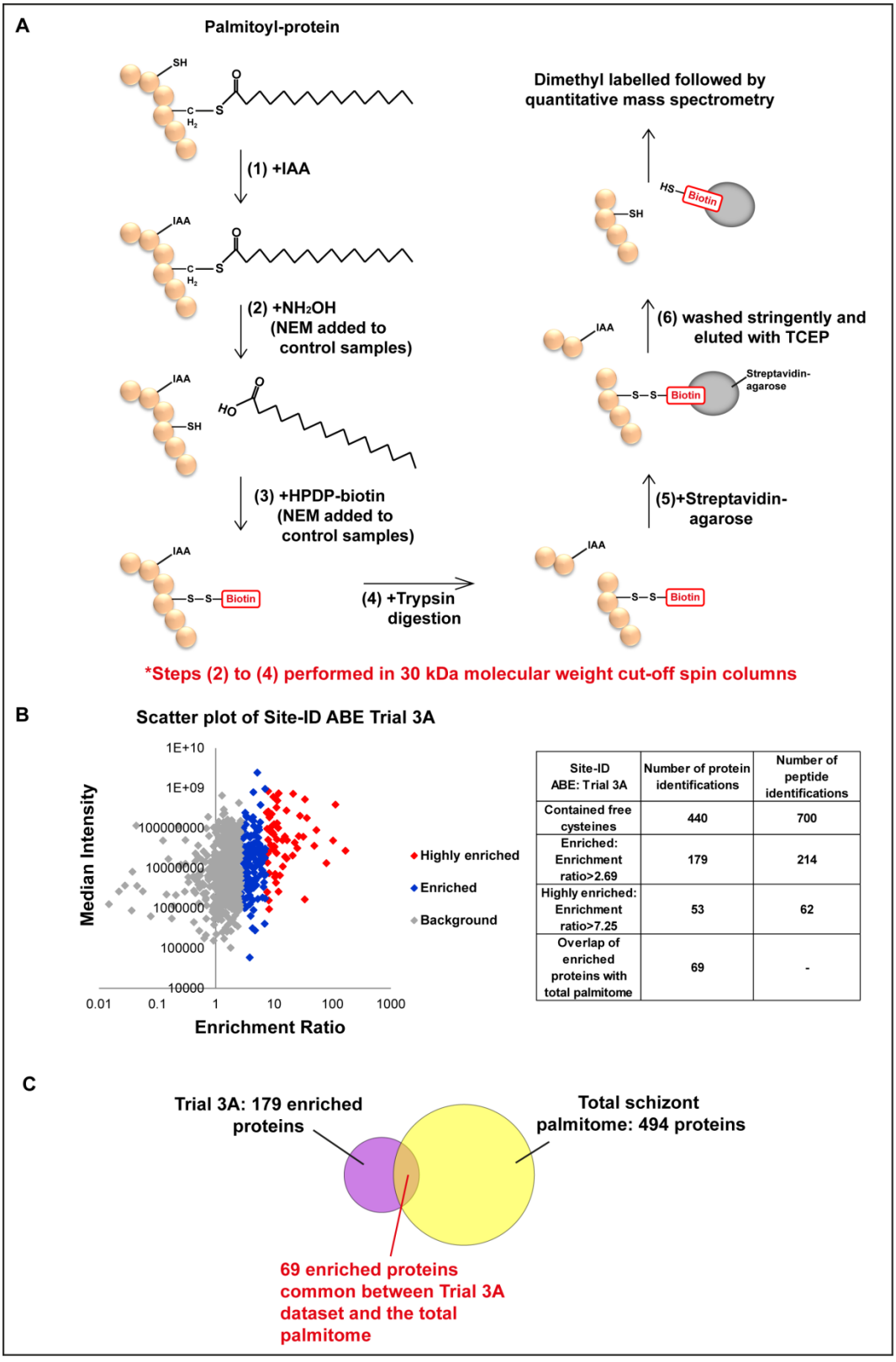
3.1.4. Development of the site-ID palmitome purification method – Trial 3A

For the next trial of the site-ID palmitome purification, approximately 10^9 *P. falciparum* strain 3D7 schizont stage parasites (200 mL *in vitro* culture at 5% haematocrit and 8-10% parasitemia) was extracted from infected erythrocytes by saponin-lysis as performed for Trial 1 and 2. Extraction of the proteome from saponin-treated parasite pellets was then performed exactly the same as in Trial 2 (described in Section 3.1.3). Initial blockage of free thiols in the proteome was performed using IAA and the hydroxylamine treatment and biotinylation steps of ABE performed in the 30 kDa MWCO spin columns, as in Trial 2 (Section 3.1.3). As always, the non-hydroxylamine-treated control sample was also prepared in parallel. In this trial purification however, during the hydroxylamine cleavage and biotinylation steps, NEM was added only into control samples, in an attempt to block any free thiols that may become exposed due to palmitate groups falling off during the ABE procedure. This was performed in the hopes that the background of cysteine-containing peptides observed in control

samples in the previous trial purifications could be reduced. Trypsin digestion of the samples in order to produce peptide fragments, streptavidin affinity purification of only biotinylated peptide fragments and elution of the peptides were all performed similar to Trial 2 (Section 3.1.2). The steps of Trial 3A of the site-ID palmitome purification are as shown in the schematic in Figure 3.3A.

Aliquots of the palmitome and control elutions from this Trial 3A purification were kept aside for label-free analysis in order to determine whether a label-free analysis would produce a better dataset (as will be described in the section below). The rest of the palmitome and control elutions were subjected to stable isotope dimethyl labelling (as for the previous trial purifications), and were analysed by LC-MS/MS. MaxQuant was used to generate the specific enrichment ratios (palmitome over control) for each of the peptide identifications.

In total, 700 peptides, which had enrichment ratios and contained a free cysteine, were identified, and this corresponded to a total of 440 protein identifications. This set of 700 peptides was grouped into enriched and highly enriched classes using the same cut-off criteria as above (detailed in Section 3.1.2). Using these cut-off criteria, 214 peptides (corresponding to 179 protein identifications) were defined as enriched, and out of these 214 peptides, 62 peptides (corresponding to 53 protein identifications) could be further defined as highly enriched. A scatter plot showing the median peptide intensities against the MaxQuant generated enrichment ratios for all 431 peptide identifications are shown in Figure 3.3B. Of the 179 proteins which had peptide identifications classified as enriched, 69 proteins were also present in the total palmitome. This indicated that 39% of the proteins classified as enriched in this dataset (69 out of 179 proteins) were present in the total palmitome, and this was equivalent to 13% of the total palmitome (69 proteins out of 494 proteins in the total palmitome) (Figure 3.3C).



Site-ID ABE: Trial 3A	Number of protein identifications	Number of peptide identifications
Contained free cysteines	440	700
Enriched: Enrichment ratio>2.69	179	214
Highly enriched: Enrichment ratio>7.25	53	62
Overlap of enriched proteins with total palmitome	69	-

Figure 3.3: Trial 3A site-ID palmitome purification by ABE. (A) The Trial 3A site-ID purification was similar to that of Trial 2. The proteome was extensively extracted from saponin-treated parasite pellets as in Trial 2, and protein samples were (1) blocked with IAA, (2) treated with hydroxylamine (NH_2OH) (except for control samples) and (3) biotinylated using HPDP-biotin. In this Trial 3A purification however, NEM was added only into control samples during steps (2) and (3) in an attempt to reduce the high background of cysteine-containing peptides seen in the control samples. (4) Trypsin digestion was then performed to digest the proteins into

peptide fragments. Steps (2) to (4) were performed in 30 kDa MWCO spin columns as before. Biotinylated peptides were purified by (5) streptavidin affinity purification, and unbound peptides removed by stringent washes. The biotinylated peptide fragments which remained bound to the beads were (6) eluted by reduction with TCEP. Aliquots of palmitome and control samples were kept aside for label-free analysis. The remaining palmitome and control samples were dimethyl labelled and analysed by quantitative mass spectrometry as before. **(B)** Scatter plot displaying the median protein intensities against the MaxQuant generated enrichment ratios for all the peptides purified by the Trial 3A site-ID purification. Peptides classified as enriched and highly enriched (according to the cut-off criteria described in the text) are highlighted in blue and red respectively. The numbers of protein and peptide identifications found are listed in the table on the right. **(C)** Venn diagram representing the overlap between the enriched proteins in the Trial 3A dataset and the total schizont palmitome.

Again, proteins previously shown to be palmitoylated, such as CRT, were enriched in this dataset. Strangely however, GAP45, a protein known to be palmitoylated and usually identified as enriched or highly enriched in the previous site-ID purifications, was identified in this dataset but did not pass the cut-off criteria to be defined as enriched.

Although the number of enriched proteins identified in this dataset had increased, the overlap with the total palmitome was not improved from the Trial 2 purification (the Trial 2 dataset purified 12% of the total palmitome while the Trial 3A dataset purified 13% of the total palmitome), and the palmitoylated protein GAP45, which had previously been consistently purified and classified as enriched, was not classified as enriched in this trial. This implied that Trial 3A was not an improvement from the previous purification, although the addition of NEM into the control samples did indeed reduce the direct identification of cysteine-containing peptides in the control samples by MaxQuant.

3.1.5. Development of the site-ID palmitome purification method – Trial 3B

As mentioned in Section 3.1.4 above, aliquots of the palmitome and control elutions from the Trial 3A site-ID purification were kept aside for label-free analysis in order to determine whether quantification using a label-free approach would be more appropriate for the site-ID palmitome purification compared to stable isotope dimethyl labelling.

