CHAPTER 3

TRANSCRIPTIONAL PROFILING OF CO-INHIBITED PfADOMETDC/ODC

3.1 INTRODUCTION

3.1.1 Transcriptional profiling of perturbed *P. falciparum* compared to other organisms

Transcriptional profiling can be used to assess the response of cells or organisms to environmental stress, which can identify feedback mechanisms, alternative pathways and metabolic buffering systems activated to cope with a perturbation [136, 187]. The data can demonstrate how specific compounds affect regulatory networks of cellular metabolism and their effects on particular metabolic pathways [188]. The approach was applied with great success in the case of *Mycobacterium tuberculosis* [138] and *S. cerevisiae* [139]. Studies on *M. tuberculosis* revealed transcriptional signatures specific to the mode of action for several antimycobacterial drugs [138] and for *S. cerevisiae* it was shown that the transcriptional changes associated with target deletion or under-expression should theoretically mimic the effect of chemically inhibiting that target [188]. The latter makes microarray studies of drug perturbed *P. falciparum* an attractive alternative compared to gene deletion, in particular since genetic manipulation of this organism faces unique technical problems and a low success rate due to its A+T-richness and intracellular location [130]. In contrast to *M. tuberculosis* [138] and *S. cerevisiae*, the correlation of perturbation-specific events in the *P. falciparum* transcriptome was initially limited to only a few studies in malaria research [179, 189-191]. This could have been related to the multistage nature of the parasite, which challenged the experimental design of microarray and other functional genomics investigations. The limited evidence of compensatory mechanisms, the paucity of transcription factors identified in the *P. falciparum* genome [119], the small amplitude of transcriptional responses observed in *Plasmodium* upon perturbation compared to other organisms, e.g. *M. tuberculosis* [137, 138], and evidence of post-transcriptional control [128, 192, 193] (as discussed in section 4.1), caused doubt regarding the role of transcriptional regulation in the parasite. It was perceived that the parasite's transcriptional control was "hard-wired" and that transcriptional profiling during environmental perturbations was futile [128, 194, 195]. This perspective was corroborated by several perturbation-investigations of the parasite that failed to detect programmed transcriptional responses [194, 196, 197].

However, evidence supporting the role of transcriptional control in the parasite is now mounting. One of the first studies of plasmodial transcriptional profiling employed serial analysis of gene expression (SAGE) and monitored the transcriptional response of asexual parasites to 6 h chloroquine treatment [179]. More than 100 transcripts were identified as differentially affected and included the increased abundance of the transcript for

the multidrug resistance gene, *pfmdr1*, which underscored its importance in chloroquine resistance. This investigation indicated the existence of a compensatory feedback mechanism, signalling a transcriptional response to chloroquine [179]. A subsequent follow-up report identified 600 drug-responsive genes after treatment with chloroquine [196]. Since then, there have been a number of microarray studies which reported transcriptional compensation in response to environmental stress in *P. falciparum*. These include the increase of the transcript for a *rcd1*$^{+}$ homologue (regarded as glucose-specific) upon glucose deprivation [189] and the increased transcript abundance of two heat shock protein orthologues after exposure to elevated temperature [190]. Other microarray studies detected perturbation-specific, albeit not compensatory, effects including the specific impairment of transcription of apicoplast genes after parasites were treated with doxycycline [62] and enrichment of the differentially affected transcripts for histone H3K9 acetylation after treatment with the histone acetyltransferase inhibitor, anarcadic acid [191]. Recently, treatment with artesunate was reported to cause a transcriptional death response, but it is not clear whether the response observed was specific or generalised cytotoxicity [198]. Finally, transcriptional profiling of sphingomyelin biosynthesis inhibition, which similar to polyamine depletion results in developmental arrest in the trophozoite stage, indicated the involvement of an previously unknown tubovesicular network export-protein that was subsequently shown to be important for lipid import and parasite growth [199]. These studies demonstrated that the malaria parasite is able to respond to environmental perturbation in the transcriptome and that perturbation-specific transcriptional responses can be distinguished from the basal “just-in-time” level of transcriptional control.

3.1.2 Transcriptomics methodologies, experimental design and data analysis

Genome sequencing of humans and various organisms has provided large amounts of data and various techniques were developed to exploit this growing body of knowledge to the maximum. These include SAGE, representational difference analysis (RDA), complementary DNA (cDNA) microarrays and oligonucleotide microarrays, all of which enable the simultaneous analysis of the transcript abundance from thousands of genes [200]. SAGE is based on the isolation of unique sequence tags from individual transcripts and serial concatenation of these tags into long DNA molecules. Sequencing of the concatemer clones reveals the individual tags, which enables the rapid identification and quantitation of the cellular transcripts [201]. In RDA, the difference between two complex genomes is determined by subtractive and kinetic enrichment of restriction endonuclease fragments that are present in the one DNA population but not in the other [202]. cDNA microarray experiments involve the competitive hybridisation of two mRNA samples that have been converted into cDNA and each labelled with its own fluorescent dye (e.g. Cy3 and Cy5). The hybridisation occurs on a glass slide or chip spotted with cDNA probes. The resulting data provide information on the relative transcript abundance of the sample genes [203]. In oligonucleotide microarray, oligonucleotides are spotted or synthesised onto glass slides instead of cDNA probes. With two-colour/channel oligonucleotide arrays the relative expression between two differentially labelled samples is determined as with cDNA

microarrays, but for single-colour/channel arrays the absolute gene expression values of a single sample are determined [203].

Operon Biotechnologies (Cologne, Germany) provides commercial 70-mer spotted arrays or array-ready oligonucleotide sets for in-house spotting. In addition to spotted arrays, various other competing technologies for commercial oligonucleotide microarrays have emerged. These include the use of either full-length cDNAs, or presynthesised/*in situ* synthesised oligonucleotide probes [204]. The pioneering work in this field was performed by Affymetrix (Santa Clara, California, USA) and consequently their GeneChips are regarded as the optimal method for transcriptional profiling. The GeneChip technology uses a combination of light-directed (photolithography) and solid-phase DNA synthesis of 25-mer oligonucleotides *in situ* on the slide. Agilent (Palo Alto, California, USA) technology synthesises 60-mer oligonucleotides *in situ* at or near the surface of the slide by inkjet printing using phosphoramidite chemistry, whereas NimbleGen (Madison, Wisconsin, USA) technology uses a digital micromirror device instead of the photolithographic masks for *in situ* synthesis of 24 to 70-mer oligonucleotides. A major advantage of oligonucleotide arrays is that they are designed *in silico* without the need for clone libraries and tedious sequencing for identification purposes [204]. The longer 60 to 70-mer oligonucleotide lengths provide higher sensitivity compared to the shorter 24 or 25-mers [204].

In microarray experimental design a key issue is whether to use direct or indirect comparisons i.e. to make the comparison within or between slides. In direct comparison the differential transcript abundance between two samples is directly measured on the same slide, but in indirect comparison the differential abundance of the two samples is determined on separate slides compared to a common reference. The common reference can be an external RNA or cDNA reference pool or a composite pool containing all the samples. In time course microarray experiments RNA is extracted from samples harvested at several time points after a form of treatment or stimulation. These samples can also be compared to a common reference or directly using a sequential or loop design (Fig. 1.9, Fig. 3.1). The experimental design usually depends on the number of hybridisations to be performed, i.e. the number of time points and the number of replicates. When the main focus is on the relative changes between the time points, the reference design is the better choice. In contrast to the sequential and loop designs, the reference design has the further advantage that the normalised data (i.e. $\log_2$-ratios) can be directly compared (through the reference), which simplifies data analysis and interpretation [203].

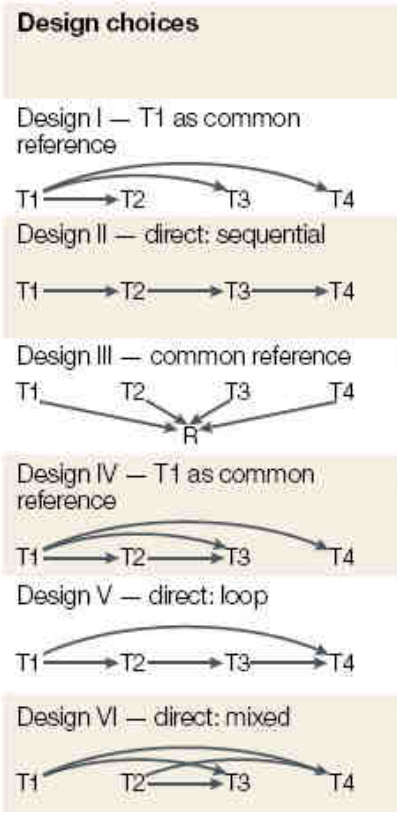


Fig. 3.1 Six designs of microarray time course experiments (image obtained from [203]). Designs I and II require only three slides, whereas the other four designs require four [203].

After hybridisation, the data are collected by scanning the two independent fluorescent images (e.g. Cy3 and Cy5) of the co-hybridised species (e.g. one for the sample and one for the reference) from each slide at high resolution. The spot relative fluorescence intensities (e.g. Cy5/Cy3) are normalised during scanning by adjusting the photon multiplier tube (PMT) settings to compensate for differences in labelling and detection efficiencies (i.e. dye bias) such that Cy5/Cy3 ≈ 1. The quality of every spot is inspected and bad quality spots are flagged and subsequently removed from the dataset. Subsequently, the data for each gene is typically reported as a $\log_2$ expression-ratio (Eq. 3.1). The data can then be explored via different analyses, such as clustering, to identify genes that are co-regulated, i.e. with similar expression profiles [200].

$$\text{GeneX } \log_2\text{-ratio} = \log_2 \left\{ \frac{\text{normalised value of treated sample}}{\text{normalised value of reference}} \right\} \qquad \text{……Equation 3.1}$$

For differential transcript abundance analysis, various software packages are available for normalisation and statistical analysis, such as the R statistical environment (http://www.r-project.org/). Data normalisation for two-colour arrays will be discussed. Within R, data are normalised, first within the same slide to remove local background artefacts or edge effects. Various diagnostic plots can be compiled to visualise the data before and after normalisation to determine the most appropriate normalisation method to apply. For closely related samples with similar expression, a scatterplot of Cy5 versus Cy3 (or their logarithms) would result in data

distribution in a straight line approximately (slope ≈ 1) if labelling and detection for both samples were equal. However, this is not necessarily true and normalisation involves calculating the best-fit slope using regression techniques such as locally weighted scatterplot smoothing (LOWESS) [200]. Global LOWESS normalisation is usually performed if the background is homogenous, but if there are significant differences in the distribution of $\log_2$-ratios among the print-tips (visualised as print-tip boxplots, Fig. 3.14), print-tip LOWESS is recommended [205]. Robust spline normalisation is a compromise between print-tip and global LOWESS [206], which uses regression splines instead of LOWESS curves to shrink the individual print-tip curves towards a common value according to Bayesian rules. This technique introduces little variation into good quality arrays with limited spatial variation (localised artefacts), but also improves the data from arrays with significant spatial variation [206]. After within-slide normalisation, between-slide normalisation is performed, such as quantile normalisation or scaling, to standardise the distribution of $\log_2$-ratios across the different slides [205]. Quantile normalisation was initially proposed for Affymetrix-style single-colour arrays and ensures that the fluorescence intensities (i.e. Cy3 or Cy5) have the same empirical distribution across both fluorescent channels for all the arrays (Fig. 3.16). Rquantile or Gquantile normalisation is particularly useful when using a reference design to normalise the common reference values across all the slides. Rquantile is used when the reference is labelled with Cy5 (red) or Gquantile when labelled with Cy3 (green). With every data analysis step diagnostic plots are compiled and evaluated to determine the suitability of the data transformation. When satisfactory distributions of $\log_2$-ratios are obtained, the replicate data are consolidated and differential transcript abundance (also referred to as differential expression, but actually the function of mRNA expression and decay) between samples can be calculated. With a reference design, the differential transcript abundance of any sample combination can be determined (e.g. treated compared to untreated or to a time zero).

Microarray data should be reported in sufficient detail, including information on the applied methodology and analysis. This ensures the simple interpretation and independent verification of the data as stipulated by the minimum information about a microarray experiment (MIAME) convention [207]. Public microarray database repositories such as ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae) and the NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) require detailed MIAME information during data deposition, which enables meta-analyses across different microarray platforms and comparison between different datasets.

Presented here is transcriptional profiling of *P. falciparum* during cytostasis using a spotted, 70-mer oligonucleotide microarray platform and employing a reference experimental design. Cytostasis was induced by polyamine depletion via co-inhibition of both catalytic sites of PfAdoMetDC/ODC with DFMO and MDL73811. These drugs are specific inhibitors of PfAdoMetDC/ODC [147] and therefore the transcriptional profiles obtained should be characteristic of the effects of polyamine-depletion on the parasite [138]. A reference point for quantitative analysis of differential expression was defined and several transcripts with

altered profiles could be detected, including some from polyamine and methionine metabolism. Transcriptome analysis revealed polyamine-specific compensatory responses to alleviate the perturbation, which supports the role of transcriptional regulation in polyamine and methionine metabolism of *Plasmodium*.

3.2 MATERIALS AND METHODS

3.2.1 Ensuring the correct treatment dosage for the transcriptomics investigation

3.2.1.1 Growth morphology studies

3D7 *P. falciparum* cultures were maintained *in vitro* as discussed in section 2.2.1. A small-scale morphology study was performed to ascertain that complete arrest occurred at the treatment concentrations and to determine the exact sampling times for the transcriptomics investigation. Parasites were synchronised by consecutive sorbitol treatments as discussed in section 2.2.2 and treated in the late schizont stage with 5 mM DFMO and 5 μM MDL73811 and the combination thereof. The parasite morphology was monitored microscopically at 6 h intervals for a complete 48 h life cycle using Giemsa-stained thin smears.

3.2.1.2 Radio-labelled substrate assays

To ensure complete enzyme inhibition of DFMO/MDL73811-treated (T) and untreated (UT) 3D7 cultures (sampled at ~10% parasitaemia, 3% haematocrit) at 5 mM DFMO and 5 μM MDL73811, the decarboxylase activities of AdoMetDC and ODC were determined after PfAdoMetDC/ODC co-inhibition according to the original method of Assaraf and colleagues [88]. In principle, the ^{14}C-labelled substrates *S*-adenosyl-L-[^{14}C]methionine (56.2 mCi/mmol, Amersham Biosciences, Buckinghamshire, England) and L-[1-^{14}C]ornithine (47.7 mCi/mmol, Amersham Biosciences) were incubated with the culture lysates and the release of $^{14}CO_2$ was measured. Cultures were sampled in the early (19 hpi) and mature trophozoite (34 hpi) stages. Samples of 10 - 15 ml were centrifuged at 2500 g for 5 min and the pellet washed three times with an equal volume of PBS, after which 500 μl was transferred to a cryotube and stored at -70°C. Uninfected erythrocytes were sampled and processed in the same way to serve as a negative control. A volume of 1 ml buffer A (40 mM Tris-HCl, 1 mM dithiothreitrol (DTT), 1 mM EDTA, 0.1 mM phenylmethanesulphonylfluoride, pH = 7.4) [88] was added, the samples freeze-thawed three times (alternating between -70°C and 37°C) and then centrifuged at 8000 g for 20 min at 4°C. The cell lysate supernatants were aspirated, mixed and kept on ice. Reactions were performed in 50 ml glass tubes in duplicate. Whatman 2 filter paper (Merck) was folded lengthwise and inserted into 2 ml open-ended microfuge tubes onto which 40 μl hydroxide of hyamine (PE Applied Biosystems, California, USA) was absorbed to trap released $^{14}CO_2$. Two hundred microlitres of cell lysate supernatant was pipetted to the bottom of the glass tubes on ice. This was followed by 50 μl of reaction mixture containing 7.2 μM ^{14}C L-ornithine (100 nCi) and 40 μM pyridoxal-5-phosphate (PLP) in buffer A, or 7.2 μM ^{14}C AdoMet (50 nCi) only, in buffer A. Filter paper-containing tubes (standing upright inside the glass tubes) were inserted at the bottom of the glass tubes, followed by a rubber stopper to prevent CO_2 escape.

The assays were allowed to take place at 37°C for 30 min (ODC) or 60 min (AdoMetDC) in a ZHWY-110X shaking water bath (Shanghai ZHICHENG Analytical Instruments Manufacturing Co., Shanghai, China). The reactions were terminated with the injection of 500 µl 30% trichloro-acetic acid and free $^{14}CO_2$ was diluted with regular CO_2, by the addition of 500 µl 0.1 M $NaHCO_3$. The tubes were once again incubated for 30 min (ODC) or 60 min (AdoMetDC) at 37°C. The filter papers were transferred into 4 ml Pony-Vial H/I tubes (PE Applied Biosystems) and 4 ml of Ultima Gold XR scintillation fluid (PE Applied Biosystems) was added. The radioactivity was determined with a Tri-Carb series 2800 TR liquid scintillation counter (PE Applied Biosystems) until at least 10000 events were recorded for each sample. The average disintegrations per minute (DPM) was calculated from the average counts per minute (CPM) corrected for quenching as determined by ^{14}C-standards. The results were analysed with QuantaSmart (PerkinElmer, Connecticut, USA) software and total activity was calculated (Eq. 3.2)

Total activity = DPM x nmol substrate / ml cell lysate x min incubation time ……Equation 3.2

3.2.2 Drug treatment for the transcriptomics investigation

Parasites were synchronised by consecutive sorbitol treatments, as described in section 2.2.2, for three generations [171]. Drug treatment with the combination of 5 mM DFMO and 5 µM MDL73811 occurred in the late schizont stage (42 hpi) at ~2% parasitaemia and 3% haematocrit. After schizogony these parasites proliferated up to a parasitaemia of about 10% in both the treated and untreated cultures. Treatment was performed in duplicate (i.e. two biological replicates, assigned A and B) alongside untreated controls. Culture medium (with and without drug) was replaced halfway through the time course (i.e. about 18 hpi), but before the first sampling, to prevent metabolic stress of the parasites. Drug treated and untreated samples of 15 ml (10% parasitaemia and 3% haematocrit) were harvested at three time points within the trophozoite stage (t_1 = 19 hpi, t_2 = 27 hpi and t_3 = 34 hpi), based on the morphology of untreated parasites after microscopic inspection of Giemsa-stained thin smears. These were centrifuged at 2500 g for 5 min, the supernatant aspirated and replaced with 15 ml PBS. The pellet was resuspended by pipette-mixing, centrifuged at 2500 g for 5 min and the supernatant aspirated once more. The PBS-washed infected erythrocyte pellet was stored at -70°C until RNA isolation could be performed.