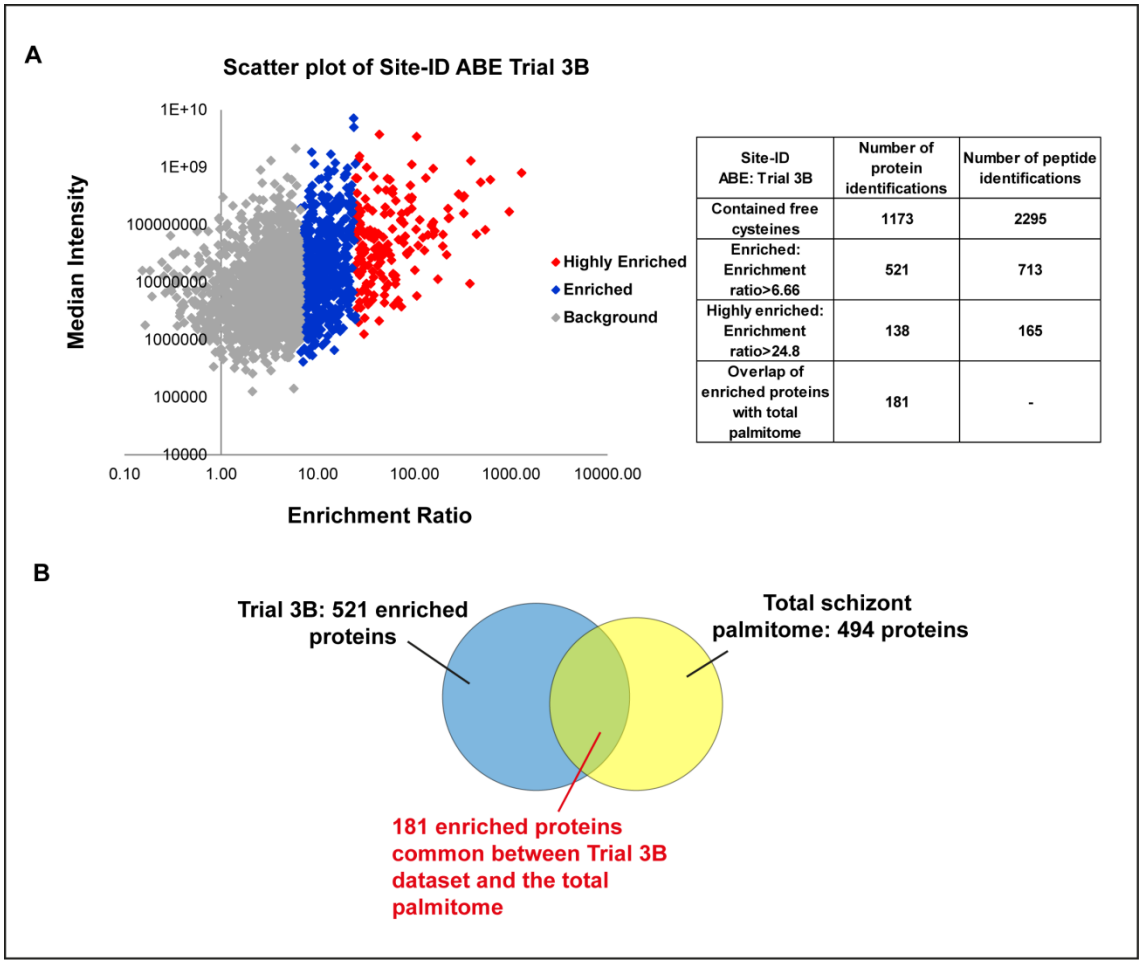
In the dimethyl labelling approach, the palmitome and control samples –each labelled with a different isopotomer of formaldehyde- were pooled together and analysed by LC-MS/MS in a single run. In contrast, for label-free analysis, the palmitome and control samples were kept separate and

were subjected to individual LC-MS/MS runs. Protein quantification was then measured by comparing the peak intensities from each individual run for a particular peptide [8]. Due to the run-to run experimental variations of LC-MS/MS, label-free quantification is thus less accurate and less reproducible, and is more variable compared to quantification with dimethyl labelling, resulting in the need for more rigorous analysis of the data acquired. However, despite the lower accuracy and lower reliability when measuring small quantitative differences, label-free analysis may provide greater depth of proteome coverage as well as a higher dynamic range of quantification [13]. Additionally, dimethyl labelling of samples requires that no buffers containing primary amines be present during the labelling reaction. Thus, samples need to be desalted prior to dimethyl labelling, and this can lead to some sample losses. No desalting is required for label-free analysis, thus reducing the loss of samples. Lastly, although dimethyl labelling is not expected to change the chemical properties of labelled peptides, it still involves modification of the peptides, and as the peptides being labelled here contain free cysteines, may result in unwanted modification at the side chains of the free cysteines by formaldehyde, which could lead to reduced identifications. With the label-free approach however, the eluted peptides are not further modified prior to mass spectrometry analysis. Thus, due to these differences, the label-free approach was tested to determine its suitability for the site-ID palmitome purifications.

The aliquots of the palmitome and control elutions from the Trial 3A site-ID purification (method described in Section 3.1.4) was analysed by LC-MS/MS without any dimethyl labelling. MaxQuant was again used to generate specific enrichment ratios (palmitome over control) for each of the peptide identifications.

In total, 2295 peptides, which had enrichment ratios and contained a free cysteine, were identified, and this corresponded to a total of 1173 proteins. This dataset of peptide identifications obtained using the label-free approach was clearly larger than previously obtained using dimethyl labelling, suggesting that a label-free analysis might provide better coverage for the site-ID palmitome purifications. The 2295 peptides was then grouped into enriched and highly enriched classes following the same cut-off criteria used for the previous site-ID trial purifications (detailed in Section 3.1.2). Using these cut-off criteria, 713 peptides (corresponding to 521 protein identifications) were defined as enriched, and out of these 713 peptides, 165 peptides (corresponding to 138 protein identifications) could be further defined as highly enriched. A scatter plot showing median peptide intensities against the MaxQuant generated enrichment ratios for all 2295 peptide identifications found in this Trial 3B site-ID palmitome purification is shown in Figure 3.4A.

Of the 521 proteins which had peptide identifications classified as enriched, 181 proteins were also present in the total schizont palmitome. This indicated that 37% of the total palmitome (181 out of 494 proteins in the total palmitome) was identified in this dataset (Figure 3.4B), suggesting that potentially more putative palmitoyl-proteins were being identified when using a label-free quantitative approach with the site-ID purification. As for the previous trial site-ID purifications, proteins shown previously to be palmitoylated, such as GAP45 and CRT, were identified as enriched in this dataset.



Site-ID ABE: Trial 3B	Number of protein identifications	Number of peptide identifications
Contained free cysteines	1173	2295
Enriched: Enrichment ratio>6.66	521	713
Highly enriched: Enrichment ratio>24.8	138	165
Overlap of enriched proteins with total palmitome	181	-

Figure 3.4: Trial 3B site-ID palmitome purification by ABE. The aliquots kept aside from the Trial 3A site-ID purification were analysed using a label-free approach by quantitative mass spectrometry, and this analysis was named Trial 3B. The scatter plot displays the median protein intensities against the MaxQuant generated enrichment ratios for all the peptides purified by the Trial 3B site-ID purification. Peptides classified as enriched and highly enriched (according to the cut-off criteria described in the text) are highlighted in blue and red respectively. The numbers of protein and peptide identifications found are listed in the table on the right. **(B)** Venn diagram representing the overlap between the enriched proteins in the Trial 3B dataset and the total schizont palmitome.

Although 37% of the total palmitome was found to be enriched in this dataset, this was equivalent to only 35% of the enriched proteins in this dataset (181 out of 521 proteins) (Figure 3.4B), which was less than previously observed in the other trial purifcations. This indicated that more than half of the proteins in this dataset were not found in the total palmitome. This could be due to the differences between the purification methods as detailed in the sections above (for example, the use of ABE purification alone in these site-ID purifications and differences in methodology such as the use of IAA instead of NEM), which would reduce the overlap between the two datasets. Additionally, another reason could be due to the fact that the analysis of the total palmitome was performed with SILAC labelling, whereas Trial 3B was performed using label-free analysis, which has been shown to provide greater proteome coverage compared to labelling approaches, including metabolic labelling [14, 15]. However, as described above, due to the nature of label-free analysis, quantification is generally more inaccurate, and although the dataset is larger, it must be more cautiously and rigorously analysed. For example, the accuracy of this label-free approach could be improved by performing further replicates and accepting only those proteins that were repeatedly identified.

In addition to the dataset of peptide identifications possessing enrichment ratios, a dataset of peptides with peak intensities found only in the palmitome (hydroxylamine-treated) sample and not in the control (non-hydroxylamine-treated) sample was also identified in this label-free trial analysis. It is possible that these peptides were truly only present in the palmitome sample, indicating that these peptides truly contained the palmitoylated cysteine. Indeed, peptides from known palmitoylated proteins, such as GAP45, were present in this 'palmitome-only' dataset. However, it is also equally possible that the corresponding peak for a particular peptide was not matched in control samples, due to run-to-run variation or differences in the retention time, which are the errors associated with label-free analysis. This was not observed in the previous site-ID trial purifications because in dimethyl labelling approaches, analysis is based on the co-elution of the differentially-labelled versions of a particular peptide. As the peptides in this palmitome-only dataset were not detected in the control samples, an enrichment ratio could not be assigned and cut-off criteria could not be applied to this dataset, making it even more difficult to distinguish between truly palmitoylated peptides and false-positive identifications. Thus, the peptide identifications in this 'palmitome-only' dataset were not further considered at this time.

Despite the lower accuracy of the label-free approach, this Trial 3B site-ID label-free palmitome purification appeared to be improved compared to the previous purifications, with increased proteome coverage and more enriched protein identifications found to be also present in the total

palmitome, suggesting that the use of label-free quantification may provide more useful information for these site-ID purifications.

3.1.6. Development of the site-ID palmitome purification method – Trial 4

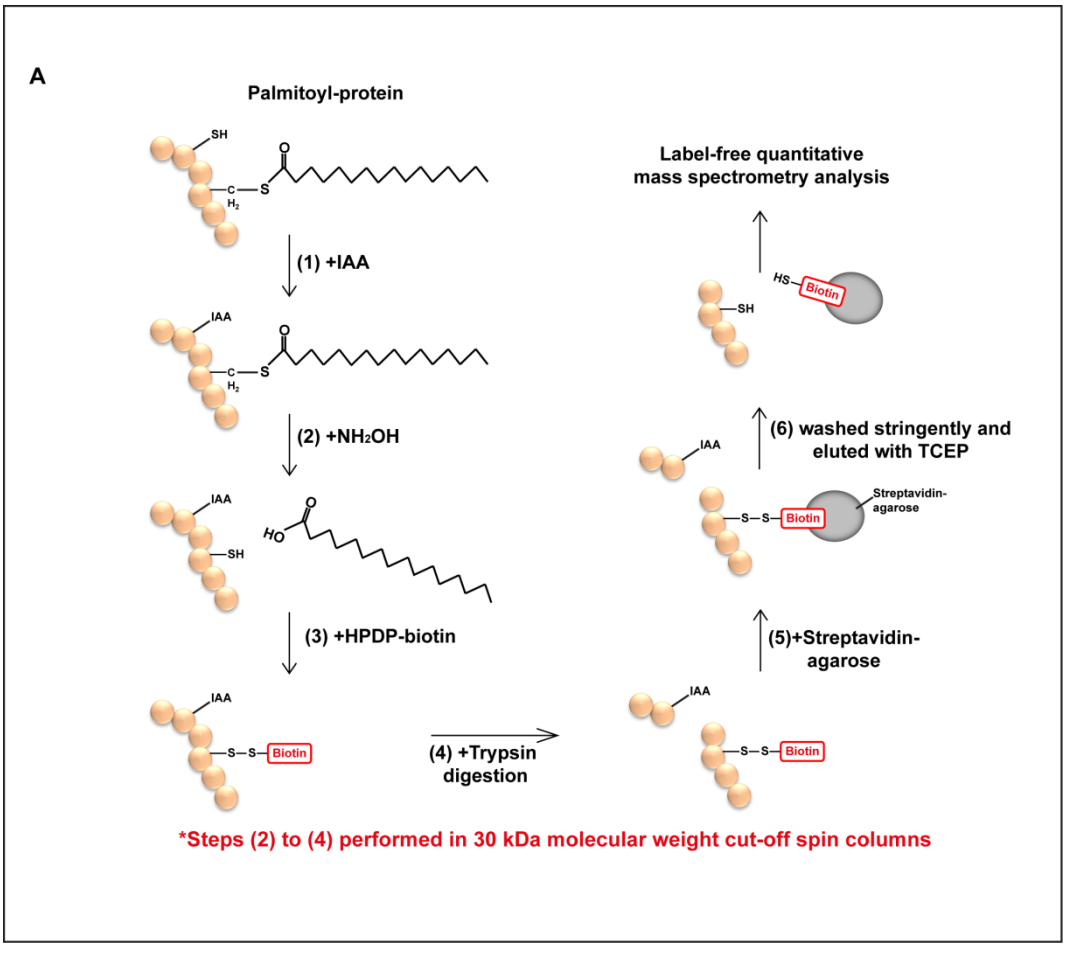
The previous trial purifications revealed that using a label-free approach appeared to be better suited for the site-ID palmitome purifications. Thus, another site-ID purification was next performed based on the Trial 3A purification method (shown in the schematic in Figure 3.3A) coupled with the Trial 3B label-free analysis (described in Section 3.1.5 above).