3.2.3 RNA isolation

RNA was isolated under RNAse-free conditions according to the Chomczynski and Sacchi [208] method with a few additional lysis steps. The monophasic, phenol/guanidine thiocynate solution in TRI-Reagent inhibits RNase activity and lyses the sample material. The homogenate is then separated into aqueous and organic phases by the addition of chloroform followed by centrifugation. RNA partitions in the aqueous phase, DNA in

the interphase and proteins in the organic phase. The RNA was then precipitated from the aqueous phase with ethanol (EtOH).

Total RNA was extracted from frozen infected erythrocyte pellets collected from 15 ml culture samples with the RNeasy (Qiagen, Germany) kit. The erythrocytes were lysed by freeze-thaw, but to ensure optimal parasite lysis three additional steps were performed. A volume of 1.2 ml lysis buffer was added to each pellet, which was then centrifuged for 2 min at 15700 g through a QIA-Shredder column (Qiagen) for physical rupture. The flow-through was split in two aliquots and 600 µl TRI-Reagent was mixed with each in a 2 ml microfuge tube. After 5 min incubation at room termperature, 400 µl chloroform was added, vortexed and incubated at room temperature for 10 min. The samples were centrifuged at 15700 g for 15 min and the upper aqueous phase of each was transferred to a clean tube without disturbing the interphase. The RNA was precipitated with 700 µl 70% EtOH; the aliquots were combined and loaded onto RNeasy colunns. The RNA was washed several times according to the kit instructions with two different buffers and a 15 min on-column treatment with 27 units (U) DNaseI (Qiagen) was performed to degrade any residual gDNA contamination of the RNA. The RNA was finally eluted with RNase-free H_2O and the concentration was determined by measuring the absorbance at 260 nm by ultraviolet (UV) spectrophotometry with a NanoDrop-1000. The absorbance is multiplied by one absorbancy unit, which equals 40 ng/µl for RNA [209]. Purity from protein contamination was estimated from the 260 nm/280 nm ratio, which should be 1.8 – 2, as proteins have maximum absorbance at 280 nm [209]. RNA was stored at -70°C to retain stability until cDNA synthesis.

RNA integrity was assessed with denaturing formaldehyde/agarose gel electrophoresis. All buffers were prepared using diethyl pyrocarbonate (DEPC)-treated H_2O (3 h treatment at 37°C with 0.1% DEPC in MilliQ H_2O and autoclaved twice to eliminate residual DEPC), glassware was rinsed with 0.5 M NaOH and disposable plastic was autoclaved twice. Gels consisting of 1.2% (w/v) agarose in FA buffer [20 mM 3-(N-morpholino)propanesulfonic acid (MOPS, SIGMA), 5 mM sodium acetate, 1 mM EDTA] with 0.67% (v/v) formaldehyde and 0.1 µg/ml ethidium bromide were prepared. Samples of 2 µg RNA were mixed with 5 x RNA loading buffer [0.16% (w/v) bromophenol blue in 20% (v/v) glycerol containing 4 mM EDTA (pH 8), 2.66% (v/v) formaldehyde and 30.84% (v/v) deionised formamide], incubated at 65°C for 5 min and then cooled on ice before loading onto the gel. The formaldehyde and the formamide in the buffers function to denature RNA secondary structures. Electrophoresis was performed using a 10 x 14 cm electrophoresis system with FA buffer. A voltage of 5 V/cm gel was applied for 2 h (~50 V) and the gel was visualised on a UV transiluminator (Spectroline TC-312 A) at 312 nm.

3.2.4 cDNA synthesis

Due to the limited amount of RNA (15 to 20 μg) obtained from the early time points, an RNA pool consisting of equal ratios of all twelve samples to be used as common reference (reference design microarray experiment) could not be compiled. The 12 μg RNA required for cDNA synthesis per sample was therefore aliquoted and the remaining RNA from both sets of biological replicates (A, B) was pooled to obtain a representative RNA reference pool (144 μg RNA in total). First-strand cDNA synthesis was initiated from 12 μg of total RNA with 775 pmol random primer 9 (New England Biolabs, Massachusetts, USA) and 250 pmol oligo-dT_{25} (Integrated DNA Technologies, Iowa, USA) by incubation at 70°C for 10 min followed by cooling on ice for 10 min. Reverse transcription and amino-allyl incorporation were performed simultaneously as described [210], but the reaction time was extended overnight at 42°C using 480 U of SuperScript III RNase H^- reverse transcriptase (Invitrogen), 10 mM DTT and a mixture of deoxynucleotide triphosphates (dNTPs) in a 2A:1C:1G:2U/T ratio [1 mM deoxyadenosine triphosphate (dATP), 500 μM deoxycytidine triphosphate (dCTP), 500 μM deoxyguanosine triphosphate (dGTP), 500 μM deoxythymidine triphosphate (dTTP) and 500 μM aminoallyl-deoxyuridine triphosphate (dUTP) from Fermentas Life Sciences, Ontario, Canada]. The high A+T mixture of dNTPs was used to provide sufficient nucleotides for A+T-rich cDNA synthesis of *P. falciparum*. Contaminating RNA was removed by hydrolysis with 0.5 M EDTA and 1 M NaOH at 65°C for 15 min and the reactions were purified with the Wizard SV Gel and PCR Clean-Up System (Promega, Wisconsin, USA) according to the manufacturer's recommendations. The kit works on the principle that DNA binds to a silica matrix in the presence of chaotropic salts [211]. These salts are then removed with an alcohol-based wash buffer and the DNA is eluted in a low-ionic-strength solution or water. In the order of 6 – 8 μg cDNA was obtained from 12 μg RNA starting material. The cDNA was divided into aliquots of 2 μg each, which were dried under vacuum and stored at -20°C to retain stability.

3.2.5 Oligonucleotide array spotting and slide post-processing

Long 70-mer oligonucleotides, previously designed based on uniqueness within the genome and proximity to the 3' end of the gene open reading frames [210], were spotted in-house at the Lewis-Sigler Institute Microarray Facility (Princeton University) onto poly-lysine coated glass slides. The oligonucleotide identifiers and sequences are provided (Supplementary CD). The oligonucleotides were cross-linked onto the slides by exposure to 60 mJ UV in a Stratalinker. Post-processing was performed as described [210] by soaking the slides in a mixture of 3xsaline sodium citrate (SSC) and 0.2% sodium-dodecylsulphate (SDS) at 65°C for 5 min followed by MilliQ H_2O for 30 s. They were rinsed in EtOH and then dried by centrifugation at 200 g for 5 min. The oligonucleotide spots were rehydrated by soaking the slides for 1 min in 0.5xSSC and the slides were then quick-dried on a heat block at 99°C for 1 - 2 s. Blocking solution was prepared by dissolving succinic acid anhydride in 1-methyl-2-pyrrolidinone to obtain a 1.6% (w/v) solution and adding 15 ml 1 M sodium borate (pH 8) immediately afterwards as buffer. The slides (on slide-rack) were plunged vigorously up and down in the

buffering solution for 30 s while keeping them beneath the level of the solution. These were incubated for another 5 min on a horizontal shaking platform at low speed and subsequently removed from the solution and the excess blocking solution drained by lifting and tilting the rack for 5 s. The slides were submerged into MilliQ H_2O and gently pushed back and forth for 5 s and then incubated for 60 s. The blocking step was repeated with the same buffered blocking solution and plunging for 30 s, followed by incubation with gentle shaking for 5 min. The slides were again rinsed in MilliQ H_2O with plunging for 10 s, followed by a brief rinse with 95% EtOH and subsequently dried by centrifugation for 5 min at 200 g.

3.2.6 Cy dye cDNA labelling

The aminoallyl-dUTPs that were incorporated during cDNA synthesis were coupled to either Cy3 (reference pool) or Cy5 (samples) fluorescent dyes (Amersham Biosciences) in 0.1 M $NaHCO_3/Na_2CO_3$ (pH = 9.0) for a minimum of 1 h in the dark at room temperature. Free dye was removed with DNA Clean and Concentrator-5 columns (Zymo Research, California. USA). The kit works on the same principle as the Wizard SV Gel and PCR Clean-Up System Wizard SV Gel and PCR Clean-Up System (described in section 3.2.4). The samples were eluted through the columns and loaded onto the membrane a second time before washing with 200 µl wash buffer. The wash step was repeated and the samples eluted with 20 µl elution buffer. The fluorescently labelled cDNA was kept in the dark for a brief period (1-2 hours) until the hybridisations were set up. Labelling efficiency was calculated (Eq. 3.3) and should ideally be at least 10 labelled nucleotides in 1000 for hybridisation purposes.

$$\text{Labelling efficiency (number of labelled nucleotides per 1000)} = \frac{\text{pmol dye} \times 324.5\ \text{pg/pmol}^a}{\text{ng DNA}} \quad \ldots\ldots\text{Equation 3.3}$$

a. The average mass of a dNTP

3.2.7 Oligonucleotide array hybridisation, washing and scanning

The methodology of Bozdech and colleagues was followed as briefly described [210]. Post-processed slides were positioned (array-side upwards) inside a hybridisation chamber and the arrays were covered by clean dust-free lifter slips (Erie Scientific Company, New Hampshire, USA). The probe mixture was prepared by combining 25 pmol of Cy5-labelled sample and 25 pmol Cy3-labelled reference with 0.5xSSC, 0.8 mg/ml polyadenylic acid (SIGMA), 26 mM HEPES (pH 7) and 0.24% (w/v) SDS. These were denatured in a heat block at 99°C for 2 min and then cooled for 5 min at room temperature. The ~50 µl labelled cDNA probe mixture was slowly injected underneath the lifter slip covering the array, taking care not to inject bubbles. Immediately after injecting all the probes, the chamber was covered and sealed by tightening the screws. The chamber was kept horizontal and submerged in a water bath at 65°C for overnight incubation (20 h). The chambers were subsequently quickly removed from the water bath and unscrewed while being kept horizontal. The slides had been submerged one by one in washing solution 1 [0.6xSSC, 0.03% (w/v) SDS] in such a way

that the lifter slip gently lifted and settled to the bottom of the dish without scratching the arrays. After all the slides were submerged and positioned inside the slide rack, it was plunged up and down for 30 s and then incubated for about 1 min. The slides were individually transferred into another rack into washing solution 2 (350 ml MilliQ H_2O with 0.06xSSC) and as before the slide rack was plunged up and down for 30 s followed by incubation for 1 min. The slides were then dried individually in a high-speed ArrayIt microarray centrifuge (TeleChem International, Inc., California, USA). The arrays were scanned with an Axon GenePix 4000A scanner (Molecular Devices, California, USA) and the images were analysed with Axon GenePix Pro 6.0 software (Molecular Devices).

3.2.8 Data analysis

3.2.8.1 Exploratory data analysis

The GenePix default flagging parameters were applied in combination with visual inspection to assess and define spot quality. Array data were stored in the Princeton University Microarray database (PUMAdb, http://puma.princeton.edu). The normalised Cy5/Cy3 $\log_2$-ratios of the data were retrieved from PUMAdb. The technical replicates were averaged (arithmetic mean) in EXCEL and data were filtered in CLUSTER 2.1.1 [212] for transcripts with 100% data present in all the time points. Hierarchical data clustering was performed using uncentered correlation and average linkage to calculate the distance between clusters in CLUSTER 2.1.1 with display in TREEVIEW 1.0.12 [212]. Hierarchical clustering is an unsupervised, agglomerative method that joins the transcript data of single genes/oligonucleotides to form groups, which are further joined and eventually results in a single hierarchical tree that can be easily visualised and interpreted [200]. The distance or similarity measure used were based on the Pearson correlation, as defined in section 2.2.3.3.1, but instead of the standard calculation (centered), uncentered correlation was performed, which assumes that the mean is zero even if not. This compensates for vectors (transcript profiles) with the same shape that are offset relative to each other by a fixed value and thus results in a standard Pearson correlation for these [212]. Average linkage indicates that the distance between the clusters were calculated using average values as opposed to minimum (single linkage or nearest neighbour) or maximum (complete linkage or furthest neighbour) values [200]. All treated and untreated data were clustered together but the treated data were also clustered with UT_{t1} (relative t_0) only. A phaseogram was compiled from mean centered, $\log_2$-ratios that were ordered in PERL according to the phase of gene expression in the 3D7 IDC transcriptome (http://malaria.ucsf.edu) [91] and displayed in TREEVIEW 1.0.12. The phase and frequency of expression of the IDC transcriptome was determined with fast Fourier transformatioin [29, 91], a complex mathematical procedure that is used to isolate individual components of signals (e.g. waves) to determine the amplitude and phase of a particular frequency component. Pearson correlation coefficients (r) are commonly used in global expression profiling [29, 62, 91] to determine the similarity between sample profiles, between replicates etc. and were calculated in EXCEL.

3.2.8.2 Differential transcript abundance analysis

3.2.8.2.1 Linear models for microarray data (LIMMA) analysis

For differential transcript abundance analysis, data quality and normalisation methods were evaluated using data diagnostic tools from the MARRAY software package in R [213] on GenePix data and flagged values received a zero weight. Background subtraction (offset = 50) and robust spline normalisation were applied within each array, followed by Gquantile normalisation between arrays due to the common reference design of the microarray experiment. Differential abundance analysis was performed with the linear models for microarray data (LIMMA) software package within R (http://www.r-project.org/, [206, 214]). With a common reference design, LIMMA is similar to ordinary analysis of variance or multiple regression except that a linear model is fitted to the data for every oligonucleotide. Significance is calculated with moderated t-statistics using a simple Bayesian model in order to make the analyses robust even for a small number of arrays [206]. Differential abundance was calculated compared to UT_{t1}, defined as relative t_0. Genes in at least one treated time point with transcript abundance greater than 1.7-fold (log_2-ratio ≥ 0.75 or ≤ -0.75) in either direction compared to relative t_0, and p-values (adjusted for multiple hypothesis false discovery rate) of less than 0.05, were regarded as differentially affected. Data within these limits of transcripts represented by multiple oligonucleotides were averaged.

3.2.8.2.2 EDGE time course analysis

Differential abundance analysis was in addition performed with EDGE 1.1.208 software [215] on log_2-ratios normalised in PUMAdb. EDGE was specifically designed for microarray time course experiments. It models the expression over time and calculates the statistical significance by considering sources of dependence over time [215]. EDGE was applied with a Q-value cut-off of 0.1% ($p<0.02$) for significance and no fold change cut-off. Since the software was designed to determine differential expression over time e.g. treated over three time points versus untreated over three time points, it had to be forced to compare all the treated time points to UT_{t1} (relative t_0) by defining the covariates such that instead of providing the replicate data of UT_{t2} and UT_{t3}, the data of UT_{t1} were repeatedly used for all three untreated time points.

3.2.8.3 Additional data analysis

The differentially affected transcripts (LIMMA dataset) were classified into functional groups using gene ontology (GO) terms obtained from the database for annotation, visualisation and integrated discovery (DAVID, http://david.abcc.ncifcrf.gov/, [216]) and PlasmoDB 5.3 (http://www.plasmodb.org, [217]). The GO identifiers were sorted in EXCEL.

The LIMMA dataset was also compared with the PlasmoDB 5.3 general feature format file (.gff file) to search for clusters of adjacently located genes with differentially affected transcripts, where a cluster was regarded as four or more genes within a window of six adjacent genes. The analysis was performed in EXCEL.

To investigate the possible enrichment of the LIMMA dataset for transcripts of proteins functionally connected to polyamine and methionine metabolism, the dataset was manually compared with the *in silico* predicted *P. falciparum* interactome [113] of PfAdoMetDC/ODC with DHPS/PPPK (PF08_0095) as control.

3.2.9 Real-time PCR validation of differential transcript abundance data

The differential abundance analysis was validated by performing the polymerase chain reaction (PCR) in real-time using a LightCycler 1.5 and FastStart DNA MasterPLUS SYBR Green I kit (Roche Diagnostics, Mannheim, Germany). Primers with a melting temperature (T_m) around 55°C and a product length of 150 - 170 bp were designed using Oligo v.6.71 software (Table 3.1). The relative abundance of three transcripts with increased (PFL1885c, PFD0285c, PFF0435w) and three with decreased (PF08_0131, PFD0830w, PFI1090w) abundance levels were confirmed. Real-time PCR was performed in triplicate on the cDNA sample with maximum fold change compared to relative t_0. Comparable starting levels of cDNA across the different samples were obtained by adjusting template concentrations according to the relative amount of a putative cyclophilin (PFE0505w). The thermocycling programme was as follows: A 10 min pre-incubation period at 95°C was followed by 35 cycles of 10 s at 95°C, 5 s at 55°C and 7 s at 72°C, according to the manufacturer's instructions. After amplification, melting curve analysis was performed to exclude primer-dimer interference. Fold change was calculated compared to relative t_0.

Table 3.1 Real-time PCR primer information

PlasmoDB	Annotation	5' - 3' Primer sequence	T_m[a]	Size (bp)
PFE0505w	Cyclophilin Fwd	AAT TCT TTG ACC ATC TTA ATC ATT C	54.8	167
	Cyclophilin Rev	CAA AAC AAT TTT ACT TCC TTG GGT TA	56.9	
PFD0285c	Lysine decarboxylase Fwd	AGA GGG ATA TGG ATT GGT AGA	55.9	161
	Lysine decarboxylase Rev	TTC TCT TCA TGT ATG ATA CAG TA	53.5	
PFF0435w	Ornithine aminotransferase Fwd	CAA CTT TGG TCC ATT CGT ACC	57.9	165
	Ornithine aminotransferase Rev	GCT ACA CCT GGG AAA TAA CTA TC	58.9	
PFL1885c	Ca/calmodulin prot kinase2 Fwd	CGC ATT GGA AGC ATT ACA TTC TA	57.1	154
	Ca/calmodulin prot kinase2 Rev	ACA TCT CAT ATT CAT TGA TGG ACT G	58.1	
PFI1090w	AdoMet synthetase Fwd	TTT AGA TTA CAA AAC GGC AGA GAT AA	56.9	160
	AdoMet synthetase Rev	AGG CAT ATA ATT CTC AGT TTC ATC AG	58.5	
PF08_0131	1-Cys-peroxiredoxin Fwd	TAC TCC CGT TTG TAC CAC TGA	57.9	162
	1-Cys-peroxiredoxin Rev	ATA TCC CAC TTA TCT AGG TTT C	54.7	
PFD0830w	DHFR/TS Fwd	AAC CTT TAA GCA ATA GGA TAA ATG	54.2	164
	DHFR/TS Rev	TTG ATA AAC AAC GGA ACC TC	53.2	

a. Melting temperature was calculated with the formula: $T_m = 69.3 + (0.41 \times \%GC) - (650/\text{primer length})$ [218].

Fwd = forward and Rev = reverse primers.

3.3 RESULTS

3.3.1 Ensuring the correct treatment dosage for the functional genomics investigations

The IC_{50}s of DFMO (IC_{50} = 0.53 mM) and MDL73811 (IC_{50} = 0.8 µM) for 3D7 *P. falciparum* under the particular laboratory conditions were determined in section 2.3.2. In a previous suppressive-subtractive

hybridisation/microarray study of DFMO-treated *P. falciparum*, a treatment-dose of 10 mM was used [219]. In the combination with MDL73811 it was decided to use only 5 mM DFMO (i.e. 9x IC_{50}), which was previously demonstrated to inhibit ODC of the chloroquine-resistant FCR-3 strain by more than 99% [154], and 5 μM MDL73811 (i.e. 6x IC_{50}), since MDL73811 is approximately 1000 times more active than DFMO against *P. falciparum in vitro* [153]. Treatment at these high concentrations was to ensure complete parasite arrest in order to prevent parasites escaping cytostasis and causing asynchrony [154]. For the same reason, the drugs were added in the schizont stage as opposed to the ring stage. On occasion, cytostastic arrest was achieved only for a subset of the parasite population treated in the early ring stage due to the synchronisation window being between 8 and 12 h, which resulted in escape and normal progression of the more mature ring forms. However, treatment in the schizont stage ensured arrest of the whole population and the addition of DFMO in the schizont stage was previously shown to affect neither merozoite invasiveness nor ring development [88, 220].

To ensure that both drugs contributed equally to the growth arrest at the dosages applied, a parasite growth morphology study with 5 mM DFMO and 5 μM MDL73811 separately and in combination was performed. This was additionally important since the IC_{50}s could be determined at an inoculum of maximum 5 (1% parasitaemia and 5% heamatocrit) as described in section 2.3.2, but to obtain enough RNA for microarray analysis, a parasite population of at least an inoculum of ~30 (10% parasitaemia, 3% haematocrit) was required [29, 108]. Parasite morphology was carefully monitored at 6 h intervals to determine the exact sampling times (just before and during growth arrest) for the functional genomics investigations. Both drugs on their own caused growth arrest from the trophozoite stage with no visible effect during the ring stage, as reported [153, 154]. The growth arrest was clear from about 27 hpi when compared to the untreated controls, but not in the samples taken 6 h prior. However, parasites treated with the drug combination appeared smaller and slightly more delayed compared to the separate treatments (Fig. 3.2), which can be attributed to the additive effect of the two drugs.

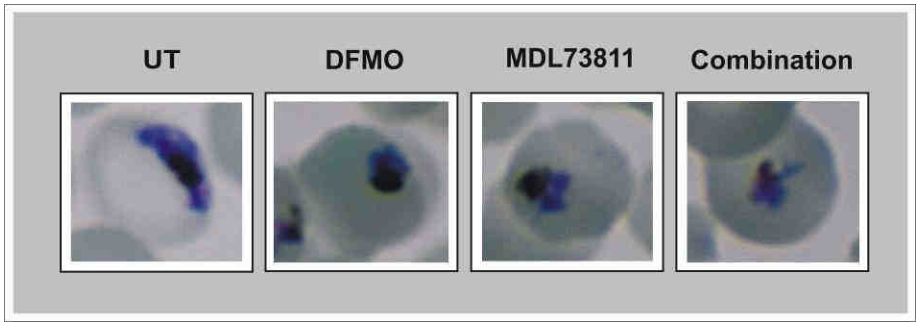


Fig. 3.2 Giemsa-stained thin smears of untreated *P. falciparum* and cultures treated with either 5 mM DFMO or 5 μM MDL73811 or the combination of 5 mM DFMO/5 μM MDL73811 result in growth arrest at ~27 hpi.

Parasite sampling times for the functional genomics investigations (after treatment with the combination of 5 mM DFMO and 5 µM MDL73811) were subsequently selected to be in the early (t_1 = 19 hpi), mid (t_2 = 27 hpi) and mature (t_3 = 34 hpi) trophozoite stages (before and during cytostasis) such that the expression period of PfAdoMetDC/ODC (12 to 40 hpi) would be spanned. The complete inhibition of AdoMetDC and ODC activities of the 3D7 parasite strain was confirmed under these treatment conditions with radio-labelled substrate (*S*-adenosyl-L-[^{14}C]methionine and L-[^{14}C]ornithine) assays of the decarboxylase activities after PfAdoMetDC/ODC co-inhibition. No decarboxylase activity above background was observed in the treated samples (T_{t1}, T_{t3}) compared to normal increasing enzyme activities of ODC and AdoMetDC in the untreated controls (UT_{t1}, UT_{t3}, Fig. 3.3), as expected from previous reports [88]. Enzyme inhibition by both DFMO and MDL73811 are specific and irreversible [147].

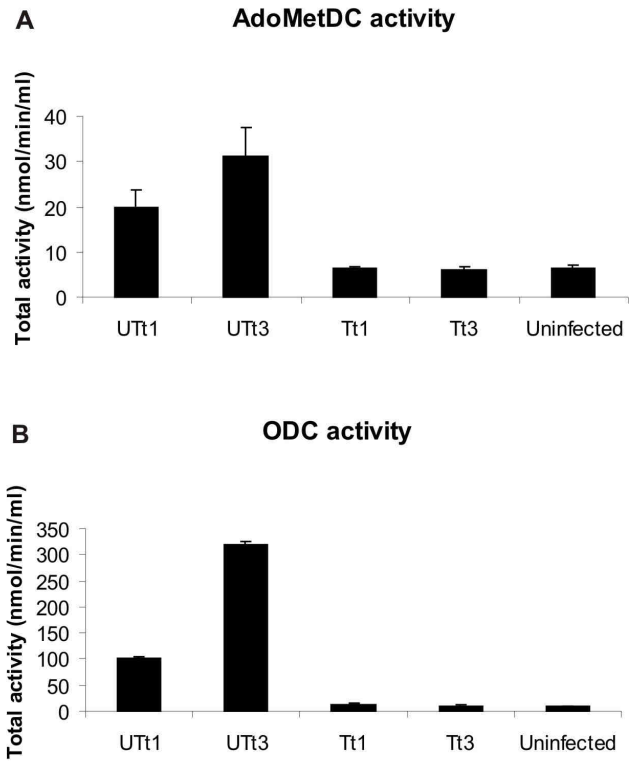


Fig. 3.3 Total activity (nmol/min/ml lysate) of **A)** AdoMetDC and **B**) ODC, based on the release of $^{14}CO_2$ from lysates of untreated and DFMO/MDL73811-treated parasites after incubation with S-adenosyl-L-[14C] methionine or L-[14C]ornithine, respectively (n=2). Cultures were sampled at t_1 = 19 hpi and t_3 = 34 hpi.

3.3.2 Transcriptomics sampling, RNA isolation and cDNA synthesis

Based on the parasite growth morphology study, DFMO/MDL73811-treatment was performed in the schizont stage at about 42 hpi (real t_0) and parasites were sampled at t_1 = 19 hpi, t_2 = 27 hpi and t_3 = 34 hpi of the next life cycle, for the isolation of RNA to be used as starting material in the subsequent microarray analysis (Fig. 3.4).

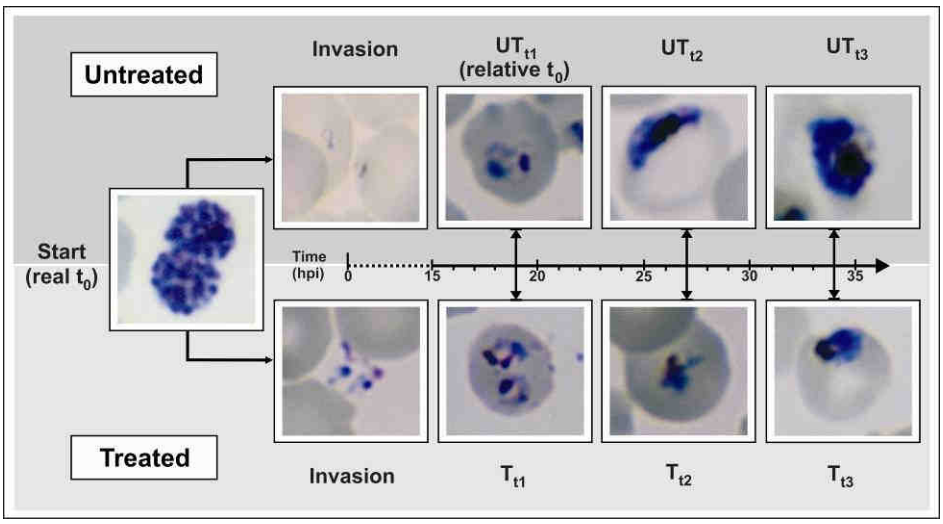


Fig. 3.4 Transcriptomics sampling times. Giemsa-stained thin smears of untreated *P. falciparum* and cultures treated with the combination of DFMO/MDL73811. Samples were taken at t_1 = 19 hpi, t_2 = 27 hpi and t_3 = 34 hpi. Growth arrest was morphologically visible from T_{t2}.

The purity and quality of the RNA were assessed with denaturing agarose/formaldehyde gel electrophoresis (Fig 3.5). The sharp definition of the 28S and 18S rRNA species, higher intensity of the 28S band and minimal smearing confirmed the integrity of the samples. Furthermore, 260/280 nm ratios between 1.9 and 2.0 were obtained, which indicated purity from protein contamination [209].

Sample	260/280	µg
t_1 = 19 hpi		
UT_A	1.945	20.4
T_A	1.973	21.4
UT_B	1.968	23.8
T_B	1.929	15.7
t_2 = 27 hpi		
UT_A	2.013	72.3
T_A	1.949	29.2
UT_B	1.982	84.1
T_B	1.953	36.2
t_3 = 34 hpi		
UT_A	1.960	88.8
T_A	1.943	32.6
UT_B	1.983	90.4
T_B	1.951	32.3

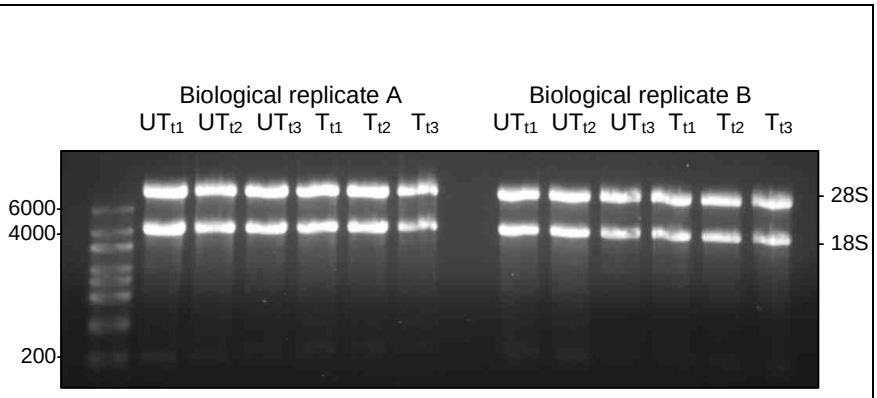


Fig. 3.5 The total RNA yield obtained from 15 ml cultures (10% parasitaemia, 3% haematocrit) over the time course is tabled and the 260/280 nm ratios are shown. RNA integrity was further assessed with denaturing agarose/formaldehyde electrophoresis where 2 µg samples were loaded onto a 1.2% agarose/formaldehyde denaturing gel with high range RNA ladder (Fermentas Life Sciences). 28S and 18S rRNA are visible, but the 5S rRNA was removed during isolation.

Note that for mRNA enrichment purposes all RNA species smaller than 200 nucleotides (e.g. 5S rRNA and tRNA) are selectively excluded by the RNeasy protocol (RNeasy manual). Cytostatic growth arrest was observed in the RNA yields obtained since the level was approximately maintained in the treated samples from t_1 to t_3 (20 - 30 µg), whereas it increased in the untreated samples (70 - 90 µg), as expected during normal development (Fig. 3.5) [221]. The cDNA synthesis protocol [210] was optimised by increasing the amount of reverse transcriptase from 150 U to 480 U and the reaction time from 120 min to overnight. In this manner the

cDNA yield from 12 µg of total RNA was increased about 3-fold (~2 µg cDNA to ~6 µg cDNA). Only 2 µg cDNA is required for sufficient labelling and hybridisation and therefore several technical replicate hybridisations could be performed from the same RNA sample.

3.3.3 Oligonucleotide microarray analysis

Transcriptional profiling of untreated versus PfAdoMetDC/ODC co-inhibited parasites was performed on non-commercial arrays spotted with 8088 70-mer oligonucleotides, which represented 5332 unique genes [210] (Supplementary CD). Note that the microarray analysis quantitates steady-state mRNA, which for any gene is a function of both the rate of transcription and the rate of mRNA decay [128], and is therefore referred to as transcript abundance here. A reference design microarray experiment was performed, i.e. the cDNA of the various samples were co-hybridised with a composite reference pool from all the samples included in the experiment. Treated and untreated samples (two biological replicates) were taken at three time points (12 samples) and two technical replicate hybridisations were performed for each. Thus a total of 24 hybridisations were performed and 20 were of acceptable quality (e.g. total number of detectable spots, limited background etc.) to be used for data analysis. Cytostasis caused arrays of treated samples to have an overall yellow appearance (similar to self-self hybridisations) compared to untreated samples at t_2 and t_3, which had a more colourful (red, yellow, green) display typical of differential expression (Fig. 3.6). The yellow appearance was due to enrichment of the reference pool with early transcripts resuting from the parasite arrest.