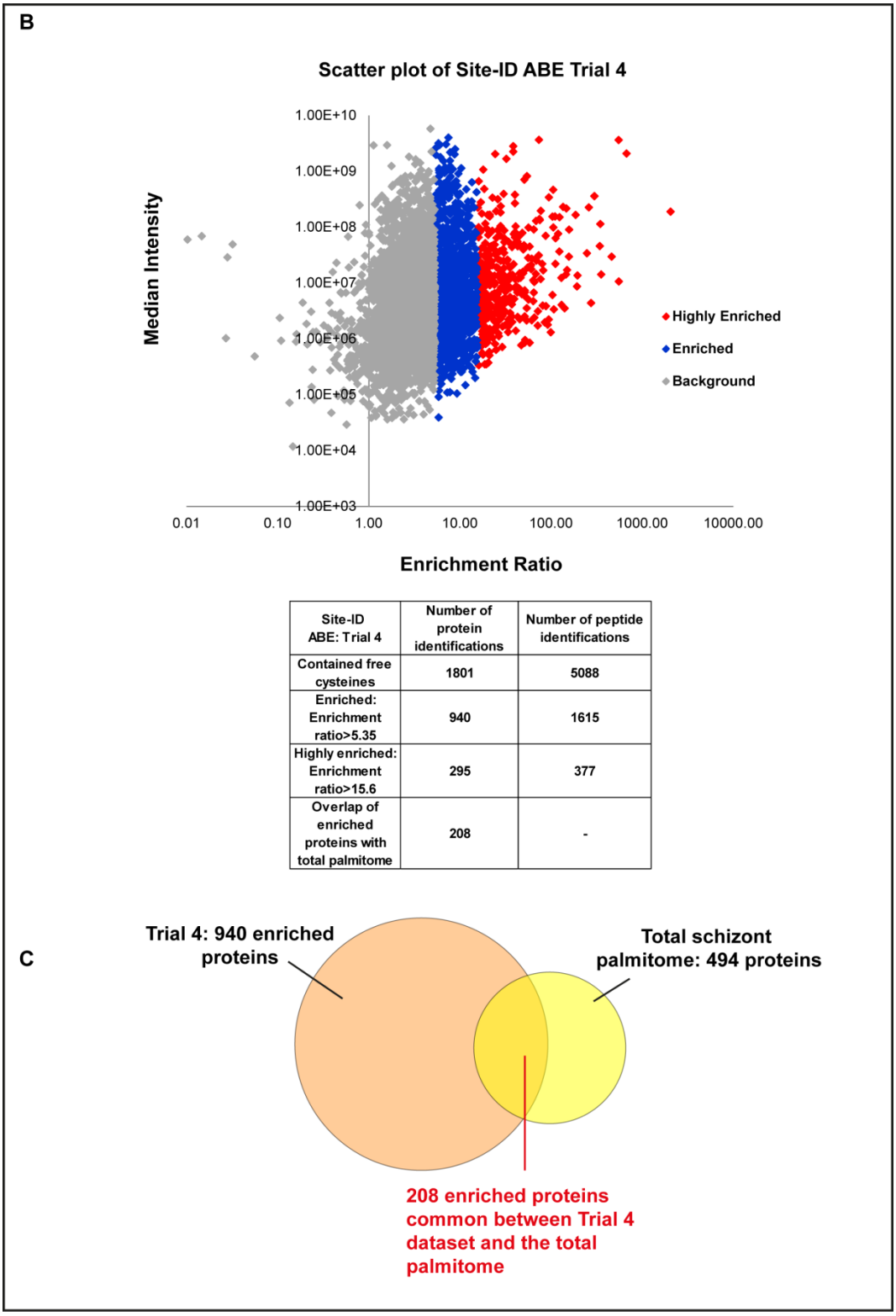
For this large-scale purification, two biological replicates and two technical replicates were performed. Each biological replicate consisted of approximately 10^9 *P. falciparum* strain 3D7 schizont stage parasites (200 mL of *in vitro* culture at 5% haematocrit and 8-10% parasitemia). Parasites were extracted from infected erythrocytes by saponin-lysis, and extraction of the proteome from saponin-treated parasite pellets was performed exactly the same as in Trial 3A. Initial blockage of free thiols was performed using IAA, and the hydroxylamine-cleavage and biotinylation steps were performed in 30 kDa MWCO spin columns as in Trial 3A. As always, the non-hydroxylamine-treated control sample was prepared in parallel using an equal amount of proteome. Although the addition of NEM into control samples during the hydroxylamine-cleavage and biotinylation steps of the Trial 3A purification appeared to reduce the high background of cysteine-containing peptides in the control sample, there was the possibility that the addition of NEM only to the control sample could end up artificially biasing the results, and might result in a dataset that did not truly represent the palmitoylated peptides in schizont stage parasites. Thus, for this Trial 4 purification, NEM was not added into control samples during the hydroxylamine-cleavage and biotinylation steps. Trypsin digestion in order to produce peptide fragments, streptavidin affinity purification of only biotinylated peptide fragments and elution of the peptides were all performed similar to Trial 3A (Section 3.1.4). The steps of this Trial 4 site-ID palmitome purification are shown in the schematic in Figure 3.5A.

The palmitome and control elutions were then analysed label-free by LC-MS/MS. MaxQuant was used to generate specific enrichment ratios (palmitome over control) from the median intensity of each peptide identification across the different replicates analysed.

In total, 5088 peptides containing enrichment ratios and a free cysteine were identified, and this corresponded to a total of 1801 proteins. This set of 5088 peptides was then grouped into enriched and highly enriched classes according to the same cut-off criteria used for the previous site-ID trial purifications (detailed in Section 3.1.2). Using these cut-off criteria, 1615 peptides (corresponding to

940 protein identifications) were defined as enriched, and out of these 1615 peptides, 377 peptides (corresponding to 295 protein identifications) could be further defined as highly enriched. A scatter plot showing the median peptide intensities against the MaxQuant generated enrichment ratios for all 5088 peptides identified in this Trial 4 site-ID palmitome purification is shown in Figure 3.5B. Of the 940 proteins which had peptide identifications classified as enriched, 208 proteins were also present in the total schizont palmitome, indicating that 42% of the total palmitome (208 out of 494 proteins in the total palmitome) was identified in this trial dataset (Figure 3.5C), which was higher than previously observed in the Trial 3B purification, suggesting that even more putative palmitoyl-proteins were being purified in this trial purification. As in the previous site-ID trial purifications, proteins known to be palmitoylated such as GAP45 and CRT, were identified and classified as enriched in this Trial 4 dataset. The presence of known palmitoyl-proteins, the larger enriched dataset obtained, and the increased number of proteins enriched in this trial that were also found in the total palmitome, indicated that this trial purification was the most improved compared to the previous trial purifications.





Site-ID ABE: Trial 4	Number of protein identifications	Number of peptide identifications
Contained free cysteines	1801	5088
Enriched: Enrichment ratio>5.35	940	1615
Highly enriched: Enrichment ratio>15.6	295	377
Overlap of enriched proteins with total palmitome	208	-

Figure 3.5: Trial 4 site-ID palmitome purification by ABE. (A) The Trial 4 site-ID purification was similar to that of Trial 3A and B. The proteome was extensively extracted from saponin-treated parasite pellets as in Trial 3A, and protein samples were (1) blocked with IAA, (2) treated with hydroxylamine (NH_2OH) (except for control samples) and (3) biotinylated using HPDP-biotin. In this Trial 4 purification however, NEM was not added into control samples during steps (2) and (3), as the addition of NEM into control samples only may skew the results obtained. (4) Trypsin digestion was then performed to digest the proteins into peptide fragments. Steps (2) to (4) were performed in 30 kDa MWCO spin columns as before. Biotinylated peptides were purified by (5)

streptavidin affinity purification, and unbound peptides removed by stringent washes. The biotinylated peptide fragments which remained bound to the beads were (6) eluted by reduction with TCEP and analysed label-free by quantitative mass spectrometry. **(B)** Scatter plot displaying the median protein intensities against the MaxQuant generated enrichment ratios for all the peptides purified by the Trial 4 site-ID purification. Peptides classified as enriched and highly enriched (according to the cut-off criteria described in the text) are highlighted in blue and red respectively. The numbers of protein and peptide identifications found are listed in the table below the plot. **(C)** Venn diagram representing the overlap between the enriched proteins in the Trial 4 dataset and the total schizont palmitome.

Although 42% of the total palmitome was enriched in the Trial 4 dataset, this was equivalent to only 22% of the Trial 4 enriched dataset (208 out of 940 proteins) (Figure 3.5C), indicating that approximately 80% of the enriched proteins in this purification were not found in the total palmitome. Additionally, this dataset was much larger than those of the previous purifications, and the classification of 940 proteins as enriched (and therefore potentially palmitoylated) may be inaccurate, considering the fact that just over 1000 proteins are upregulated in schizont stages [12].

These issues could be due in part to the nature of the label-free analysis, which is more prone to inaccurate quantification and run-to-run variability (as described in detail in Section 3.1.5). Another reason could be the high background presence of cysteine-containing peptides in the control samples, which was observed in the previous trial purifications and was also present in this Trial 4 dataset. This background, which could be due to incomplete blockage of free thiols during the initial treatment with IAA, or due to palmitate groups falling off during the other chemical steps of ABE, was not reduced in this Trial 4 purification, as NEM was not added to the control samples during the purification. The presence of this background in the control samples could thus affect the enrichment ratios and subsequently influence the cut-off criteria for enrichment, especially when the cut-off criteria were based on the median of the enrichment ratios. This indicates that more stringent analysis and more rigorous cut-off criteria may be needed in order to obtain useful information from this dataset.

Similar to what was observed in the Trial 3B site-ID purification (Section 3.1.5), a dataset of peptides that had peak intensities identified only in palmitome samples and not in control samples were also observed in this Trial 4 dataset. As for the Trial 3B purification, due to the lack of enrichment ratios for these peptide identifications, this dataset was not further analysed, although it must be noted that the presence of truly palmitoylated peptides within this dataset cannot be discounted.

In summary, five different trials of the site-ID palmitome purification were attempted in order to determine the optimum conditions required. All five trials purified peptides from a number of proteins which had been previously shown to be palmitoylated, indicating that the basic methodology was correct. All five trial purifications also produced enriched datasets which overlapped with the total schizont palmitome to different degrees. Analysis of the datasets from all trial purifications indicated that the use of a label-free approach improved the dataset of enriched proteins obtained. However, due to the inherent lower accuracy of label-free analysis, more stringent analysis of the datasets obtained is required. The best purification obtained out of the five trial purifications was considered to be the Trial 4 site-ID purification, as multiple biological and technical replicates were performed in this purification, resulting in an extensive dataset of peptide identifications and high overlap with the total schizont palmitome, although due to the large size of the dataset and the fact that the a label-free approach was used, more rigorous analysis of this dataset is still required. A summary of all five trial purifications and the main analyses performed on the five datasets is shown in Table 3.1.