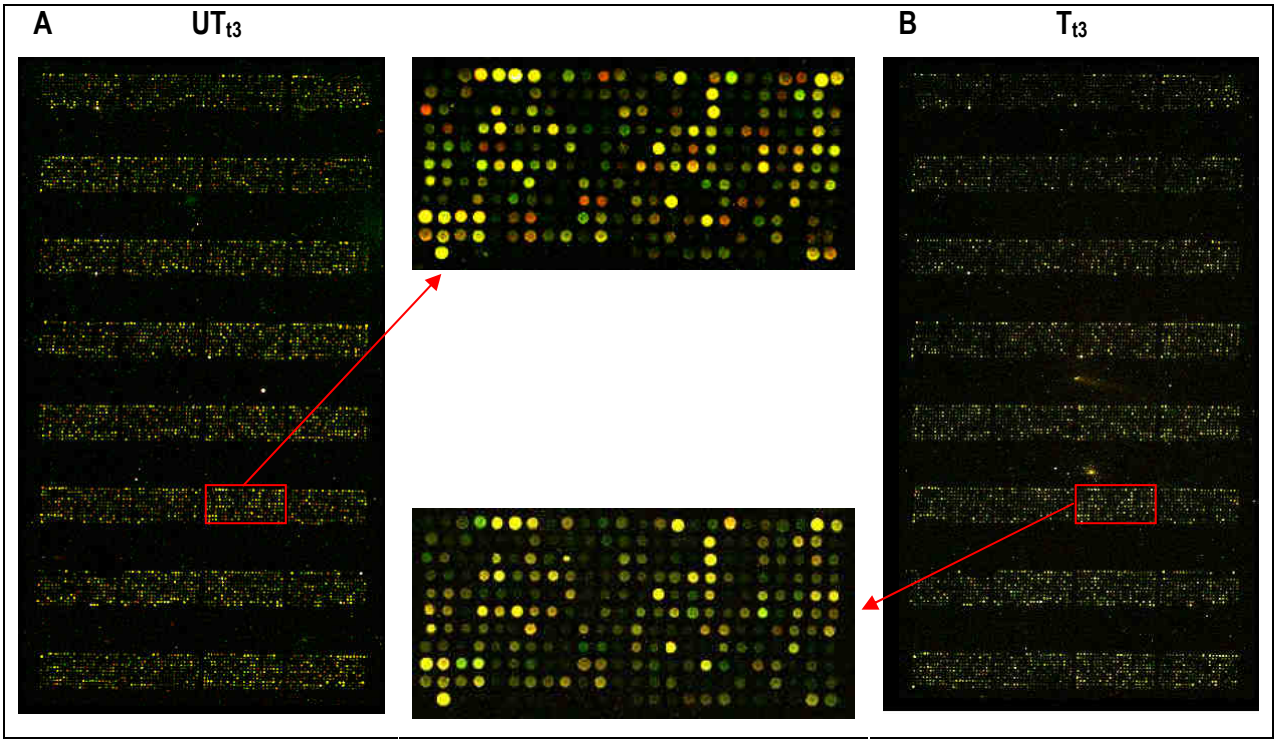


Fig. 3.6 Typical 70-mer oligonucleotide spotted arrays of **A)** UT$_{t3}$ and **B)** T$_{t3}$ with one block enlarged. The slide from T$_{t3}$ had and overall yellowish appearance (similar to self-self hybridisation) due to enrichment of the reference pool with early transcripts from arrested parasites, whereas the slide from UT$_{t3}$ was colourful as is typical of differential expression on microarray.

3.3.4 Exploratory data analysis

3.3.4.1 Hierarchical clustering of data with those of other perturbations

Exploratory data analysis was mainly performed in CLUSTER 2.1.1 with hierarchical clustering of normalised log_2-ratios, which was then displayed in TREEVIEW 1.0.12. The PfAdoMetDC/ODC co-inhibition data were clustered with microarray data of *P. falciparum* exposed to a series of perturbations, including several antimalarial drugs and environmental stressors (Fig. 3.7, data used by courtesy of M. Llinás, unpublished data) sampled 1, 2, 4, 6, 8, 12, 16 and 24 h post-treatment (treatment occurred at invasion). The PfAdoMetDC/ODC co-inhibition data are at the far right with the two biological replicate sets (A, B) displayed separately. The perturbation controls for the Llinás perturbation data are at the far left. Because of the different periods of sampling of the PfAdoMetDC/ODC co-inhibition data (t_1 = 19 hpi, t_2 = 27 hpi and t_3 = 34 hpi) and the Llinás perturbation data (1 – 24 hpi), clustering between genes (Fig. 3.7) was not informative, but when clustering between arrays (Fig. 3.8) was performed, UT_{t1} and T_{t1} as well as T_{t2} and T_{t3} respectively, clustered together and were relatively close on the hierarchical tree. UT_{t2} and UT_{t3} also clustered together but completely separately from the rest of the data along with the latest control (24 h) of the Llinás perturbation data. Biological replicates always clustered together. These results indicated the reproducibility of the PfAdoMetDC/ODC co-inhibition data in terms of replicates, but more importantly it confirmed that the data obtained were comparable with other *P. falciparum* datasets (based on the clustering together of the untreated controls) e.g. the 3D7 IDC transcriptome [91]. Furthermore, the relative closeness of the UT_{t1} with T_{t1}, T_{t2} and T_{t3} again illustrated the effect of cytostasis, whereas UT_{t2} and UT_{t3} clustered on their own branch at the bottom of the tree due to normal parasite progression. The data shown in Fig. 3.8 are those for the transcripts of PfAdoMetDC/ODC in all the different perturbations.

Hierarchical clustering was performed with both the complete set of PfAdoMetDC/ODC data (treated and untreated) and the treated parasite data with UT_{t1} only (Supplementary CD and website at http://genomics-pubs.princeton.edu/PfAdoMetDC_ODC). The latter comparison revealed two large arrest clusters that were maintained at either high (red) or low (green) transcript abundance with a correlation coefficient of 0.94 for both clusters. Clusters of transcripts with differential abundance over the time course were also revealed, including one that contained several polyamine pathway-related transcripts (Fig. 3.9).

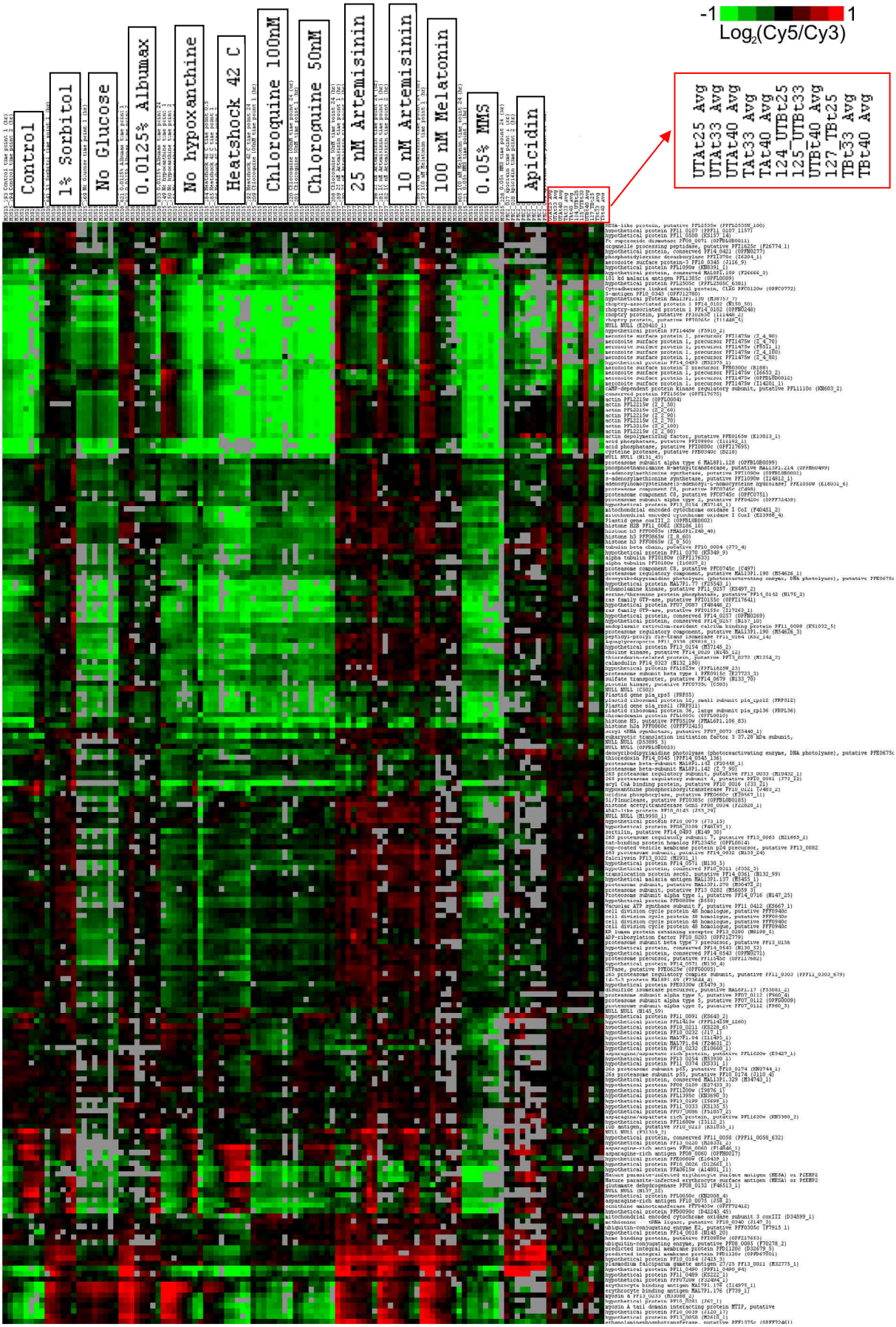


Fig. 3.7 Hierarchical data clustering between genes of the PfAdoMetDC/ODC co-inhibition data (t_1 = 19 hpi, t_2 = 27 hpi and t_3 = 34 hpi) and the Llinás perturbation data (1 to 24 hpi) was not informative due to the different sampling periods of the two sets (M. Llinás, unpublished data). MMS = methyl methane sulfonate.

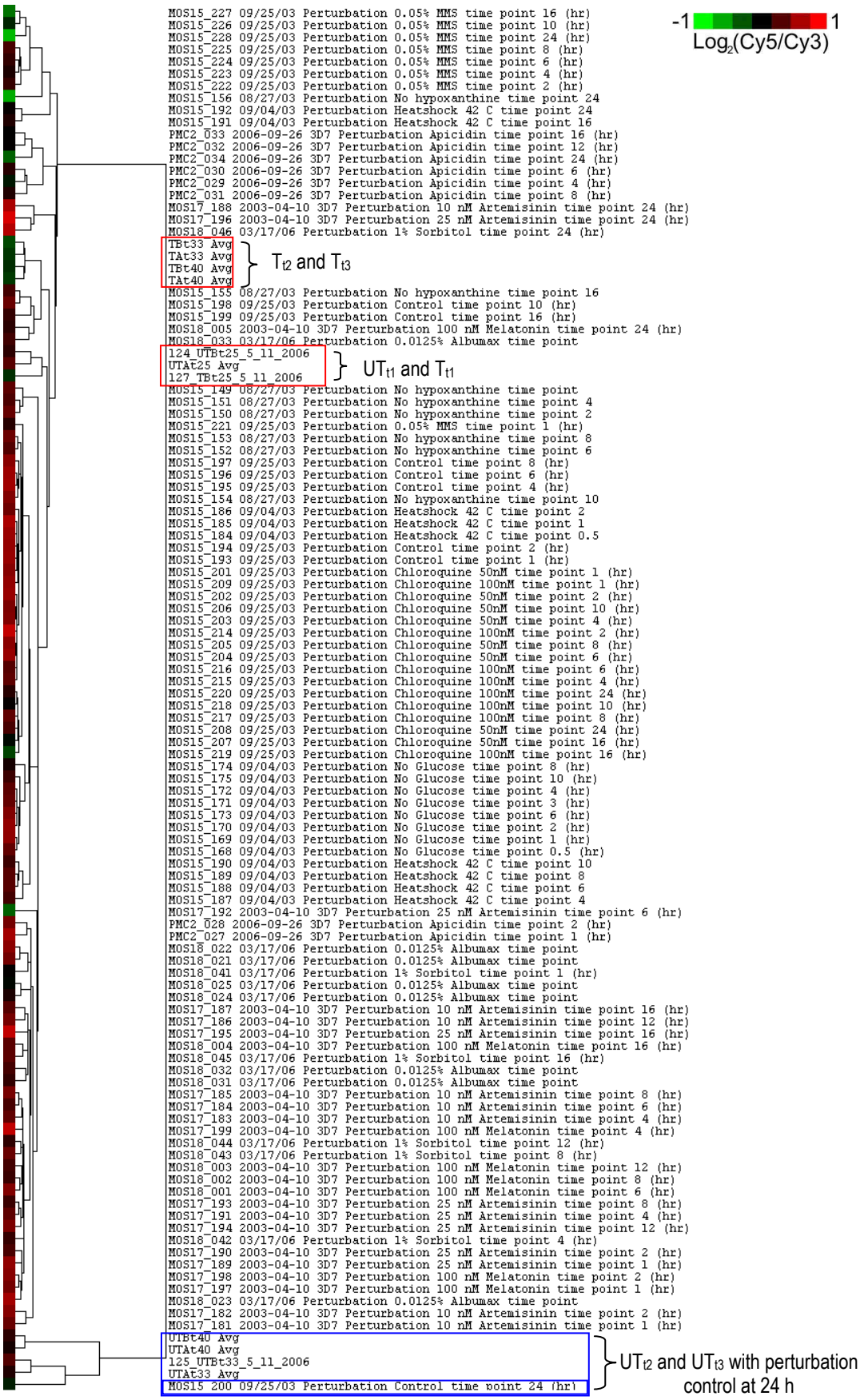


Fig. 3.8 Hierarchical data clustering between arrays of the PfAdoMetDC/ODC co-inhibition data and the Llinás perturbation data indicate the effect of cytostasis with the clusters of UT_{t1} and T_{t1} as well as T_{t2} and T_{t3} being relatively close, whereas UT_{t2} and UT_{t3} clustered completely separately, albeit with the latest control of the Llinás perturbation data (M. Llinás, unpublished data). The data shown are those for the transcript of PfAdoMetDC/ODC. The sampling times indicated are post-treatment.

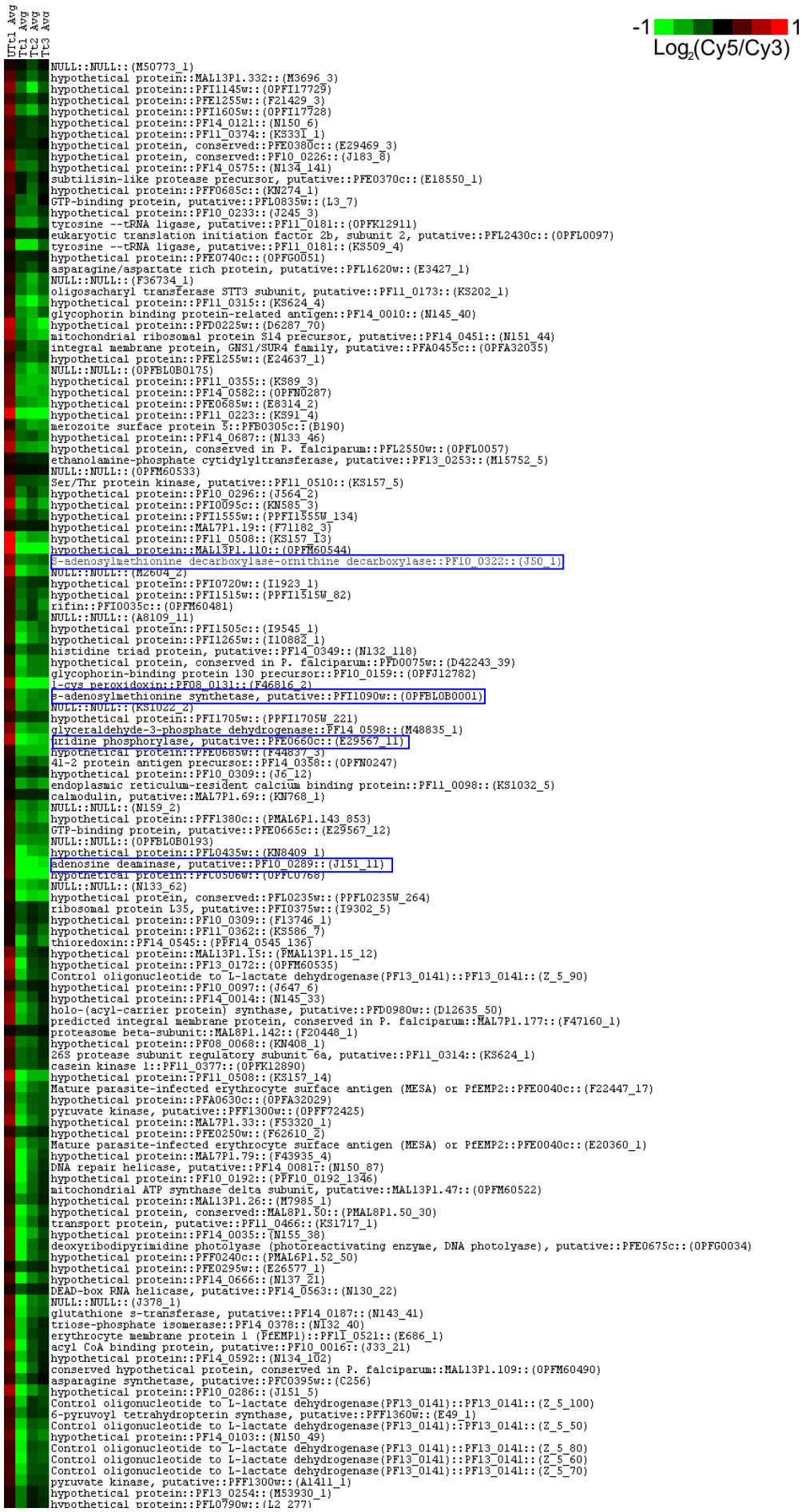


Fig. 3.9 A tight cluster (correlation coefficient = 0.89) containing several polyamine pathway transcripts (framed), including PfAdoMetDC/ODC, produced by hierarchical data clustering of UT_{t1} with T_{t1}, T_{t2} and T_{t3}.

3.3.4.2 Hierarchical clustering of data related to polyamine and methionine metabolism

Exploratory hierarchical clustering of the transcripts from polyamine and methionine metabolism on their own resulted in two relatively tight clusters (correlation values of 0.78 and 0.94, respectively, Fig. 3.10) indicating that gene expression of this pathway does not occur at the same time for all transcripts, i.e. it is not all in phase. The time points chosen combined with the transcriptional arrest appear to reveal a biphasic nature of polyamine and methionine metabolism gene expression, but a more sophisticated bioinformatics approach and inclusion of all the genes involved (e.g. all methyltransferases etc.) may indicate more phases. Gene expression for RNA and DNA synthesis is normally in phase in *P. falciparum* [222], but transcription of the genes of the glutathione and thioredoxin systems [222], and the pentose phosphate pathway [223], were also demonstrated not to be all in phase with at least two expression peaks in the IDC.

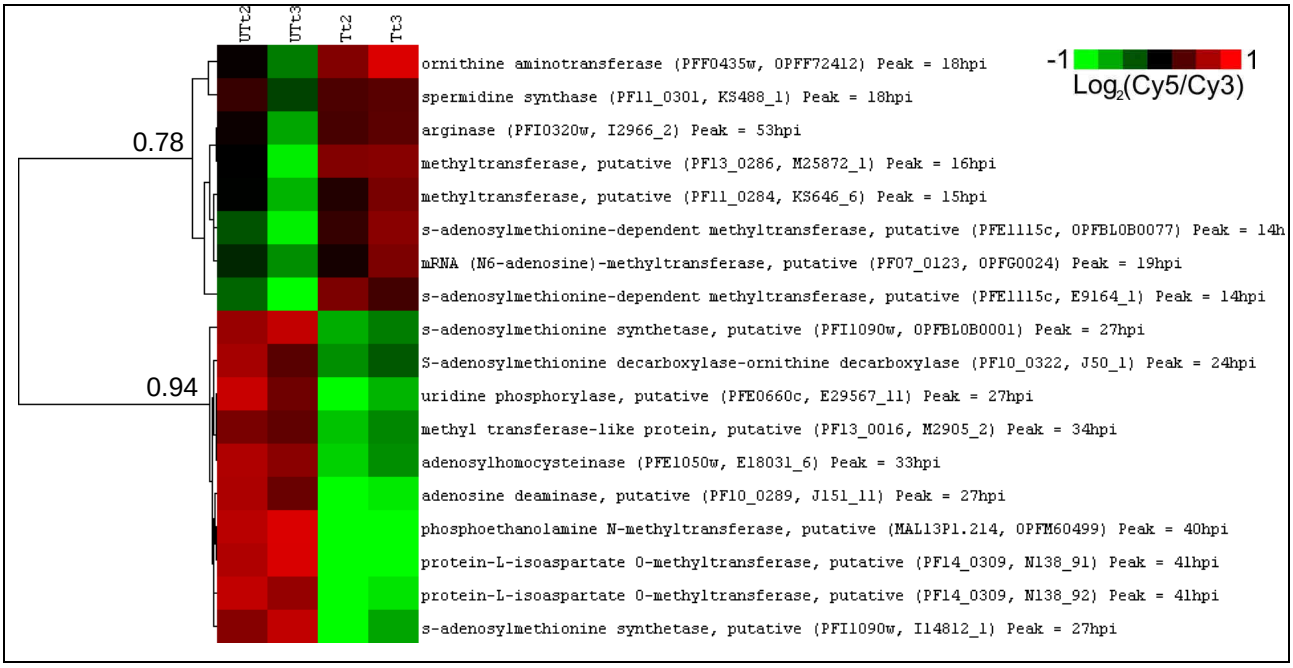


Fig 3.10 "Biphasic" segregation of the expression/peak abundance of polyamine and methionine metabolism transcripts visible as two relatively tight clusters due to the transcriptional arrest of treated samples (T_{t2}, T_{t3}) and normal progression of untreated controls (UT_{t2}, UT_{t3}). This division was particularly clear at t_2 and t_3, as demonstrated.

3.3.4.3 Phase-ordering and correlation calculations

Fast Fourier analysis was applied to calculate the apparent phase and frequency of gene expression in the IDC transcriptome [29, 91]. By ordering the PfAdoMetDC/ODC co-inhibition data according to the phase of expression, as determined for the 3D7 strain (Fig. 1.8) [91], cytostasis was revealed as a generalised transcriptional arrest across T_{t1} to T_{t3} with normal IDC progression visible in UT_{t1} to UT_{t3} (Fig. 3.11A). In general, expression of genes that were already transcribed before the effect of the treatment (IDC peak expression before or around t_1) were unchanged, whereas genes that were not yet transcribed (IDC peak expression in second half of 48 h cycle e.g. t_2 and t_3) showed no or low transcript abundance. Not surprisingly,

growth arrest occurred after the transcriptional arrest and was morphologically visible only from T_{t2} with parasites appearing small and distressed compared to their untreated counterparts (Fig. 3.4)

In order to use the IDC transcriptome as a reference time line (i.e. of transcript expression times), Pearson correlation was used to align the PfAdoMetDC/ODC co-inhibition data with each of the one hourly time points of the 3D7 IDC transcriptome [91]. UT_{t1} (19 hpi) had the highest correlation at 14 - 15 hpi with T_{t1} to T_{t3} following closely thereafter at 15 - 16 hpi (Fig. 3.11B) as opposed to UT_{t2} and UT_{t3}, which correlates at 33 hpi and 44 hpi of the 53-hour 3D7 IDC cycle [91], respectively. This analysis corroborated the transcriptional arrest and estimated the approximate time thereof as 15 - 16 hpi (according to the 3D7 IDC transcriptome), which correlates with the start of PfAdoMetDC/ODC expression (transcript produced from 12 – 40 hpi).

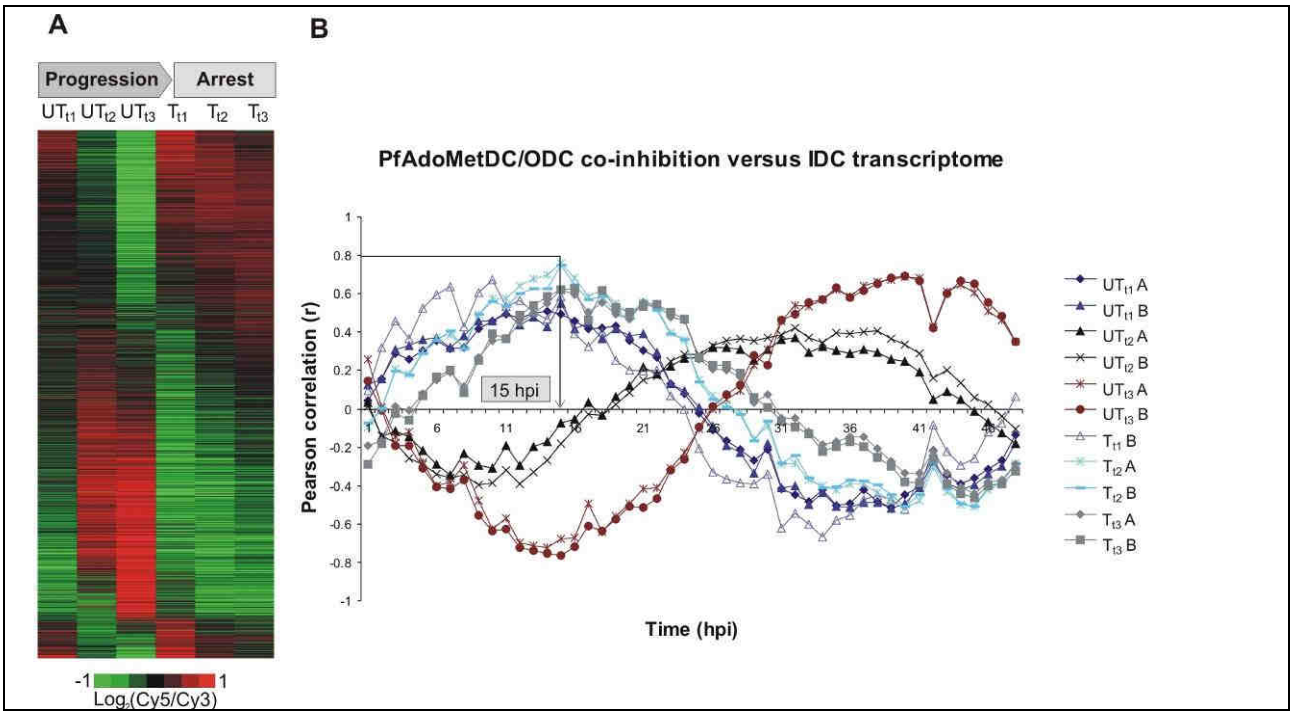


Fig. 3.11 A) A phaseogram depicting the transcriptional profiles of untreated versus PfAdoMetDC/ODC co-inhibition data over three time points (t_1 to t_3) by ordering 3206 oligonucleotides according to the phase of expression. Transcriptional arrest is visible in T_{t1} to T_{t3}. **B)** Pearson correlation between the PfAdoMetDC/ODC co-inhibition data and the one hourly time points of the 3D7 IDC transcriptome. The treated samples have a correlation profile similar to UT_{t1} (relative t_0) with the highest correlation at ~15-16 hpi, illustrating the approximate time of transcriptional arrest. Data of the respective biological replicates (A, B) are shown separately (T_{t1} had only one biological replicate due to technical difficulty).

Pearson correlation within the PfAdoMetDC/ODC co-inhibition data also indicated transcriptional arrest, since relatively close correlations were observed between the treated parasite data (T_{t1} versus T_{t2}: r = 0.77, T_{t2} versus T_{t3}: r = 0.89, T_{t1} versus T_{t3}: r = 0.61) and each of these compared to UT_{t1} (r ranging from 0.57 to 0.72). In contrast, data at t_2 and t_3 were uncorrelated (r = 0.07) and anti-correlated (r = -0.61) respectively compared to the matched untreated parasite data (Table 3.2). This is due to the transcriptional arrest of treated and normal progression of untreated parasites. In comparison, in the absence of transcriptional arrest, the array data of doxycycline-treated versus untreated *P. falciparum* were still highly correlated (r = 0.8) after 55 h [62].

Table 3.2 Pearson correlation within the PfAdoMetDC/ODC co-inhibition transcript data

Comparison	Pearson correlation (r)
T_{t1}:UT_{t1}	0.72
T_{t2}:UT_{t2}	0.07
T_{t3}:UT_{t3}	-0.61
T_{t2}:UT_{t1}	0.72
T_{t3}:UT_{t1}	0.57
T_{t1}:UT_{t2}	0.77
T_{t2}:UT_{t3}	0.89
T_{t1}:UT_{t3}	0.61

Thus, three different analyses (phase ordering, Pearson correlation with the 3D7 IDC transcriptome [91] and within the PfAdoMetDC/ODC co-inhibition data) confirmed the transcriptional arrest observed as a result of the perturbation. Data reproducibility was calculated and technical replicates had a correlation of 0.93 and biological replicates a correlation of 0.88 on average. Reproducibility of the biological replicate can be visualised in the comparison with the 3D7 IDC transcriptome where the replicate sinusoidal curves obtained can almost be superimposed (Fig. 3.11B).

3.3.5 Differential transcript abundance analysis

Comparison of the clustered PfAdoMetDC/ODC co-inhibition data with the IDC transcriptome indicated a few genes with the same expression time but with an opposite response to the perturbation, e.g. the increased abundance of the transcript for ornithine aminotransferase (OAT, PFF0435w) and the decreased abundance of that for enolase (PF10_0155), though both have normal peak expression at 18 hpi. Hence the transcriptional arrest was not merely due to the normal mRNA abundance level of IDC co-expressed genes being maintained, but transcripts of specific genes with the same expression time were differentially affected. To identify the transcripts most affected by the perturbation, a quantitative approach was followed with LIMMA as well as EDGE analysis. However, the global transcriptional arrest caused by PfAdoMetDC/ODC co-inhibition negated the direct comparison of treated and untreated data at t_2 and t_3, as this standard approach would have detected mainly growth/stage differences. Therefore, in differential transcript abundance analysis all treated time points were compared to UT_{t1} defined as a relative t_0. Comparison to the real t_0 would also primarily have identified stage differences, since drug treatment was performed in the late schizont stage of the life cycle, but cytostasis and sampling occurred in the early trophozoite stage of the next cycle.

3.3.5.1 Data normalisation

Microarray data quality and normalisation methods were evaluated using data diagnostic tools from the MARRAY software package in R [213] on GenePix data (R-script on Supplementary CD). The red (Cy5) and green (Cy3) background of every array was inspected for global or localised effects and presence in one or both channels (Fig. 3.12). Local artefacts and edge effects (visible at the top in the red background image, Fig 3.12) indicated background subtraction and that a spatial normalisation approach such as print-tip LOWESS should be used as opposed to global normalisation. Background subtraction with an offset of 50 was used to

avoid negative or zero-corrected intensities [206]. However, when print-tip LOWESS was applied the red/green density plots (shown in Fig. 3.16) indicated a serious error, appearing as data degeneracy of one or more arrays. This was subsequently solved when the prerequisite of print-tip LOWESS was considered, namely that it requires data from at least 150 spots [206]. The spotter used has 32 print-tips, which each spotted 264 spots (11 rows of 24 spots each), but in a few blocks only ~120 spots had a signal because of low intensity transcripts. Thus, print-tip LOWESS was not justified, in which case global LOWESS or robust spline normalisation is recommended [206]. To avoid the global approach as explained above, robust spline normalisation, which is a compromise between print-tip and global LOWESS [206], was applied within arrays to remove systematic error such as dye bias.

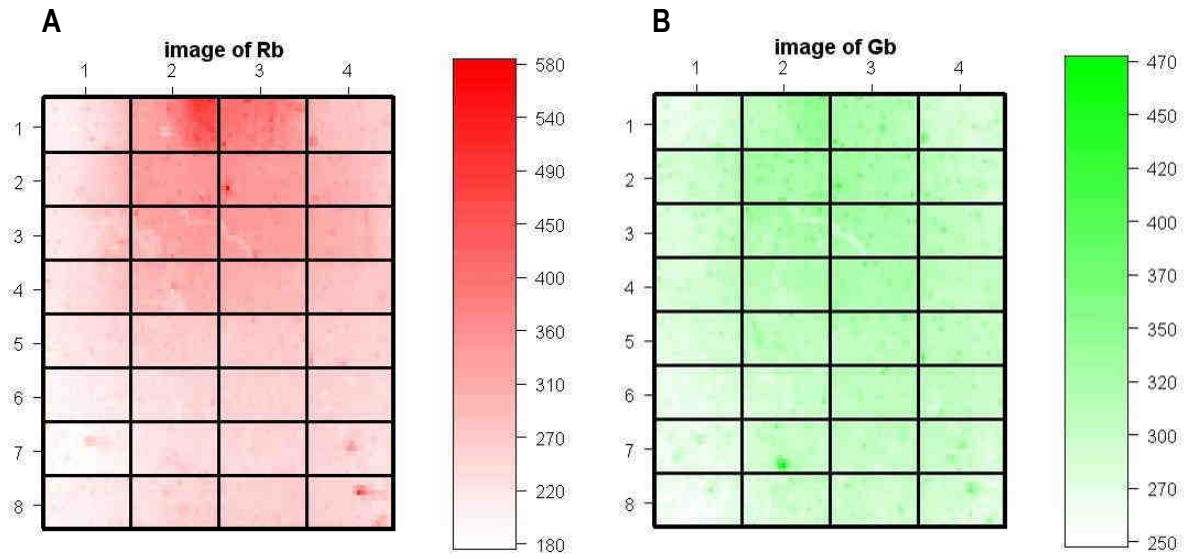


Fig 3.12 A) Red and **B)** green background images of a typical array (111_TAt33) presented in MARRAY. Note the localised artefacts in both images and edge effect towards the top of the array in the red background image.

Gquantile normalisation was applied between arrays to ensure the same distribution of spot intensities in the green (Cy3) channel across all the arrays (adjusting the red or Cy5 channel accordingly) without changing the $\log_2$-ratios (M-values). Gquantile between-slide normalisation is specifically applicable to common reference microarray experiments where the reference is labelled with Cy3, whereas Rquantile is applied in experiments where the samples are labelled with Cy3 and the reference with Cy5.

Log_2-ratios (M) versus average intensities (A) plots were used to confirm that the applied normalisation had the desired effect. On these plots, the centre of distribution of $\log_2$-ratios should ideally be zero, the ratios should be independent of spot intensity and the fitted line should be parallel to the intensity axis [205]. After performing normalisation, the data spread of the example complied better with the ideal (Fig. 3.13).

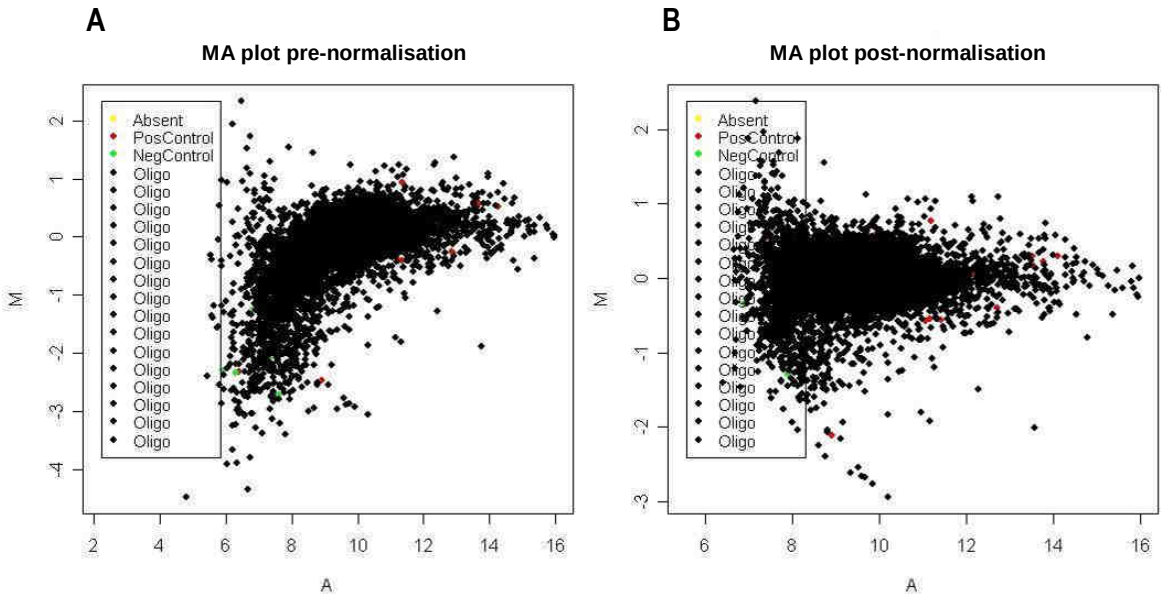


Fig 3.13 MA-plots from a typical array (125_UTBt33) **A)** before and **B)** after data transformation (background subtraction, robust spline and Gquantile normalisation). Note that the post-normalisation data comply better with the ideal, i.e. the centre of distribution of the M-values is around zero and should a line be fitted through the data, it will be parallel to the A-axis.