Site-ID ABE purification	Method	Proteins Identified	Overlap with total palmitome
Trial 1	Full ABE protocol, pulled down with streptavidin-agarose, trypsin digestion on beads, eluted and dimethyl labelled	**158** peptides (**103** proteins) with ratios. **41** peptides (**32** proteins) defined as enriched, **9** of these peptides (**6** proteins) could be further defined as highly enriched.	**26** enriched protein IDs also found in total palmitome (494 proteins) - 81% of the enriched dataset was equivalent to **5%** of the total palmitome.
Trial 2	Blocked with IAA, ABE and trypsin digestion performed in spin columns, pulled down with streptavidin-agarose, eluted and dimethyl labelled.	**431** peptides (**291** proteins) with ratios. **139** peptides (**117** proteins) defined as enriched, **28** of these peptides (**24** proteins) could be further defined as highly enriched.	**58** enriched protein IDs also found in total palmitome (494 proteins) - 50% of the enriched dataset was equivalent to **12%** of the total palmitome.
Trial 3A	Same method as Trial 2, but initial block was with IAA, followed by NEM added only to control during ABE steps, dimethyl labelled	**700** peptides (**440** proteins) with ratios. **214** peptides (**179** proteins) defined as enriched, **62** peptides (**53** proteins) could be further defined as highly enriched.	**69** enriched protein IDs also found in total palmitome (494 proteins) - 39% of the enriched dataset was equivalent to **13%** of total palmitome.
Trial 3B	Same samples from 3A but not dimethyl labelled - label free analysis	**2295** peptides (**1173** proteins) with ratios. **713** peptides (**521** proteins) defined as enriched, **165** peptides (**138** proteins) could be further defined as highly enriched.	**181** enriched protein IDs also found in total palmitome (494 proteins) - 35% of the enriched dataset was equivalent to **37%** of total palmitome.
Trial 4	Same method as 3B: Initial block with IAA, but no NEM further added into control during ABE steps, not dimethyl labelled - label free analysis. 2 biological replicates and 2 technical replicates	**5088** peptides (**1801** proteins) with ratios. **1615** peptides (**940** proteins) defined as enriched, **377** peptides (**295** proteins) could be further defined as highly enriched.	**208** enriched protein IDs also found in total palmitome (494 proteins) - 22% of the enriched dataset was equivalent to **42%** of total palmitome.

Table 3.1: Summary of all five trial site-ID purifications and the analyses performed on the five datasets.

3.2. Analysis of the overlaps between the five trial site-ID palmitome purifications

As described in the sections above, label-free analysis of the site-ID palmitome purifications appeared to provide the most peptide identifications classified as enriched and the largest overlap with the total schizont palmitome. However, due to the relatively high background of cysteine-containing peptides found in the control samples, as well as the reduced accuracy of label-free quantification, more stringent analysis of the label-free datasets was required. One way to tackle this would be to investigate the overlaps between all of the trial purifications performed during the development of the site-ID purification method, in order to determine which peptides were consistently purified in all the trial purifications. If a particular peptide was consistently identified in the different trial purifications and was consistently classified as enriched, that peptide could be more confidently considered a putative palmitoylated peptide.

3.2.1. Overlap between all five trial site-ID palmitome purifications (Trial 1, 2, 3A, 3B and 4)

The first overlap analysed was the overlap between proteins identified in all five trial site-ID purification datasets. In this analysis, only proteins classified as enriched in all five datasets were considered. It is possible that proteins classified as enriched in one trial purification could be classified as not enriched in another purification, due to the differences in methodology and to whether the analysis was performed label-free or with dimethyl labelling (as described in Section 3.1 above). However, to reduce the complexity of this analysis, only the enriched protein datasets were considered at this time.

Only 6 enriched proteins were found to be common between all five trial site-ID purifications, and all of these proteins were also found in the total schizont palmitome. The small size of this overlap is not unexpected given that the Trial 1 dataset is an outlier with respect to size, with only 32 proteins classified as enriched (Section 3.1.2). Repeating the overlap analysis with proteins classified as enriched in all the later datasets (Trial 2, 3A, 3B and 4) revealed 33 enriched protein identifications in common. Of these 33 common enriched proteins, 22 proteins were also found to be present in the total schizont palmitome (Figure 3.6A).

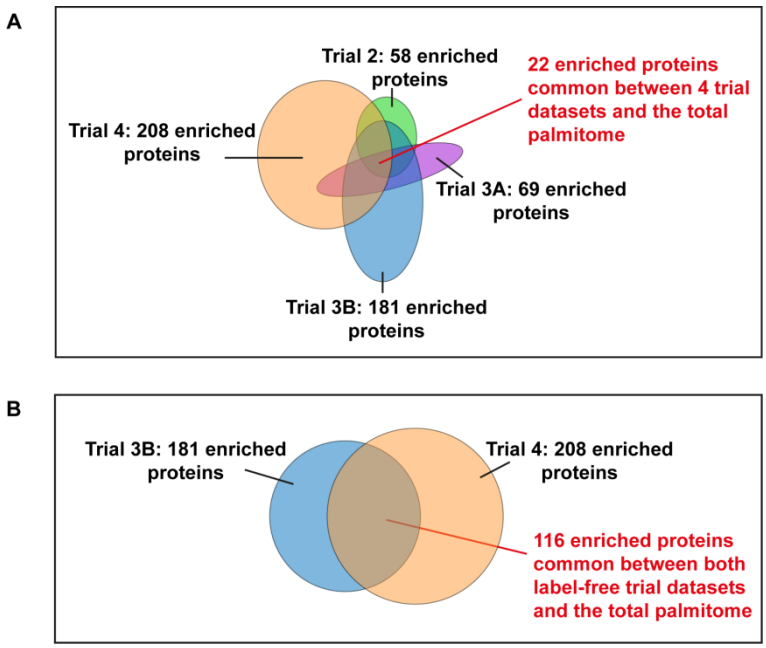


Figure 3.6: Overlaps between the different trial site-ID palmitome purifications and the total palmitome. Venn diagrams representing the overlaps between proteins classified as enriched in each of the different trial site-ID palmitome purifications. All the enriched proteins listed here for each trial dataset were also found to be present in the total schizont palmitome. The numbers of enriched proteins common between the trial site-ID datasets and the total schizont palmitome are indicated in red. (A) Overlap between 4 trial site-ID palmitome purifications (excluding Trial 1) and the total palmitome. (B) Overlap between the two label-free trial site-ID palmitome purifications and the total palmitome.

To assess whether the same peptides were consistently enriched for these 22 proteins, the peptides for all 22 enriched proteins common between the 4 site-ID trial purifications, as well as the total palmitome, were compared. For these proteins, 94 unique peptides passed the cut-off criteria for enrichment in at least one of the 4 trial datasets (for peptides which were identified multiple times, only one peptide was considered, and the peptide with the higher enrichment ratio was chosen), but only 14 unique enriched peptides were found to be present in all 4 trial datasets. As these peptides appeared to be consistently purified by all 4 trial site-ID purifications, and passed the cut-off criteria for enrichment used for each dataset, these peptides should be considered as high confidence palmitoylated peptides, and the cysteine residue present in each of these 14 common enriched peptides considered as high confidence sites of palmitoylation. For those peptides with more than one cysteine residue present, the cysteine which is palmitoylated is unknown and will have to be determined through experimental means. Table 3.2 shows all 14 enriched peptides common between the 4 datasets, along with their enrichment ratios in each dataset. These 14 peptides also include 2 peptides which were also classified as enriched in the Trial 1 dataset, and the enrichment

ratios of these 2 peptides in the Trial 1 purification is also shown in Table 3.2. It must be noted however that as mentioned above, ABE can result in the false enrichment of highly abundant proteins. Thus, the presence of peptides from highly abundant proteins, such as heat shock 70 kDa protein (PF08_0054) and glyceraldehyde-3-phosphate dehydrogenase (PF14_0598), within the dataset of 14 enriched common peptides may also be due to the high abundance of these proteins.

Although 14 common enriched peptides were found in this analysis, this degree of overlap might appear rather small. However, it should be noted that the different trial purifications had very different methodologies, as they were part of a methods development process. These relatively large differences in sample preparation, extraction and labelling will have a major impact on the subset of the proteome that is captured, making the identification of only 14 high confidence palmitoylation sites not completely surprising.

Protein	Protein name	Sequence	Enrichment Ratio				
			Trial 1	Trial 2	Trial 3A	Trial 3B	Trial 4
PF08_0035	conserved Plasmodium protein, unknown function	NYNTMVNDHNIN**C**TK	74.86	53.12	11.07	52.68	9.09
		NVN**C**NYELINK	51.87	155.19	33.88	32.27	6.06
PF08_0054	heat shock 70 kDa protein	LVNF**C**VEDFKR	-	4.86	3.89	13.29	15.08
PF11_0351	heat shock protein hsp70 homologue	AKLEEL**C**HDLLK	-	6.33	12.00	614.39	10.41
PF13_0338	cysteine-rich surface protein	TIGFV**C**PVK	-	10.50	2.93	8.47	7.97
PF14_0578	conserved Plasmodium protein, unknown function	NS**C**ETLLESEQNLLK	-	12.97	3.20	59.40	7.71
PF14_0598	glyceraldehyde-3-phosphate dehydrogenase	QLIVSNAS**C**TTN**C**LAPLAK	-	5.86	3.21	12.84	44.76
PFD1165w	Serine/Threonine protein kinase, FIKK family	LIT**C**SYNSDK	-	10.70	31.82	94.82	158.42
PFF0435w	ornithine aminotransferase	EQLDE**C**TEIIVK	-	4.95	7.51	10.32	26.17
PFF0690c	organic anion transporter	YV**C**FSGLDK	-	242.35	10.47	26.70	38.75
		IDENIIDG**C**GK	-	53.01	11.45	27.12	9.68
PFI0875w	Heat shock protein 70 (HSP70) homologue	LKDLEAV**C**QPIIVK	-	6.28	9.47	23.68	38.68
PFI1445w	High molecular weight rhoptry protein-2	N**C**TLLTDFMK	-	6.52	5.20	53.43	6.45
PFL1300c	conserved Plasmodium protein, unknown function	NIETQDLNLN**C**P**C**WNR	-	5.63	4.68	198.62	43.24

Table 3.2: The 14 enriched peptides identified in all 4 trial site-ID palmitome purifications. These peptides were found to be present in either all 5 trial purifications, or in 4 of the trial purifications (excluding Trial 1). The enrichment ratios in each trial dataset for each of the peptides is shown here, with enrichment ratios above the cut-off criteria for highly enriched peptides shown in red, and enrichment ratios above the cut-off criteria for enriched peptides shown in blue. The putative palmitoylated cysteines are highlighted in bold and in green.

3.2.2. Overlap between the two label-free trial site-ID palmitome purifications (Trial 3B and 4)

A large part of the divergence between the trial datasets may be because two of the datasets were dimethyl labelled, while the remaining two were analysed using label-free approaches, which produces far larger datasets. Therefore, in the next analysis, only the overlap between the two label-free trial datasets (Trial 3B and 4) were analysed. As before, only proteins classified as enriched in each of the datasets were considered.

In this analysis, 324 enriched protein identifications were found to be common between the 2 label-free trial purifications considered. Of these 324 enriched proteins, 116 proteins were also found to be present in the total palmitome (Figure 3.6B). Gene ontology (GO) term analysis of the 116 common enriched proteins revealed an enrichment for proteins involved in localisation and transport (Figure 3.7A), which was previously observed for the total palmitome and for palmitomes in other systems [4, 16]. Thus, the degree of overlap between the 2 label-free trial datasets appeared to be better than the previous overlaps analysed, and may represent a more extensive dataset of putative palmitoylated proteins, as evidenced by GO term analysis of this overlap, which was similar to GO term analyses of other palmitomes.

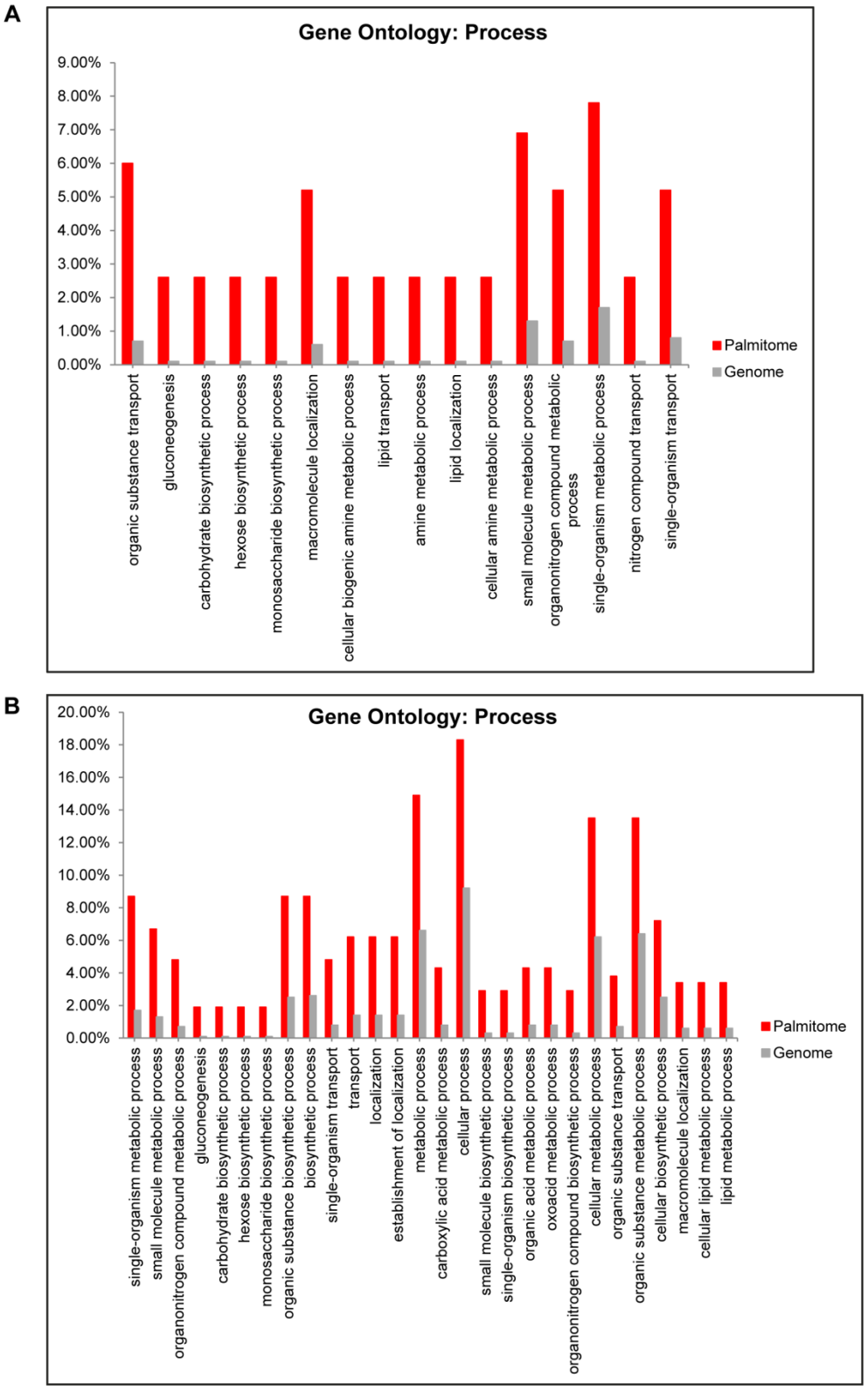


Figure 3.7: Gene ontology (GO) term analysis of enriched proteins in the overlapping trial site-ID purification datasets. GO analysis is of biological process annotations and is presented in comparison to corresponding percent genome values. p values are ≤0.05 for all displayed terms. **(A)** Enriched proteins present in both label-free datasets (Trial 3B and 4) and the total palmitome. **(B)** Enriched proteins present in both the Trial 4 dataset and the total palmitome. GO term analysis was performed using the following online tool: Generic Gene Ontology (GO)Term Finder (http://go.princeton.edu/cgi-bin/GOTermFinder).

The 116 enriched proteins which were common between the 2 label-free site-ID purifications, as well as the total palmitome, corresponded to a total of 323 unique peptides, which were classified as enriched in at least one of the 2 label-free trial datasets. Of these 323 unique peptides, a total of 142 unique enriched peptides were found to be present in both trial datasets. Although these peptides were present in both trial datasets considered here, and managed to pass the cut-off criteria for enrichment in both datasets, it must be noted that both the datasets analysed here were quantified through label-free methods, which is less accurate and more prone to error (as described in detail in Section 3.1 above), and were also affected by issues with the relatively high background of cysteine-containing peptides in control samples (Section 3.1.6). Thus, the peptides found to be common between both these datasets can be considered medium confidence palmitoylated peptides, and the cysteine residues present within these peptides regarded as medium confidence putative palmitoylation sites. Table 3.3 displays a selected set of enriched peptides common between the 2 trial datasets, with their corresponding protein identifications, which were also present in the total palmitome, along with the enrichment ratios of each peptide in each trial dataset.

Protein	Protein name	Peptide Sequence	Enrichment Ratios	
			Trial 3B	Trial 4
MAL13P1.117	DHHC-type zinc finger protein, putative	VYESNNIFF**C**K	329.46	186.76
MAL13P1.214	phosphoethanolamine N-methyltransferase	VLDIGSGLGGG**C**MYINEK	16.24	22.80
MAL7P1.27	chloroquine resistance transporter	IGGGS**C**IGK	24.62	7.13
MAL8P1.32	nucleoside transporter, putative	VGLLTGYG**C**STLYK	42.95	14.88
PF07_0033	Cg4 protein	LTAYYENTPDLPSN**C**IK	9.64	25.33
		EIVLSYPPTFTN**C**QK	6.93	7.97
PF08_0035	conserved Plasmodium protein, unknown function	NYNTMVNDHNIN**C**TK	52.68	9.09
		NVN**C**NYELINK	32.27	6.06
PF10_0039	membrane skeletal protein IMC1-related	ISSEEIKPAG**C**C**S**AA**C**T	441.70	248.58
PF10_0220	phospholipid scramblase 1, putative	I**C**IPMNMK	157.02	54.60
		CC**S**YNFNLFDPSNNK	39.14	72.16
		SILSPMQT**C**K	14.53	9.75
PF11_0168	moving junction protein	EI**C**EPQNGLIDETLTK	31.13	9.65
PF13_0338	cysteine-rich surface protein	TIGFV**C**PVK	8.47	7.97
PF14_0102	rhoptry-associated protein 1, RAP1	FGT**C**IGSFGEHHLR	122.68	27.78
PFC0120w	Cytoadherence linked asexual protein 3.1	VLDMM**C**DHESVYYSEK	326.77	30.83
		LFESLIQ**C**IEK	12.01	8.65
PFE0785c	metabolite/drug transporter, putative	FIG**C**GEEYMPLR	14.82	75.68
PFD1165w	Serine/Threonine protein kinase, FIKK family	LIT**C**SYNSDKDAK	94.82	158.42
		EIC**D**NEN**C**VNLEDIK	8.94	7.41
		L**C**DFANSAPIYTYNNR	7.59	12.41
PFF0675c	myosin E	AYEN**C**FFENR	50.93	7.22
		IGNMNEIS**C**YK	91.40	6.02
PFI1730w	cytoadherence linked asexual protein 9(CLAG9)	KE**C**NIYESDR	6.74	81.45
PFL1090w	glideosome-associated protein 45	SVTP**C**DMNK	43.53	18.14

Table 3.3: A selected set of enriched proteins found to be common between both label-free trial site-ID datasets (Trial 3B and Trial 4), as well as the total palmitome, along with a selected set of enriched peptides common between both the trial datasets. The enrichment ratios in each trial dataset for each of the peptides is shown here, with enrichment ratios above the cut-off criteria for highly enriched proteins shown in red and enrichment ratios above the cut-off criteria for enriched proteins shown in blue. The putative palmitoylated cysteines are shown in bold and in green.

3.2.3. Overlap between Trial 4 site-ID palmitome purification and the total palmitome

As described in Section 3.1.6, the Trial 4 site-ID purification was considered the best purification, with multiple biological and technical replicates performed, a large dataset of enriched peptides obtained and the greatest coverage of the total palmitome. However, it was also noted that this dataset had to be treated with caution due to the label-free quantification used, as well as the high background of cysteine-containing peptides in the control samples, thus indicating that the results of this dataset cannot be accepted by itself. Overlap of the Trial 4 dataset with the Trial 3B dataset (described in Section 3.2.2 above) yielded some useful information, with a total of 142 medium confidence palmitoylation sites determined. Nevertheless, as the Trial 4 dataset was larger than the Trial 3B dataset, and overlapped more with the total palmitome, some palmitoylated peptides may have been missed simply due to the lower proteome coverage of the Trial 3B purification. Furthermore, in the Trial 3B purification, NEM was added into control samples in an attempt to reduce the background in the control, which as explained in Section 3.1.6, may instead inadvertently bias the results. Thus, an analysis based on just the Trial 4 dataset may provide further useful information on the sites of palmitoylation.

As an analysis of the peptides purified by the Trial 4 purification based solely on the cut-off criteria for enrichment could possibly have many false positive identifications due to the effect of the high background in control samples, only those peptides corresponding to proteins which were also found in the total palmitome were considered here.

As described in Section 3.1.6, 208 proteins classified as enriched in the Trial 4 dataset were found to be also present in the total schizont palmitome (Figure 3.5C). GO term analysis of the 208 common enriched proteins revealed an enrichment of proteins involved in localisation, transport and metabolic processes (Figure 3.7B). The GO term analysis of all proteins classified as enriched in just the Trial 4 purification resulted in the enrichment of proteins involved mainly in metabolic processes (data not shown). Although enrichment for proteins involved in metabolic processes was still observed in the GO term analysis of the 208 proteins common between the Trial 4 dataset and the total palmitome, the enrichment of proteins involved in localisation and transport seen here, which have been previously observed in other palmitomes, implied that this set of 208 common proteins may be a more accurate representation of palmitoylated proteins.