Robust spline normalisation in addition had a smoothing effect on the distribution of the log-ratios among the print-tip groups (Fig. 3.14), which reduced technical bias as a result of, for example, the unequal deposition of oligonucleotides or different size spots resulting from the different print-tips.

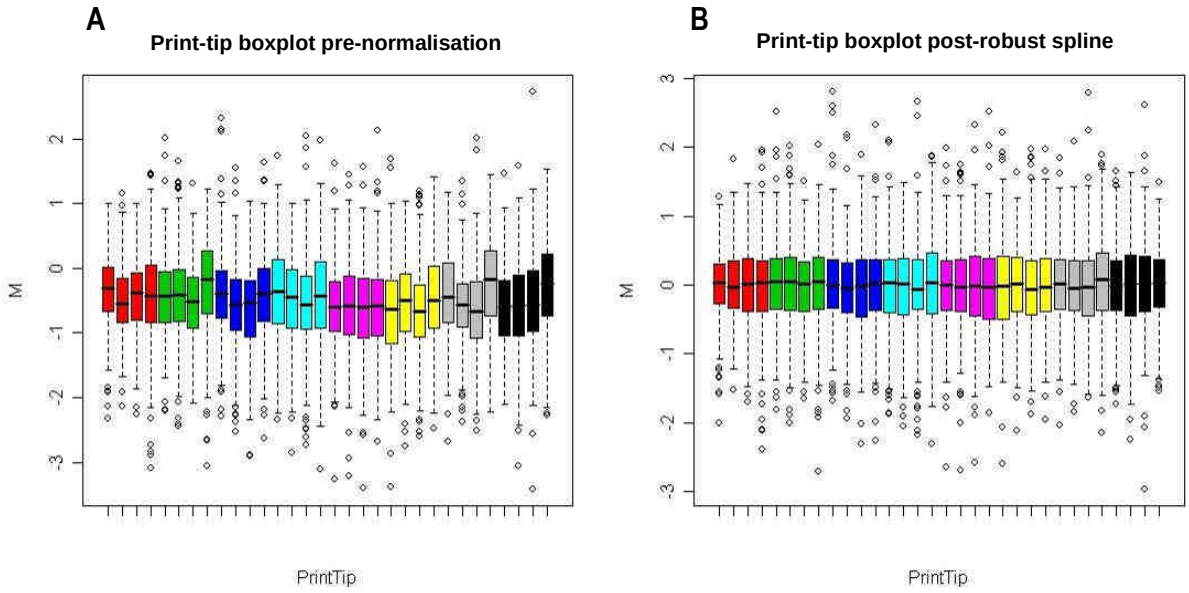


Fig 3.14 Print-tip boxplots of the log_2-ratio distributions of the 32 blocks (equivalent to the 32 print-tips) from a typical array (116_TBt25), **A)** before and **B)** after data transformation (background subtraction, robust spline normalisation).

Log_2-ratio boxplots and intensity boxplots across all arrays (Fig. 3.15) illustrated that the log_2-ratios remained unchanged although intensities across the different arrays were equalised after normalisation.

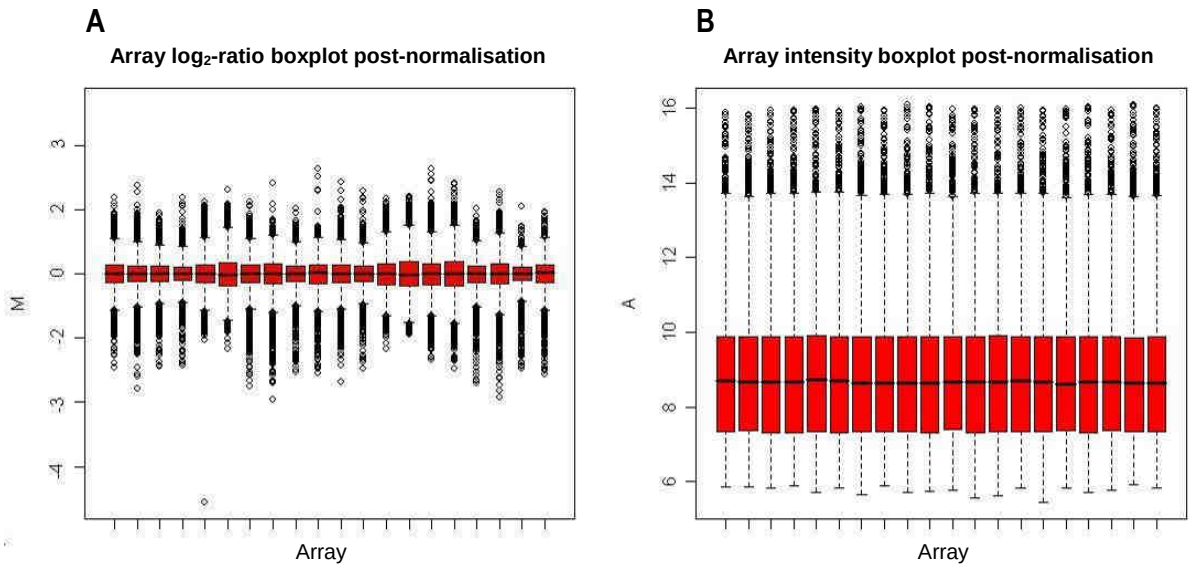


Fig 3.15 A) Boxplots of log_2-ratios (M) and **B)** intensities (A) across all arrays post-normalisation (background subtraction, robust spline and Gquantile normalisation). Although the intensities across the different arrays were equalised, the log_2-ratios are independent and did not change.

Red/green (Cy5/Cy3) density plots indicated a dye bias towards the green channel for several arrays prior to normalisation but this was improved revealing only one density curve for the green channel subsequent to Gquantile normalisation (Fig. 3.16). Thus, all the arrays were normalised according to the green (Cy3, reference pool) channel, enabling subsequent comparison of the log_2-ratios (Cy5/Cy3).

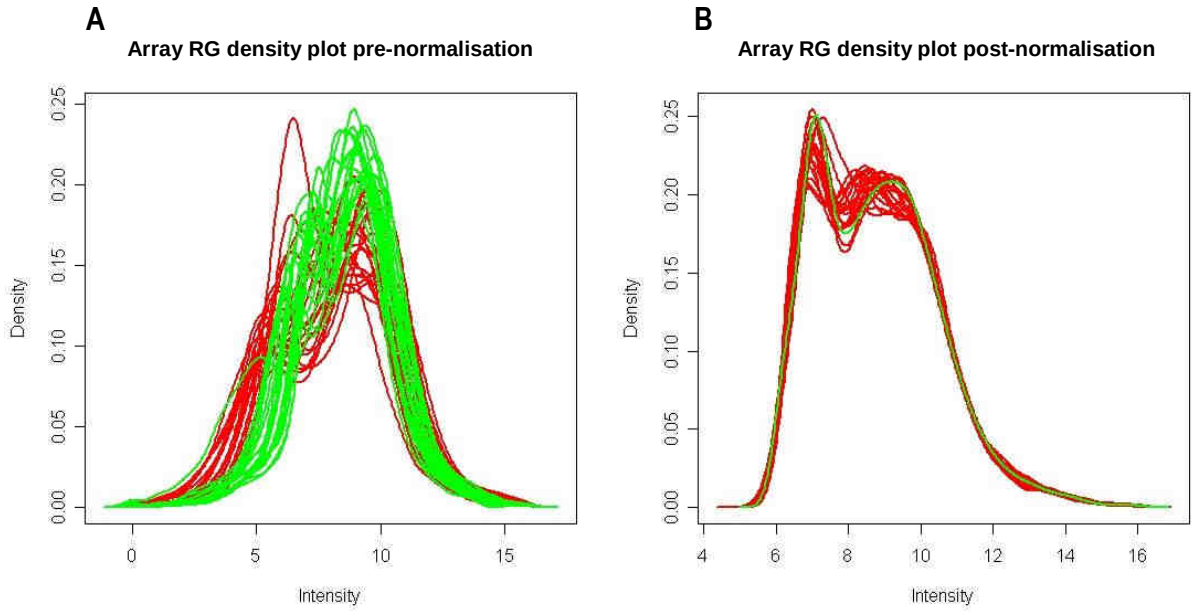


Fig 3.16 Red/Green density plots of all the arrays **A)** before and **B)** after data transformation (background subtraction, robust spline, Gquantile normalisation). Note that after Gquantile normalisation only one curve is visible for the green channel in all the arrays.

3.3.5.2 LIMMA data analysis

After obtaining a satisfactory intensity distribution within and across the arrays according to the diagnostic plots, differential transcript abundance between the samples was calculated for all the oligonucleotides in

LIMMA. Over the time course, the abundance of 538 transcripts from 5332 unique genes represented on the array, was significantly differentially affected (1.7-fold change in either direction; $\log_2$-ratio ≥ 0.75 or ≤ -0.75; $p<0.05$) compared to the relative t_0 (Appendix A). Of these, 171 transcripts were increased (up to 3.2-fold) and 377 were decreased (down to 5-fold), with the transcripts of 10 apparently unrelated genes falling in both categories by displaying an increase in t_1 as well as a decrease in t_2 and/or t_3. A selected subset of the differentially affected transcripts is presented in Table 3.3, including eight transcripts from polyamine and methionine metabolism and three methyltransferases. Note that several of the transcripts (marked with *) with decreased abundance in the table also clustered together during the exploratory analysis (Fig. 3.9) e.g. PF10_0322, PFI1090w, PFE10_0289, PFE0660c, PF14_0598, PF08_0131, P13_0141, PFF1300w, PF14_0018, PFE0675c, PF14_0187, PF14_0378. The complete LIMMA dataset of 538 transcripts is presented in Appendix A.

Table 3.3 Biological functions of a subset of the transcripts differentially affected according to LIMMA as a result of PfAdoMetDC/ODC co-inhibition

PlasmoDB ID	Annotation	Fold change[a]	IDC time of peak expression (hpi)[b]
Polyamine and methionine metabolism			
PF10_0322*	PfAdoMetDC/ODC	-1.9	24
PFD0285c	Lysine decarboxylase, putative	2.8	25
PFF0435w	Ornithine aminotransferase	1.9	18
PFI1090w*	AdoMet synthetase	-2.4	27
PFE1050w	Adenosylhomocysteinase	-1.9	33
PF10_0289*	Adenosine deaminase, putative	-2.6	27
PFE0660c*	Uridine phosphorylase, putative	-3.2	27
PF10_0340	Methionine-tRNA ligase	-1.7	33
Methyltransferases			
MAL13P1.214	Phosphoethanolamine N-methyltransferase, putative	-2.8	40
PF14_0309	Protein-L-isoaspartate O-methyl transferase, putative	-1.9	41
PF14_0526	Generic methyltransferase	-2.9	37
Potential polyamine associated effects			
PF14_0316	DNA topoisomerase II	-1.7	50
PFL1885c	Calcium/calmodulin-dependent protein kinase 2, putative	2.4	50
PF07_0065	Zinc transporter, putative	-2.4	40
Oxidative stress defence			
PF08_0131*	1-Cys-peroxiredoxin	-3.2	34
PF14_0192	Glutathione reductase	-1.7	34
PF14_0187*	Glutathione S-transferase	-1.8	-
Energy metabolism			
Oxidative phosphorylation			
CoI	Cytochrome oxidase I, putative	-2.0	-
CoxI	Cytochrome oxidase I, putative	-1.7	53
CoxIII_2	Mitochondrial encoded cytochrome oxidase subunit 3	-2.0	-
PF11_0412	Vacuolar ATP synthase subunit F, putative	-1.8	34
MAL7P1.75	Mitochondrial ATP synthase F1, epsilon subunit, putative	-1.8	-
PFE0970w	Cytochrome c oxidase assembly, putative	-1.7	24
PF13_0121	Dihydrolipoamide succinyltransferase	-1.7	27
Glycolysis			
PF10_0155	Enolase	-2.0	18
PF13_0141*	L-lactate dehydrogenase	-1.8	18
PF14_0378*	Triose-phosphate isomerise	-1.8	18
PF14_0598*	Glyceraldehyde-3-phosphate dehydrogenase	-2.1	27
PFF1300w*	Pyruvate kinase	-1.9	28
DNA replication			

PF11_0117	Replication factor C subunit 5, putative	-2.1	34
PF11_0087	Rad51 homologue, putative	-2.0	35
PF13_0095	DNA replication licensing factor mcm4-related	-2.2	42
PF13_0291	Replication licensing factor, putative	-1.8	34
PF14_0081*	DNA repair helicase, putative	-1.7	-
PF14_0112	POM1, putative	-2.0	37
PF14_0254	DNA mismatch repair protein Msh2p, putative	-1.8	32
PF14_0314	Chromatin assembly factor 1 p55 subunit, putative	2.4	-
PF14_0601	Replication factor C3	-2.0	34
PFB0180w	5'-3' Exonuclease, putative	-1.9	26
PFB0895c	Replication factor C subunit 1, putative	-2.0	33
PFD0470c	Replication factor A protein, putative	-2.5	34
PFD0685c	Chromosome associated protein, putative	-2.0	40
PFD0830w	Dihydrofolate reductase-thymidylate synthase	-2.0	33
PFD0950w	Ran binding protein 1	2.0	-
PFE0675c*	DNA photolyase	-2.2	36
PFF1470c	DNA polymerase epsilon, catalytic subunit A, putative	-1.7	36
PFI0235w	Replication factor A-related protein, putative	-1.8	33
PFI0530c	DNA primase, large subunit, putative	-1.8	35
Transcription factors			
PF11_0241	Hypothetical protein with Myb-like domains	1.8	-
PFL0465c	C2H2-type zinc-finger transcription factor, krox1	2.0	-
PFE1245w	CCCH-type zinc-finger protein	1.7	26
PFD0560w	Hypothetical protein with a TATA-box -like domain	1.7	32
PFE0415w	Transcription factor IIB, putative	-1.8	-
Translation			
MAL13P1.327	Ribosomal protein S17 homologue, putative	1.7	22
PF07_0080	40S ribosomal protein S10, putative	1.9	16
PF10_0038	Ribosomal protein S20e, putative	2.3	15
PF11_0454	Ribosomal protein, 40S subunit, putative	2.0	-
PF13_0014	40S ribosomal protein S7 homologue, putative	1.8	15
PF13_0171	60S ribosomal protein L23, putative	2.0	13
PF13_0228	40S ribosomal subunit protein S6	1.8	13
PF14_0205	Ribosomal protein S25	2.4	23
PF14_0231	Ribosomal protein L7a, putative	1.8	21
PF14_0579	Ribosomal protein L27, putative	2.1	21
PF14_0709	Ribosomal protein L20, putative	-2.6	32
PFB0455w	Ribosomal L37ae protein, putative	1.7	16
PFC0535w	60S ribosomal protein L26, putative	2.0	-
PFC1020c	40S ribosomal protein S3A, putative	1.7	15
PFE0185c	60S ribosomal subunit protein L31, putative	1.8	14
PFI1585c	30S ribosomal protein S6-like protein, putative	1.8	-
Cell cycle mediators			
PF13_0328	Proliferating cell nuclear antigen	-3.3	40
PF14_0604	Hypothetical protein with cyclin homology	-1.7	2
PFL1330c	Hypothetical protein with cyclin homology	1.8	37

a. Average fold change calculated at the time point of maximum change

b. Transcript peak expression time according to the 3D7 IDC transcriptome [91]

***** Transcripts that clustered together after hierarchical clustering (Fig. 3.9)

(-) No data available

The majority (70%) of the differentially affected transcripts were decreased, equally so for polyamine and methionine metabolism with the abundance of only two transcripts, OAT and LDC (PFD0285c), being increased ~2-fold and 2.8-fold, respectively (Fig 3.17). The transcript level for PfAdoMetDC/ODC, the protein of which was targeted by DFMO and MDL73811, was decreased by ~2-fold (Fig. 3.17). Thus, despite the transcripts for LDC and PfAdoMetDC/ODC being expressed at approximately the same time in the IDC (25 hpi and 24 hpi respectively), the transcript for LDC was increased and that of PfAdoMetDC/ODC was decreased,

which illustrates the differential effect of the co-inhibition on the abundance of specific transcripts, despite the generalised transcriptional arrest. Other affected transcripts from polyamine and methionine metabolism include AdoMet synthetase (PFI1090w), adenosylhomocysteinase (PFE1050w), a putative adenosine deaminase (PF10_0289), uridine phosphorylase (PFE0660c) and methionine-tRNA ligase (PF10_0340). The transcripts for putative methionyl-tRNA formyltransferase (MAL13P1.67) and pyroline carboxylate reductase (MAL13P1.284) could not be detected.

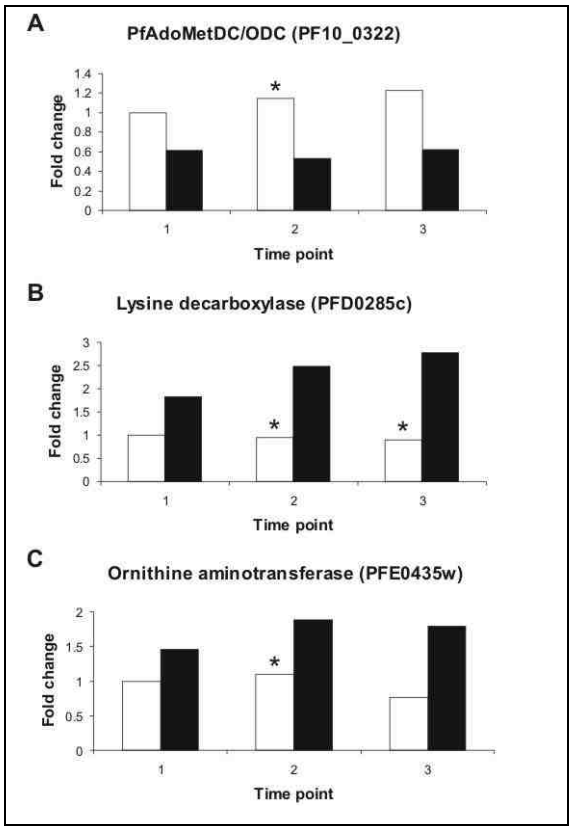


Fig. 3.17 Transcript profiles of **A)** PfAdoMetDC/ODC, **B)** lysine decarboxylase and **C)** ornithine aminotransferase compiled from the PfAdoMetDC/ODC co-inhibited (black bar) and untreated (white bar) *P. falciparum* cultures sampled at t_1 = 19 hpi, t_2 = 27 hpi and t_3 = 34 hpi according to untreated parasite morphology. Fold change was calculated compared to relative t_0 ($p<0.05$ except where indicated with *)

Several of the transcripts with decreased abundance translate to proteins that are known to require polyamines for optimal functioning, protection or gene expression in other organisms (Table 3.3). These, for example, include the transcript for DNA topoisomerase II (PF14_0316) [224], three oxidative stress defence transcripts [225] and transcripts involved with zinc transport and energy metabolism [226], which were all decreased. In addition, the transcript for a putative calcium/calmodulin-dependent protein kinase 2 (PFL1885c) was increased 2.4-fold and it is known that polyamines inhibit this protein [227]. Furthermore, the transcript abundance of several transcription factors was increased (PF11_0241, PFL0465c, PFE1245w, PFD0560w) and polyamines have been shown to act as transcriptional repressors. This may be indirectly via the regulation of specific transcription factors [226] e.g. c-Myc was shown to increase upon polyamine depletion [228].

3.3.5.3 EDGE time course analysis

Differential transcript abundance analysis was performed in addition with EDGE 1.1.208 software [215]. EDGE was specifically designed to calculate the significance of differential abundance in microarray time course experiments [215]. EDGE identified 718 significantly differentially affected unique transcripts (Fig. 3.18, Appendix B) compared to relative t_0 versus the 538 identified by LIMMA. The EDGE results include the transcripts for LDC, OAT and spermidine synthase (PF11_0301), but it do not indicate fold change or even direction of change. There was only about 35.6% overlap between the EDGE and LIMMA differential abundance data sets, but EDGE uses a default Q-value cut-off of 0.1% ($p<0.02$) for significance and no fold change cut-off compared to the $p<0.05$ significance and 1.7-fold change cut-offs specified in LIMMA. However, since EDGE was written for proper time course analysis with parallel time point comparison, the software had to be forced to compare all three treated time points with the relative t_0, which potentially interfered with the statistical calculations. For this reason and because of the advantage of definite fold change cut-offs with LIMMA, the LIMMA dataset of 538 was regarded as the dataset of choice to be used for further analysis.

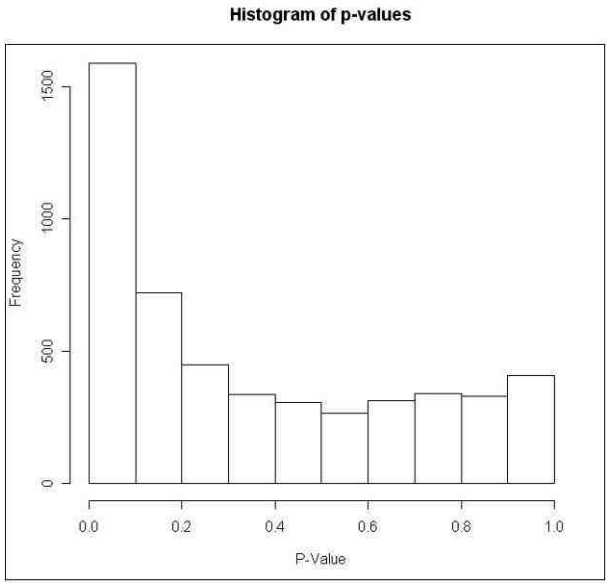


Fig 3.18 EDGE output in the form of a histogram of p-values of the oligonucleotides with changed expression. Note that the abundance of more than 1500 of the 5056 oligonucleotides analysed (including repeats and NULL-controls) changed during the time course with $p<0.1$. However, when the Q-value significance cut-off of 0.1% ($p<0.02$) was used, 718 unique transcripts (823 oligonucleotides) showed significant differential abundance.

3.3.6 GO assignment of differentially affected transcripts

The 538 transcripts identified by LIMMA were classified into 14 functional groups (Fig. 3.19) using GO terms obtained from DAVID and PlasmoDB. Most transcripts with increased abundance were related to RNA metabolism (9%), translation (10%) and host/parasite interaction (11%), whereas those with decreased abundance mostly represented DNA (7%) and primary metabolism (8%, including carbohydrate, lipid and energy metabolism). The increased abundance of transcripts associated with host/parasite interaction (including surface antigens) is regarded as a general stress response [189, 196]. Transcripts related to the

mitochondria (energy metabolism) and apicoplast (lipid metabolism), including organellar translation, generally decreased, whereas those related to ribosomal translation increased. As expected with growth arrest, cell cycle regulators were affected and three cyclin-associated transcripts had differential transcript abundance. The limited annotation status (only ~40%) of the *P. falciparum* genome [96] was reflected in the dataset with 51% of the transcripts encoding hypothetical proteins with unknown biological function.

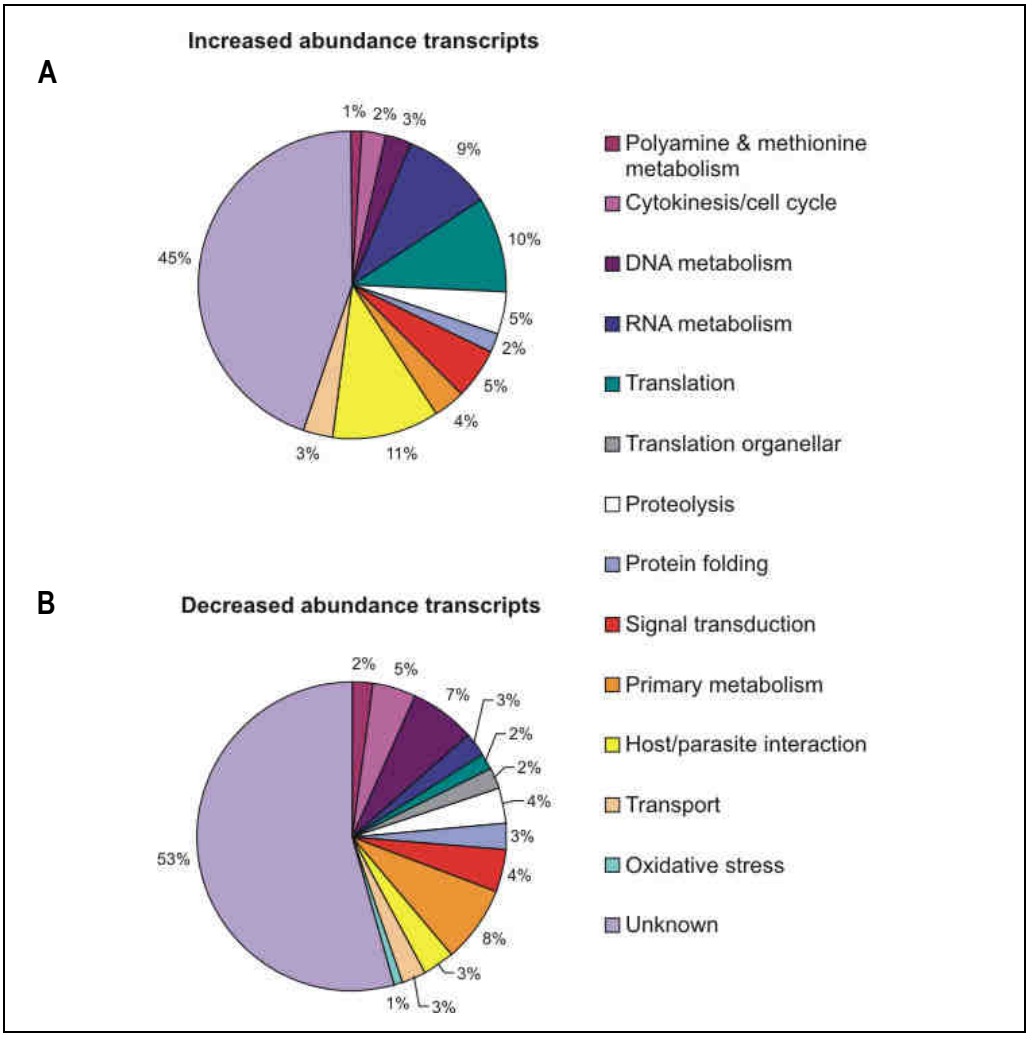


Fig. 3.19 Functional classification of transcripts with **A)** increased and **B)** decreased differential abundance (1.7-fold change, $p<0.05$) due to PfAdoMetDC/ODC co-inhibition, as determined by LIMMA analysis. The percentages indicated were calculated from the total number of increased (171) or decreased (377) transcripts. The terms used include a number of sub-processes i.e. DNA metabolism (including replication), RNA metabolism (including transcription), primary metabolism (carbohydrate, lipid and energy metabolism) etc., whereas the unknown category mainly include hypothetical ORFs.

3.3.7 Finding adjacently located genes with differentially affected transcripts

During data analysis, it was noticed that several transcripts of genes physically located adjacently or in close proximity at the chromosomal level were differentially affected. The LIMMA dataset was therefore evaluated for clusters of adjacently located genes, where a cluster was defined as four or more genes within a window of six adjacent genes of which the transcripts were present within the dataset. Seven such adjacent gene

clusters were found among the decreased abundance transcripts and none among the increased abundance transcripts. The seven clusters were distributed over chromosomes 7, 10 and 11, respectively (Table 3.4).

Table 3.4 Adjacent gene clusters with decreased abundance transcripts

Chromosome	Cluster 1 [a]	Cluster 2 [a]	Cluster 3 [a]
MAL7	MAL7P1.6 MAL7P1.7 PF07_0005 PF07_0006	MAL7P1.170 MAL7P1.173 MAL7P1.176 PF07_0129 MAL7P1.177	-
MAL10	PF10_0014 PF10_0015 PF10_0016 PF10_0017 PF10_0019 PF10_0020 PF10_0021 PF10_0022 PF10_0023 PF10_0024 PF10_0025	PF10_0152 PF10_0153 PF10_0154 PF10_0155	-
MAL11	PF11_0037 PF11_0039 PF11_0040 PF11_0041	PF11_0364 PF11_0365 PF11_0367 PF11_0368	PF11_0503 PF11_0504 PF11_0505 PF11_0508 PF11_0509

a. A cluster was defined as four or more genes within a window of six adjacent genes of which the transcripts were present within the dataset. Seven such gene clusters were detected, namely two on MAL 7, two on MAL10 and three on MAL11.

Particularly significant was a cluster of eleven genes on chromosome 10 (PF10_0014 to PF10_0025) that lie back-to-back on the same strand, with only one gene (PF10_0018) in the entire stretch of ~41,000 bases not present in the LIMMA dataset (Fig. 3.20). PF10_0018 produces a low abundance transcript, which was most likely affected in the same way, but was not detected.

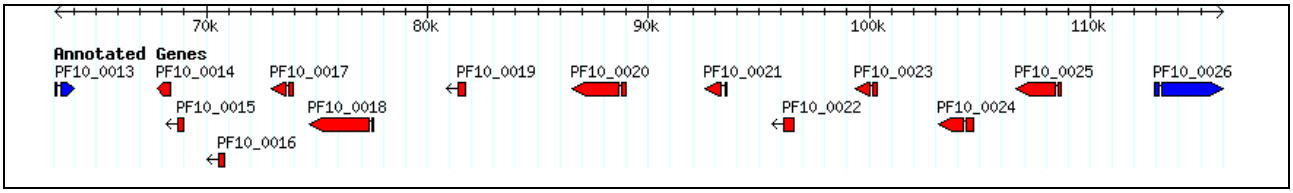
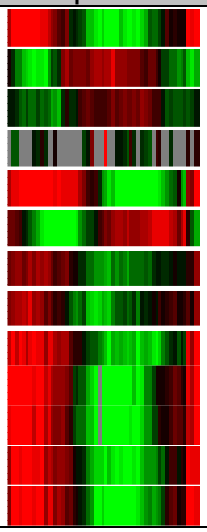


Fig. 3.20 Eleven gene cluster from chromosome 10 (PF10_0014 to PF10_0025) where the transcripts of eleven adjacently located genes (red arrows) were all decreased except for the low abundance transcript of PF10_0018, which could not be detected due to low signal intensity. Figure constructed within PlasmoDB 5.3.

Decreased transcription of co-localised genes may be due to a common transcription factor that became non-functional in the absence of polyamines. These twelve genes (including PF10_0018) were therefore analysed for motifs identified in other organisms with Transfaq and RSA tools, via the MADIBA database (http//www.bi.up.ac.za/MADIBA/MADIBA.html), as well as MEME (http://meme.sdsc.edu/) and Weeder (http://159.149.109.16/Tool/Ind.php), but a meaningful motif could not be identified. Finding known motifs in

the *P. falciparum* genome is a well-known problem, which is ascribed to divergence from other organisms and the high A+T-rich content of the genome [128, 229]. TATA-box-like motifs are characteristic of regulatory regions in other organisms but occur at random in the *P. falciparum* genome [128]. However, the genes within the identified clusters are not all co-expressed under normal conditions according to the 3D7 IDC transcriptome [91] (Table 3.5) and nuclear expression of contiguous genes is rarely co-regulated in *P. falciparum* [29]. A more likely explanation for the co-regulation may therefore be partial DNA unwinding due to polyamine depletion [230], thereby causing genes previously buried within the nucleosome particle to become more exposed to e.g. oxidative damage.

Table 3.5 IDC mRNA expression profiles of the eleven gene cluster from chromosome 10

PlasmoDB ID	Oligo ID	IDC time of peak expression (hpi)	IDC expression profile
PF10_0015	opfj12802	53	
PF10_0016	j3_21	28	
PF10_0017	33_20	37	
PF10_0018	j33_18	0	
PF10_0019	j33_16	10	
PF10_0020	j33_15	42	
PF10_0021	j33_12	9	
PF10_0022	j33_11	6	
PF10_0023	j33_10	53	
PF10_0024	j33_6	5	
	kn1473_2	4	
PF10_0025	f67629_1	6	
	j33_5	6	

The data above were obtained from the 3D7 IDC transcriptome (http://malaria.ucsf.edu) [91].

3.3.8 PfAdoMetDC/ODC-interactome data comparisons

In contrast to most other environmental perturbations of *P. falciparum* that have been evaluated on the transcriptome level, e.g. glucose deprivation [189], heat-shock [190] and even chloroquine treatment [179, 196], the exact protein and pathway targeted by DFMO/MDL73811 in this study were known, which enabled specific transcripts to be evaluated. The presence of eight transcripts from polyamine and methionine metabolism in the LIMMA dataset indicated the potential enrichment for transcripts from proteins functionally connected to this pathway or to PfAdometDC/ODC. To investigate this observation further, the LIMMA dataset was compared with the *in silico* predicted interactome data of PfAdoMetDC/ODC [113]. These networks were constructed using among others the IDC transcriptome [29, 113]. The transcripts of 23% of all the predicted binding partners of PfAdoMetDC/ODC were present in the LIMMA dataset, including 60% (12/20) of the top 20 scored (highest prediction) proteins (Table 3.6). To assess the possibility of random overlap, the LIMMA dataset was also compared with the interactome data of another bifunctional protein from outside the polyamine pathway and unaffected by the treatment, namely DHPS/PPPK (PF08_0095). The transcripts of

only about 10% of the predicted binding partners of this protein and only 10% (2/20) of the top 20 scored proteins were present in the dataset of 538 (Table 3.6). The interactome is a theoretical interaction prediction that still requires experimental verification, but this analysis may indicate enrichment of the LIMMA dataset for transcripts of proteins functionally connected to PfAdoMetDC/ODC or polyamine and methionine metabolism. The complete interactomes of these proteins and data comparison are provided in Appendix C.