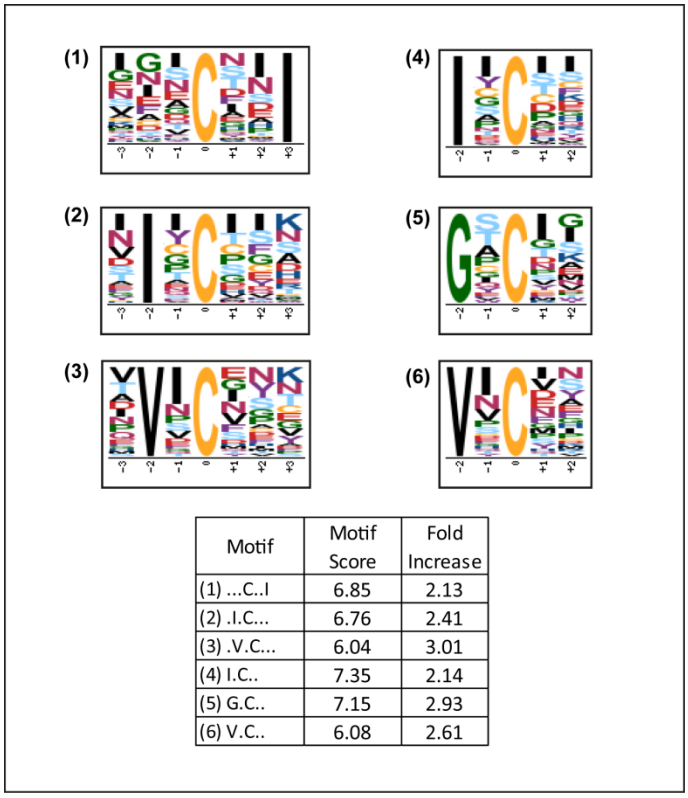
The 208 enriched proteins present in both the Trial 4 purification and the total palmitome corresponded to a total of 383 unique enriched peptides (which included the 142 unique enriched peptides already determined from the analysis of the overlap between the 2 label-free datasets and

the total palmitome described in Section 3.2.2 above). Although the dataset of 208 enriched proteins did appear to be a more representative list of palmitoyl-proteins, the peptide data was still only from one round of purification (albeit with two biological and two technical replicates). Thus, the 383 unique enriched peptides described here can also be regarded as medium confidence palmitoylated peptides, with the cysteine residues present in each peptide regarded as medium confidence palmitoylated sites. Table 3.4 displays a selected set of enriched proteins common between the Trial 4 dataset and the total palmitome, along with a selected set of the enriched peptides, and the enrichment ratios of each peptide in the Trial 4 dataset.

Protein	Protein Names	Peptide Sequence	Enrichment Ratio
MAL13P1.117	DHHC-type zinc finger protein, putative	VYESNNIFF**C**K	186.76
MAL13P1.310	calpain	KKEE**C**NIIENVEGNNVGNK	12.69
MAL7P1.229	Cytoadherence linked asexual protein	TVTT**C**YWFPSPIK	6.61
MAL7P1.27	chloroquine resistance transporter	IGGGS**C**IGK	7.13
PF07_0033	Cg4 protein	ITAYYENTPDIPSN**C**IK	25.33
		CEII**C**DIADSNIGGR	17.53
PF10_0039	membrane skeletal protein IMC1-related	ISSEEIKPAG**C**C**C**SAA**C**T	248.58
PF10_0220	phospholipid scramblase 1, putative	**C**C**C**SYNFNIFDPSNNK	72.16
		I**C**IPMNMK	54.60
		SEF**C**YR	44.16
		SIISPMQT**C**K	9.75
PF11_0168	moving junction protein	EI**C**EPQNGIIDETITK	9.65
PF13_0338	cysteine-rich surface protein	**C**MISFDEGNNWK	27.03
		SIIII**C**SNIHDWK	9.86
		TIGFV**C**PVK	7.97
PF14_0102	rhoptry-associated protein 1, RAP1	FGT**C**IGSFGEHHIR	27.78
PF14_0164	NADP-specific glutamate dehydrogenase	F**C**QAFMNEIYR	106.37
		QTAVVSGSGNVAIY**C**VQK	35.78
		HIGP**C**TDVPAGDIGVGGR	10.96
		SNNIIY**C**PSK	7.68
		VAESYIEQG**C**F	6.67
PF14_0455	multidrug resistance protein 2 (heavy metal transport family)	**C**NGEYTEMWNMQSK	14.84
PFB0815w	Calcium-dependent protein kinase 1	QIISGI**C**YIHK	6.27
PFA0130c	Serine/Threonine protein kinase, FIKK family, putative	IS**C**DPI**C**IEK	349.65
PFC0120w	Cytoadherence linked asexual protein 3.1	VIDMM**C**DHESVYYSEK	30.83
		IFESIIQ**C**IEK	8.65
PFC0185w	membrane skeletal protein IMC1-related	N**C**FIFNINR	19.88
PFD0895c	Bet3 transport protein, putative	SDISF**C**EDFEETVNVIAK	29.27
PFE0785c	metabolite/drug transporter, putative	FIG**C**GEEYMPIR	75.68
PFI1730w	cytoadherence linked asexual protein 9(CLAG9)	E**C**NIYESDR	81.45
PFL1090w	glideosome-associated protein 45	SVTP**C**DMNK	18.14

Table 3.4: A selected set of enriched proteins found to be common between the Trial 4 site-ID dataset and the total palmitome, along with a selected set of corresponding enriched peptides. The enrichment ratios from the Trial 4 purification is shown here for each of the peptides, with enrichment ratios above the cut-off criteria for highly enriched proteins shown in red and enrichment ratios above the cut-off criteria for enriched proteins shown in blue. The putative palmitoylated cysteines are highlighted in bold and in green.

The sequences of all these 383 unique enriched peptides were also analysed by motif-x, an online software tool which looks for over-represented sequence motifs within sequence data (http://motif-x.med.harvard.edu) [17], in order to determine whether any conserved motifs existed around the palmitoylated cysteines. No significant consensus motif for palmitoylation was determined from this analysis. However, there did appear to be a significant prevalence for the presence of hydrophobic, branched amino acids, such as isoleucine and valine, one or two amino acids upstream or downsteam of the putative palmitoylated cysteine. As one of the classes of proteins commonly palmitoylated are proteins that are palmitoylated close to TM domains, it is not surprising that the palmitoylated cysteine appears to be often close to hydrophobic amino acids. Indeed, analysis of the 208 proteins corresponding to the 383 unique enriched peptides revealed that 27% of the enriched proteins contained TM-domains (58 out of 208 proteins). The sequence motifs found to be overrepresented by the motif-x software is as shown in Figure 3.8.



Motif	Motif Score	Fold Increase
(1) ...C..I	6.85	2.13
(2) .I.C...	6.76	2.41
(3) .V.C...	6.04	3.01
(4) I.C..	7.35	2.14
(5) G.C..	7.15	2.93
(6) V.C..	6.08	2.61

Figure 3.8: Over-represented sequence motifs within the Trial 4 site-ID palmitome purification dataset. The sequences of the 383 unique enriched peptides of the Trial 4 dataset, which corresponded to protein identifications also present within the total schizont palmitome, were analysed by motif-x software (http://motif-x.med.harvard.edu) and 6 over-represented sequence motifs were found. The motif score is calculated by the software and is the sum of the -log(probability) of each fixed position in the motif. The motif score roughly correlates with statistical significance, where motifs with higher motif scores are more statistically significant (although all motifs found by the motif-x software is statistically significant) [17]. The fold increase measures the enrichment of the particular motif in the sequence dataset of interest compared to the background [17] (which in this case was sequences of the entire *P. falciparum* proteome obtained from *Plasmo*DB). The motif-x software was run using the following standard parameters: significance threshold - 0.000001, occurrences threshold - 20, width - 7 (for motifs 1 to 3) or 5 (for motifs 4 to 6).

In summary, the site-ID palmitome purification method developed here has successfully allowed the purification of palmitoylated proteins and can localise the putative palmitoylated cysteine. However, due to the reasons discussed in detail in the sections above, the whole enriched datasets from each of the different trial purifications might not be accepted as completely accurate when considered individually. Additionally, the excessive background of cysteine-containing peptides in control samples may disappointingly have a negative effect on the definition of enrichment in some of the datasets. Nevertheless, analysis of the overlaps between these different datasets still produced valuable information, providing sets of high confidence and medium confidence palmitoylation sites.

Although it cannot be stated that these sites are definitely palmitoylated, and the definition of the palmitoylation sites identified as high confidence or medium confidence is mostly arbitrary, the data presented here can still act as a guide to which cysteines may be possibly palmitoylated, and these can then be further validated experimentally. Indeed, the site-ID data from the Trial 4 purification was used to guide experiments in other parts of this project (described in Chapter 6). A summary of the analysis of the overlaps between the different trial purifications is shown in Table 3.5, and is represented visually by the Venn diagrams in Figure 3.6.

Site-ID Datasets	Overlaps between datasets	Overlap with total palmitome	Unique enriched peptides
All 5 datasets (Trial 1, 2, 3A, 3B and 4)	6 enriched protein IDs common	6 enriched protein IDs also found in total palmitome (494 proteins) - 100% of overlapping protein IDs equivalent to 1.2% of total palmitome	Total of 34 unique enriched peptides - only 2 peptides found in all 5 datasets with corresponding protein IDs found in the total palmitome.
4 datasets (Trial 2, 3A, 3B and 4)	33 enriched protein IDs common	22 enriched protein IDs also found in total palmitome (494 proteins) - 67% of overlapping protein IDs equivalent to 4.5% of total palmitome.	Total of 94 unique enriched peptides - only 14 peptides found in all 4 datasets with corresponding protein IDs also found in the total palmitome.
2 datasets (Trial 3B and 4) - label free datasets	324 enriched protein IDs common	116 enriched protein IDs also found in total palmitome (494 proteins) - 35% of overlapping protein IDs equivalent to 24% of total palmitome	Total of 323 unique enriched peptides - 142 peptides found in both datasets with corresponding protein IDs also found in the total palmitome.
1 label-free dataset (Trial 4)	-	208 enriched protein IDs also found in total palmitome (494 proteins).	Total of 383 unique enriched peptides with corresponding protein IDs also found in the total palmitome.

Table 3.5: Summary of the analyses performed on the overlaps between the different trial site-ID palmitome purifications.

3.3. Trial site-ID palmitome purification from *Plasmodium berghei* schizonts and *Toxoplasma gondii* tachyzoites

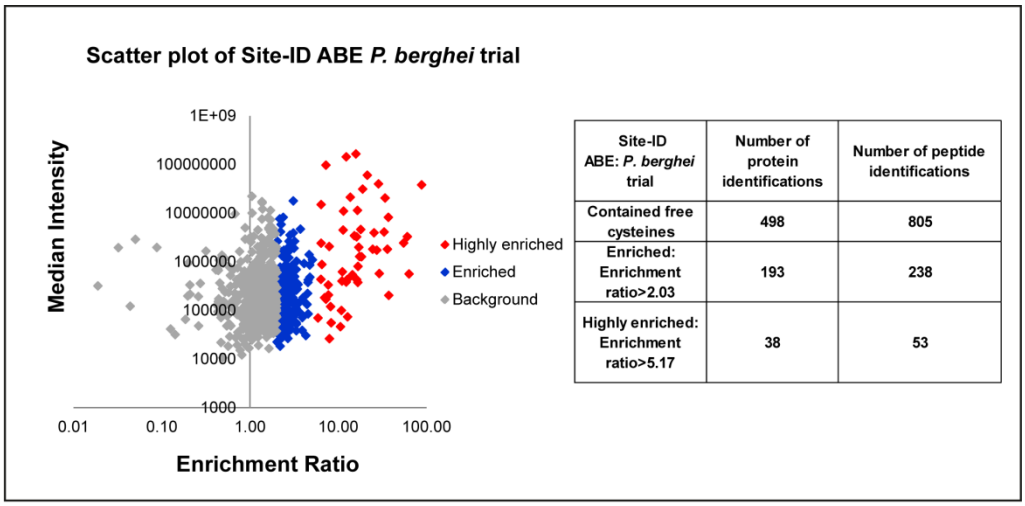
As mentioned in Section 3.1 above, the site-ID palmitome purification method was developed based on ABE, and without SILAC metabolic labelling, in order to allow the use of this site-localising method on other stages of the parasite life cycle, or on other Apicomplexan species, which are not amenable to metabolic labelling. As detailed in the sections above, the site-ID purification method did successfully purify palmitoyl-proteins and identified some putative palmitoylated cysteines in *P. falciparum* schizonts. In order to determine whether this site-ID purification could be used to determine sites of palmitoylation in other related Apicomplexan parasites, the final site-ID purification method (as performed in the *P. falciparum* Trial 4 site-ID purification) was applied to both the schizont stages of the rodent malaria parasite, *Plasmodium berghei*, and on the tachyzoite stages of the related Apicomplexan species, *Toxoplasma gondii.*

3.3.1. Trial site-ID palmitome purification from P. berghei schizonts

Approximately 10^9 *P. berghei* strain ANKA 2.33 schizont stage parasites was extracted from infected mouse blood by saponin-lysis as described in the Materials and Methods (infection of mice with *P. berghei* strain ANKA 2.33 and isolation of blood from infected mice was kindly performed by Katarzyna Modrynska, Billker and Rayner lab). Extraction of the proteome from saponin-treated parasite pellets and all other steps of the site-ID purification were performed exactly as for the *P. falciparum* Trial 4 site-ID palmitome purification (described in Section 3.1.6 and shown in the schematic in Figure 3.5A). Palmitome and control elutions were analysed using the label-free approach by LC-MS/MS as before. MaxQuant was used to generate specific enrichment ratios (palmitome over control) for each of the peptide identifications.

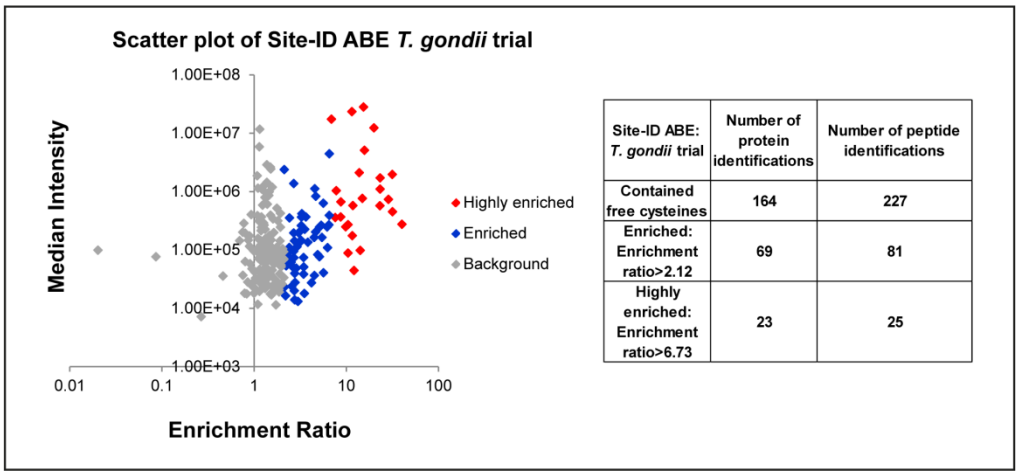
In total, 805 peptides, which had enrichment ratios and contained a free cysteine, were identified in this *P. berghei* trial purification, and this corresponded to a total of 498 proteins. This set of 805 peptides was then grouped into enriched and highly enriched classes following the same cut-off criteria used for the *P. falciparum* trial site-ID datasets (detailed in Section 3.1.2). Using these cut-off criteria, 238 peptides (corresponding to 193 protein identifications) were defined as enriched, and out of these 238 peptides, 53 peptides (corresponding to 38 protein identifications) could be further defined as highly enriched. A scatter plot showing the median peptide intensities against the MaxQuant generated enrichment ratios for all the 805 peptides identified in this trial purification is shown in Figure 3.9A.

A



Site-ID ABE: *P. berghei* trial	Number of protein identifications	Number of peptide identifications
Contained free cysteines	498	805
Enriched: Enrichment ratio>2.03	193	238
Highly enriched: Enrichment ratio>5.17	38	53

B



Site-ID ABE: *T. gondii* trial	Number of protein identifications	Number of peptide identifications
Contained free cysteines	164	227
Enriched: Enrichment ratio>2.12	69	81
Highly enriched: Enrichment ratio>6.73	23	25

Figure 3.9: Trial site-ID palmitome purification by ABE for *Plasmodium berghei* and *Toxoplasma gondii*. The site-ID trial purification for *P. berghei* and *T. gondii* was performed identical to that of the *P. falciparum* Trial 4 site-ID palmitome purification, and purified peptides were analysed label-free by quantitative mass spectrometry. **(A)** Scatter plot displaying the median protein intensities against the MaxQuant generated enrichment ratios for all the peptides purified by the *P. berghei* site-ID purification. Peptides classified as enriched and highly enriched (according to the cut-off criteria described in the text) are highlighted in blue and red respectively. The numbers of protein and peptide identifications found are listed in the table on the right. **(B)** Scatter plot displaying median protein intensities against MaxQuant generated enrichment ratios for all the peptides purified by the *T. gondii* site-ID purification. Peptides classified as enriched and highly enriched (according to the cut-off criteria described in the text) are highlighted in blue and red respectively. The numbers of protein and peptide identifications found are listed in the table on the right.

The dataset from the *P. berghei* trial site-ID purification was smaller than previously observed for *P. falciparum.* In fact, the median protein intensities appeared to be in general lower than that of the previous *P. falciparum* purifications (Figure 3.9A compared to Figure 3.5B), suggesting that the overall amount of protein present was lower. This indicates that further optimisation of the amount of initial parasite material required for the purification must still be performed.

Despite the smaller size of the *P. berghei* trial dataset, known palmitoylated proteins such as GAP45 and MTIP, were found to be present and was classified as enriched in this dataset. The presence of known palmitoyl-proteins was encouraging, as this suggested that putative palmitoyl-proteins were indeed being purified in this trial purification. A summary of the analysis performed on the *P. berghei* trial dataset is shown in Table 3.6.

Site-ID ABE purification	Method	Proteins Identified
P. berghei trial	Similar to *P. falciparum* Trial 4 - label free	**804** peptides (**498** proteins) with ratios. **237** peptides (**193** proteins) defined as enriched, **52** of these peptides (**38** proteins) further defined as highly enriched.
T. gondii trial	Similar to *P. falciparum* Trial 4 - label free	**227** peptides (**164** proteins) with ratios. 81 peptides (**69** proteins) defined as enriched, **25** of these peptides (**23** proteins) further defined as highly enriched.

Table 3.6: Summary of the trial site-ID purifications for *P. berghei* and *T. gondii*, and the analyses performed on the two datasets.

Of the 193 enriched *P. berghei* proteins identified in this dataset, 177 enriched *P. berghei* proteins had homologues in *P. falciparum* (the list of *P. berghei* and corresponding *P. falciparum* homologues was kindly provided by Lia Chappell, Berriman and Rayner lab). This list of 177 *P. falciparum* homologues of the *P. berghei* enriched proteins was compared with the *P. falciparum* Trial 4 site-ID enriched dataset, as this purification was the most similar to the *P. berghei* trial purification in terms of methodology and was also analysed using label-free quantification. Of the 177 homologues, 100 homologues were found to be present in the *P. falciparum* Trial 4 dataset, indicating that 56% of the *P. berghei* trial dataset overlapped with the *P. falciparum* Trial 4 dataset.

A selected set of 8 *P. berghei* proteins, and their peptides classified as enriched in this trial purification is shown in Table 3.7, along with the corresponding *P. falciparum* homologues, and their peptides which were classified as enriched in the *P. falciparum* Trial 4 purification. In this set of homologous pairs, the same homologous peptide was purified from both site-ID purifications for four of the homologous pairs. For the other four homologous pairs however, different peptides appeared to be enriched in the *P. berghei* dataset compared to their homologues in the *P. falciparum* dataset. However, as the *P. berghei* dataset was from only a single purification, it is unknown at this time whether different peptides are palmitoylated in some of the homologues of the different species, or the enrichment of different peptides was due to false positive identification or the inaccurate quantitation of the label-free approach. Further replicates are still required before it can be determined whether the same or different peptides are palmitoylated in the homologous pairs of these two *Plasmodium* species. Nevertheless, the purification of the same peptides for some of the enriched proteins, the good overlap with the *P. falciparum* Trial 4 dataset, and the presence of known palmitoyl-proteins in this *P. berghei* dataset indicates that this site-ID purification can be used for site-ID palmitome purification in *P. berghei*, although further optimisation is still required.

P. berghei				*P. falciparum* homologue			
Protein	Protein Name	Sequence	Enrichment Ratio	Protein	Protein Name	Sequence	Enrichment Ratio
PBANKA_141970	zinc finger protein, putative	VYESNNIFI**C**QGK	33.13	MAL13P1.117	DHHC-type zinc finger protein, putative	VYESNNIFF**C**K	186.76
PBANKA_111530	glideosome-associated protein 40, putative (GAP40)	FIG**C**GEEYMPIR	25.53	PFE0785c	metabolite/drug transporter, putative	FIG**C**GEEYMPIR	75.68
PBANKA_124060	membrane skeletal protein, putative	IA**CC**SGDNVFD	18.95	PFE1285w	membrane skeletal protein IMC1-related	NIY**C**SYA	9.01
		ENNY**C**TTVS**C**K	6.42			ESIMEVPEVN**C**PHIDISK	8.14
PBANKA_011220	myosin-like protein, putative	GYENI**C**YASFAK	88.27	PFF0675c	myosin E	QNFI**C**EFPSNYK	10.77
		F**C**TIHIENEEIK	2.23			AYEN**C**FFENR	7.22
						IGNMNEIS**C**YK	6.02
PBANKA_143760	glideosome-associated protein 45, putative	STTP**C**DIDKMDETAK	12.32	PFL1090w	glideosome-associated protein 45	SVTP**C**DMNK	18.14
PBANKA_130290	copper transporter, putative	SGNDE**C**KPIIDINHIGSEGK	4.77	PF14_0369	copper transporter putative	ISFMYN**C**WK	10.52
		CDVDPHKEIYGDPA**CC**G**C**	2.29			**C**SPDDIDVHK	8.60
PBANKA_132640	glyceraldehyde-3-phosphate dehydrogenase, putative	FP**C**EVTPTEGGIMVGNK	2.34	PF14_0598	glyceraldehyde-3-phosphate dehydrogenase	**C**QVDVV**C**ESTGVFITK	56.39
						QIIVSNAS**C**TTN**C**IAPIAK	44.76
						CAISNIIPASTGAAK	30.32
						VPIGTVSVVDIV**C**R	6.44
PBANKA_070770	conserved Plasmodium protein, unknown function	NFISGM**CC**GGAK	33.93	PF08_0035	conserved Plasmodium protein, unknown function	NYNTMVNDHNIN**C**TK	9.09
		HEDAIESNVT**C**FSK	21.37			NVN**C**NYEIINK	6.06
		YD**C**TNPTNYIQNYN**C**MK	17.50				
		CCNIPNF**C**R	17.08				

Table 3.7: A selected set of enriched proteins purified in the *P. berghei* trial site-ID palmitome purification, along with their *P. falciparum* homologues, present in the *P. falciparum* Trial 4 site-ID palmitome purification. The enriched peptides are shown for each enriched protein in both *P. berghei* and *P. falciparum* datasets, along with the enrichment ratios of each peptide. Enrichment ratios above the cut-off criteria for highly enriched proteins are shown in red and enrichment ratios above the cut-off criteria for enriched proteins are shown in blue. The putative palmitoylated cysteines are highlighted in bold and in green.

3.3.2. Trial site-ID palmitome purification from T. gondii tachyzoites

Next, the site-ID palmitome purification method was attempted on the tachyzoite stages of *T. gondii* parasites (a stage in the *T. gondii* life cycle which is similar to the schizont stages of *Plasmodium*). The proteome was extracted from approximately 10^8 *T. gondii* RH strain tachyzoite stage parasites (kindly provided by Karine Frenal, Soldati-Favre lab) using the same extensive extraction methods as before (described in the Materials and Methods and in Section 3.1.3). All other steps of the site-ID purification were performed exactly as for the *P. falciparum* Trial 4 site-ID palmitome purification (described in Section 3.1.6 and shown in the schematic in Figure 3.5A). Palmitome and control elutions were analysed label-free by LC-MS/MS as before, and MaxQuant was used to generate specific enrichment ratios (palmitome over control) for each of the peptide identifications.

In total, 227 peptides with enrichment ratios and containing a free cysteine were identified in this *T. gondii* trial purification, and this corresponded to a total of 164 proteins. The 227 peptides were grouped into enriched and highly enriched classes according to the same cut-off criteria used for the *P. falciparum* trial purifications (detailed in Section 3.1.2). Using these cut-off criteria, 81 peptides (corresponding to 69 protein identifications) were defined as enriched, and out of these 81 peptides, 25 peptides (corresponding to 23 protein identifications) could be further defined as highly enriched. A scatter plot showing the median protein intensities against the enrichment ratios for all the 227 peptides identified in this trial purification is shown in Figure 3.9B.

The *T. gondii* trial dataset was smaller than the previous datasets, and the average peptide intensity was again lower than previous purifications. In fact, the average peptide intensity was even lower than that of the *P. berghei* trial purification, indicating that an overall low amount of protein was present and further optimisation of the starting parasite material used in the purification is still required. Despite the small size of this dataset, proteins known to be palmitoylated in other systems, such as GAP45 and Bet3 transporter, were identified in this dataset. A summary of the analysis performed in the *T. gondii* trial dataset is shown in Table 3.6.

Additionally, the amino acid sequences of a selected set of seven *T. gondii* proteins, classified as enriched in this dataset, was used in a protein-protein BLAST search in order to determine potential *P. falciparum* homologues for these proteins. These *T. gondii* proteins, and their peptides enriched in this trial purification, are shown along with their corresponding potential *P. falciparum* homologues, and their peptides enriched in the *P. falciparum* Trial 4 purification, in Table 3.8. The same homologous peptide appeared to be enriched in both datasets for one of the homologous pairs. Conversely, different peptides were enriched in each dataset for two of the homologous pairs. For

the remaining 5 homologous pairs, both the same homologous peptides, as well as additional different peptides, were enriched in the two datasets. However, as mentioned in Section 3.3.1 above, further replicates are still required before any conclusions can be made.

T. gondii				*P. falciparum* homologue			
Protein	Protein Name	Sequence	Enrichment Ratio	Protein	Protein Name	Sequence	Enrichment Ratio
TGGT1_006420	transport protein particle component Bet3, putative	TGMGG**C**D**C**FR	5.65	PFD0895c	Bet3 transport protein, putative	SDISF**C**EDFEETVNVIAK	29.27
						VN**C**YFVK	10.26
TGGT1_033520	glyceraldehyde-3-phosphate dehydrogenase, putative	SSDVIVSNAS**C**TTN**C**IAPIAK	3.30	PF14_0598	glyceraldehyde-3-phosphate dehydrogenase	**C**QVDVV**C**ESTGVFITK	56.39
		VPVPDVSVVDIT**C**K	2.70			QIIVSNAS**C**TTN**C**IAPIAK	44.76
						CAISNIIPASTGAAK	30.32
						VPIGTVSVVDIV**C**R	6.44
TGGT1_078320	conserved hypothetical protein	**C**G**C**DIGDQHDENE**C**PI**C**R	15.39	PFL1090w	glideosome-associated protein 45	SVTP**C**DMNK	18.14
		SVVGYTVTP**C**DMAAIDETAK	11.52				
TGGT1_101540	conserved hypothetical protein	FIG**C**GEEYMPMR	28.75	PFE0785c	metabolite/drug transporter, putative	FIG**C**GEEYMPIR	75.68
TGGT1_107040	enolase, putative	IEESIGSD**C**QYAGAGFR	3.55	PF10_0155	enolase	SDQMVIPVP**C**INVINGGSHAGNK	94.67
						NA**C**NAIIIK	16.24
						IIGMN**C**TEQK	10.41
TGGT1_112260	cell division protein, putative	AVASE**C**SANFVSIK	2.23	PFF0940c	cell division cycle protein 48 homologue, putative	K**C**PVAENVPIDFIAQK	111.19
		IDEVGYDDIGG**C**R	2.20			HSTI**C**IIINDNDIDEGK	35.85
						TAGFSGADIAEI**C**QR	16.98
TGGT1_112840	heat shock protein, putative	IVDF**C**VQDFK	2.93	PF08_0054	heat shock 70 kDa protein	IVNF**C**VEDFKR	15.08
						NSIENY**C**YGVK	9.84
						EKIQPAEIET**C**MK	7.28

Table 3.8: A selected set of enriched proteins purified in the *T. gondii* trial site-ID palmitome purification, along with their *P. falciparum* homologues (homologues were determined by protein-protein BLAST search) found to be present in the *P. falciparum* Trial 4 site-ID palmitome purification. The enriched peptides are shown for each enriched protein in both *T. gondii* and *P. falciparum* datasets, along with the enrichment ratios of each peptide. Enrichment ratios above the cut-off criteria for highly enriched proteins are shown in red and enrichment ratios above the cut-off criteria for enriched proteins are shown in blue. The putative palmitoylated cysteines are highlighted in bold and in green.

Although further optimisation is still required for the *T. gondii* site-ID palmitome purification, the presence in this dataset of known palmitoylated proteins, as well as the presence of *P. falciparum* homologues enriched in previous datasets, indicate that this purification method can be used to determine sites of palmitoylation in *T. gondii* parasites.

For both the *P. berghei* and *T. gondii* trial purifications, further optimisation and certainly more replicates are still required before any truly relevant site-ID data can be obtained. Unfortunately, this was unable to be achieved before the writing of this dissertation, due to time constraints and due to the availability of mass spectrometry facilities. Nevertheless, these two trial purifications still provide valuable preliminary information on the sites of palmitoylation in *P. berghei* schizonts and *T. gondii* tachyzoites.

Conclusion

In this chapter, a method for the identification of palmitoylation sites within the palmitome of an organism was developed based on the ABE method of palmitoyl-protein purification. This site-ID palmitome purification method was successful in purifying palmitoylated peptides (containing the putative palmitoylated cysteine) in *P. falciparum, P. berghei* and *T. gondii* parasites. However, the continued presence of a relatively high background of cysteine-containing peptides in control samples, and the use of the less accurate label-free quantification method, have made analysis of the data more difficult, and require the treatment of datasets to be more stringent and rigorous. Although this method still requires some optimisation and refinement (for example, the reduction of the background of cysteine-containing peptides in control samples would be beneficial), some high and medium confidence palmitoylation sites have been identified in *P. falciparum* schizonts, and these data can now be validated experimentally and used to guide future experiments, such as those carried out in Chapter 6.

References:

1. Drisdel, R.C., et al., *Assays of protein palmitoylation.* Methods, 2006. **40**(2): p. 127-134.
2. Roth, A.F., et al., *Global analysis of protein palmitoylation in yeast.* Cell, 2006. **125**(5): p. 1003-1013.
3. Martin, B.R. and B.F. Cravatt, *Large-scale profiling of protein palmitoylation in mammalian cells.* Nat Methods, 2009. **6**(2): p. 135-138.
4. Jones, Matthew L., et al., *Analysis of Protein Palmitoylation Reveals a Pervasive Role in Plasmodium Development and Pathogenesis.* Cell Host & Microbe, 2012. **12**(2): p. 246-258.
5. Nirmalan, N., P.F. Sims, and J.E. Hyde, *Quantitative proteomics of the human malaria parasite Plasmodium falciparum and its application to studies of development and inhibition.* Mol Microbiol, 2004. **52**(4): p. 1187-1199.
6. Prieto, J.H., et al., *Large-scale differential proteome analysis in Plasmodium falciparum under drug treatment.* PLoS One, 2008. **3**(12): p. e4098.
7. Kovanich, D., et al., *Applications of stable isotope dimethyl labeling in quantitative proteomics.* Analytical and Bioanalytical Chemistry, 2012. **404**(4): p. 991-1009.
8. Zhu, W., J.W. Smith, and C.-M. Huang, *Mass Spectrometry-Based Label-Free Quantitative Proteomics.* Journal of Biomedicine and Biotechnology, 2010. **2010**: p. 1-6.
9. Boersema, P.J., et al., *Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics.* Nat Protoc, 2009. **4**(4): p. 484-494.
10. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.
11. Gardner, M.J., et al., *Genome sequence of the human malaria parasite Plasmodium falciparum.* Nature, 2002. **419**(6906): p. 498-511.

12. López-Barragán, M.J., et al., *Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum.* BMC Genomics, 2011. **12**(1): p. 587.
13. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review.* Analytical and Bioanalytical Chemistry, 2007. **389**(4): p. 1017-1031.
14. Li, Z., et al., *Systematic Comparison of Label-Free, Metabolic Labeling, and Isobaric Chemical Labeling for Quantitative Proteomics on LTQ Orbitrap Velos.* Journal of Proteome Research, 2012. **11**(3): p. 1582-1590.
15. Patel, V.J., et al., *A comparison of labeling and label-free mass spectrometry-based proteomics approaches.* J Proteome Res, 2009. **8**(7): p. 3752-9.
16. Wilson, J.P., et al., *Proteomic Analysis of Fatty-acylated Proteins in Mammalian Cells with Chemical Reporters Reveals S-Acylation of Histone H3 Variants.* Molecular & Cellular Proteomics, 2010. **10**(3): p. M110.001198-M110.001198.
17. Schwartz, D. and S.P. Gygi, *An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets.* Nature Biotechnology, 2005. **23**(11): p. 1391-1398.