Table 3.6 Interactome data comparisons

Score[a]	Protein ID	Description	Present in LIMMA dataset
PfAdoMetDC/ODC top 20 binding partners			
9.53	PF11_0317	Structural maintenance of chromosome protein, putative	
8.31	PFE0195w	P-type ATPase, putative	
7.98	PFA0390w	DNA repair exonuclease, putative	
6.62	MAL8P1.99	Hypothetical protein	Yes
6.62	PF11_0427	Dolichyl-phosphate b-D-mannosyltransferase, putative	
6.62	PF07_0129	ATP-dependent acyl-coa synthetase	Yes
6.62	PFA0590w	ABC transporter, putative	Yes
5.9	PF10_0260	Hypothetical protein	
5.9	PF13_0348	Hypothetical protein	
5.7	PF14_0053	Ribonucleotide reductase small subunit	Yes
4.71	PFD0685c	Chromosome associated protein, putative	Yes
4.71	PFC0125w	ABC transporter, putative	Yes
4.71	PF14_0709	Ribosomal protein L20, putative	Yes
4.71	PF08_0131	1-Cys peroxiredoxin	Yes
4.71	PF11_0117	Replication factor C subunit 5, putative	Yes
4.71	PF11_0181	Tyrosine tRNA ligase, putative	Yes
4.71	PFB0180w	5'-3' Exonuclease, N-terminal resolvase-like domain, putative	Yes
4.71	PFL2180w	50S Ribosomal protein L3, putative	
4.71	PF14_0097	Cytidine diphosphate-diacylglycerol synthase	
4.71	PF14_0081	DNA repair helicase, putative	Yes
DHPS/PPPK top 20 binding partners			
10.32	PF13_0140	Dihydrofolate synthase/folylpolyglutamate synthase	
8.31	PFL0740c	10 kd chaperonin, putative	
8.31	PF11_0258	Co-chaperone GrpE, putative	
8.31	PF13_0180	Chaperonin, putative	
7.98	PF08_0006	Prohibitin, putative	
7.98	PFL1475w	Sun-family protein, putative	
5.96	PF13_0234	Phosphoenolpyruvate carboxykinase	
5.96	PF11_0188	Heat shock protein 90, putative	
5.96	PF14_0656	U2 snRNP auxiliary factor, putative	
5.96	PF14_0242	Arginine n-methyltransferase, putative	
5.9	PFB0953w	Hypothetical protein	Yes
5.9	MAL7P1.209	AAA family ATPase, putative	
5.9	PFF0945c	Long-chain fatty-acid Co-A ligase and oxalyl Co-A decarboxylase	
5.9	PFE0060w	Hypothetical protein	
5.9	PF11_0076	Hypothetical protein	
5.9	PFF0775w	Pyridoxal kinase-like protein	
5.9	PF10_0013	Hypothetical protein	
5.9	MAL8P1.124	Hypothetical protein	
5.9	PF14_0705	Hypothetical protein	
5.9	PFE1245w	Zinc finger protein, putative	Yes

a. Probability score for functional linkage

3.3.9 Real-time PCR validation of differential transcript abundance data

The accuracy of the microarray data was validated with real-time PCR of three increased and three decreased transcripts including LDC, OAT and AdoMet synthetase. At the time, two complete hybridisation sets (unamplified and amplified) had been performed (amplified results not shown) and very limited quantities of cDNA were available. Several plasmodial "housekeeping" genes from other studies were considered to be used as loading control to obtain comparable starting levels e.g. ribosomal protein L37A, chromatin binding protein, Ser/Thr protein kinase [231] and LDH [179]. However, transcript levels of these either fluctuate during the IDC or were affected by PfAdoMetDC/ODC co-inhibition (Appendix A). However, the $\log_2$-ratios of a putative cyclophilin (PFE0505w) remained relatively unchanged across the different samples in the PfAdoMetDC/ODC co-inhibition data (Fig. 3.21A) and in the IDC transcriptome (Fig. 3.21B), and was used as "housekeeping" or cDNA loading control.

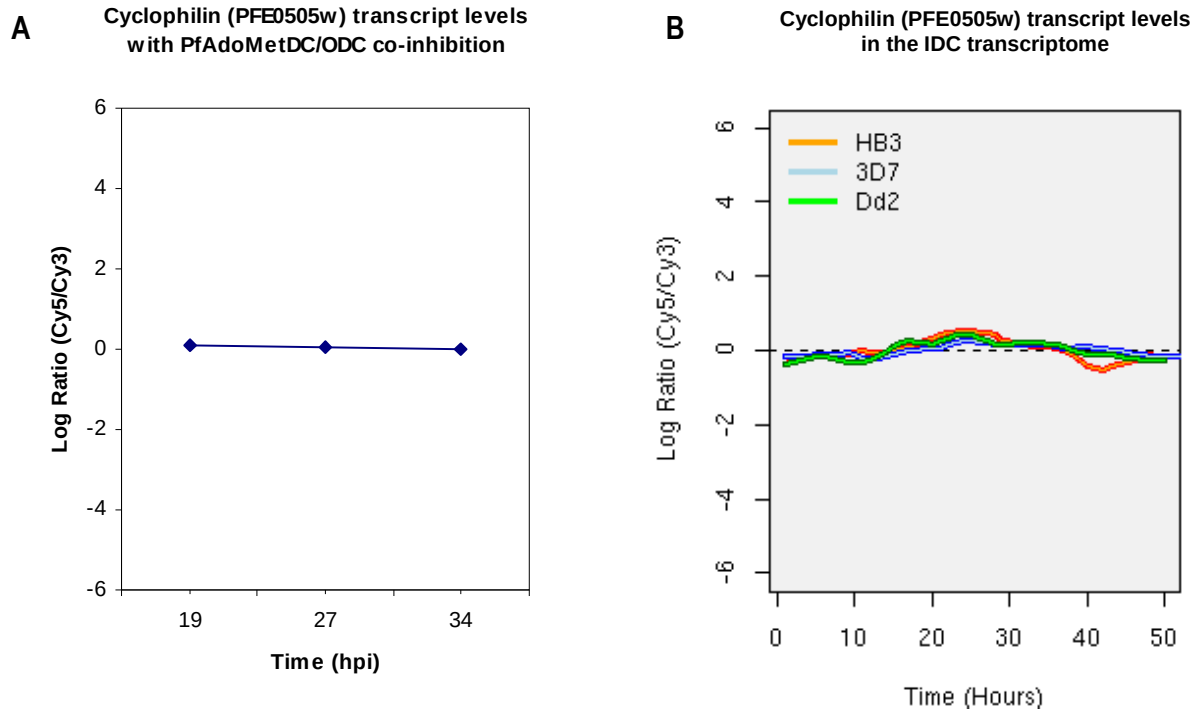


Fig. 3.21 The relatively constant transcription profile of the putative cyclophilin (PFE0505w) in **A)** the PfAdoMetDC/ODC co-inhibition data and **B)** the IDC transcriptome of three parasite strains (image in B obtained from http://malaria.ucsf.edu, [91]) enabled its use as "housekeeping" or cDNA loading control.

A five-part cDNA dilution series (1, 1/10, 1/20, 1/50 and 1/100) of the putative cyclophilin was prepared. The amplification process was visualised in real-time by plotting fluorescence against the cycle number on a logarithmic scale (Fig. 3.22). The cycle at which the sample fluorescence of a specific amplification product crosses the threshold for detection (above background) is the cycle threshold (C_t) and is indicative of the abundance of that specific cDNA species within the sample.

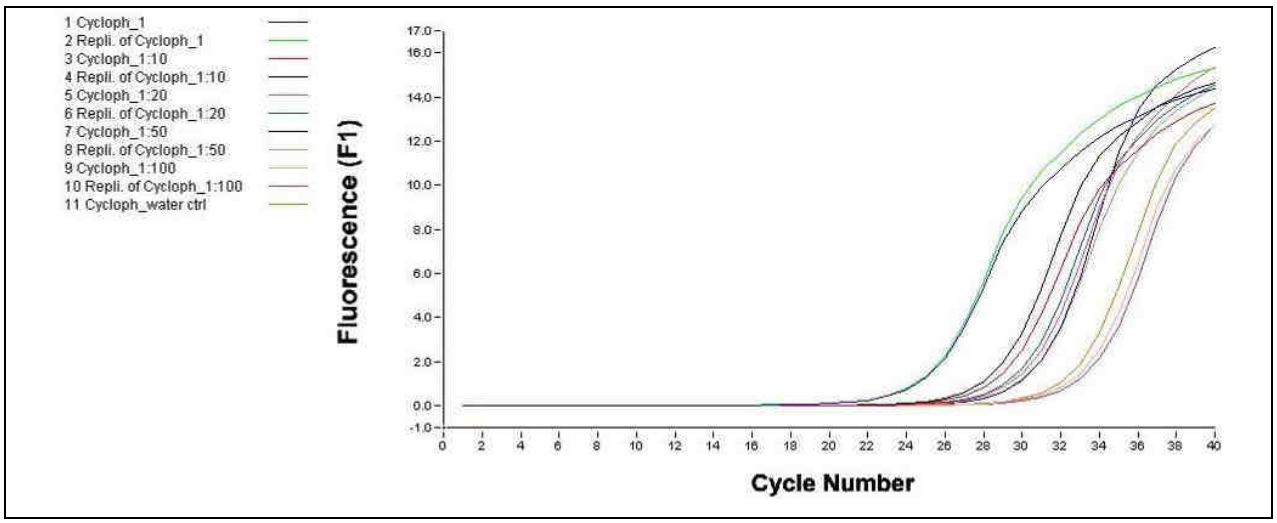


Fig. 3.22 A real-time PCR plot obtained for a five-part cDNA dilution series (1, 1/10, 1/20, 1/50 and 1/100) of the putative cyclophilin. The 1/100 dilution curve was detected in the same cycle as the negative water control, which indicated primer-dimer formation.

Melting curve analysis was performed to distinguish the correct amplification products from primer-dimers, since the SYBR Green dye intercalates non-specifically into all double-stranded DNA (Fig.3.23).

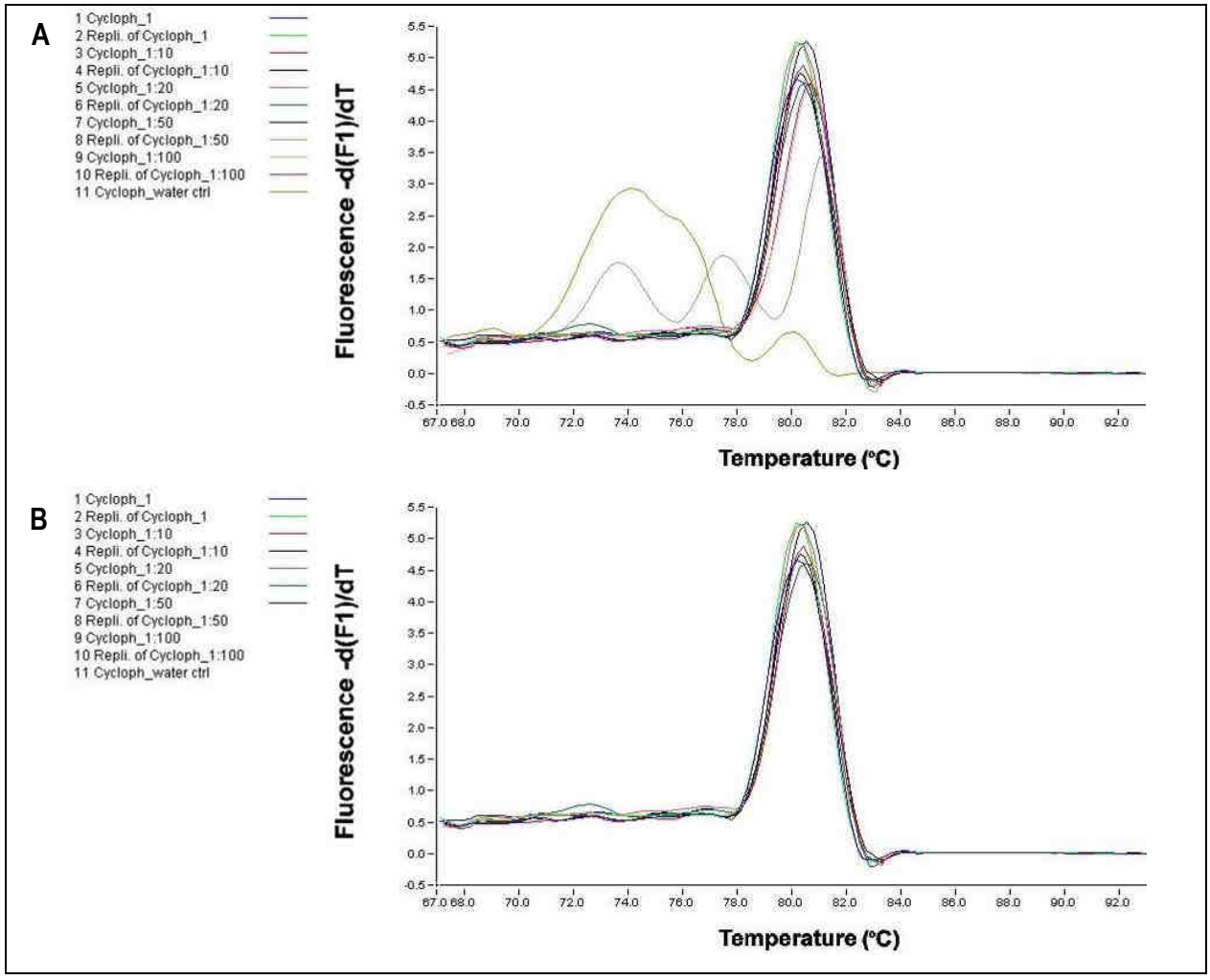


Fig. 3.23 Melting curve analysis of the amplification product of cyclophilin. **A)** The five-part cDNA dilution series confirmed primer-dimer formation at the higher sample dilutions (1/100 and 1/50 replicate). **B)** These values were subsequently excluded from the analysis and a 1/5 sample dilution were routinely used.

The short primer-dimers dissociate at a lower temperature than the 160-170 bp gene-specific products, which usually had a $T_m \approx 80°C$ (Fig. 3.23). Primer-dimers were detected in the very dilute samples (i.e. 1/50 to 1/100) and these data were excluded from the analysis. Sample dilutions of 1/5 were routinely used.

A standard curve was compiled from the C_t-values of the putative cyclophilin dilution series (Fig. 3.24) and was used to estimate the amount of cDNA of the six transcripts of interest (Table 3.7). The amount of cDNA obtained for the specific transcripts was divided by the amount of the putative cyclophilin in that same sample to normalise for the variation in sample concentration and resulted in a relative cDNA ratio or fold change.

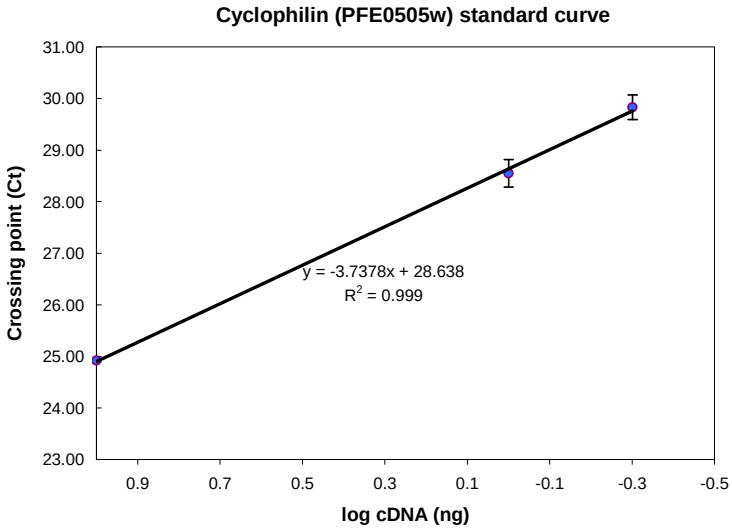


Fig. 3.24 A standard curve of the putative cyclophilin (PFE0505w).

Fold change values obtained with the two different mRNA quantitation techniques i.e. microarray and real-time PCR, were relatively consistent for the six transcripts tested (Table 3.7), which confirmed the reliability of the microarray analysis. Relatively low abundance transcripts within the dataset were validated by the inclusion of LDC and DHFR/TS. However, LDC cDNA levels were so low that they could not be assessed in the 1/5 dilution without primer-dimer formation and were therefore measured in the undiluted cDNA stock to obtain reliable data. Similarly, for DHFR/TS, of which the transcript abundance decreased with PfAdoMetDC/ODC co-inhibition (Table 3.3), cDNA levels could not be measured in the sample of maximum fold change (T_{t1}), but were subsequently detected in T_{t3}.

Table 3.7 Microarray data validation with real-time PCR

PlasmoDB ID	Annotation	Time point[a]	Fold change to relative t_0	
			Oligo array (±SD)[b]	Rt PCR (±SD)[c]
PFL1885c	Calcium/calmodulin-dependent protein kinase 2	t_3	2.43 (±0.18)	1.74 (±0.40)
PFD0285c[d]	LDC	t_3	2.78	1.71 (±0.19)
PFF0435w	OAT	t_3	1.88	2.14 (±0.30)
PF08_0131	1-Cys-peroxiredoxin	t_3	0.35	0.42(±0.14)
PFD0830w[d]	DHFR/TS	t_1	0.50 (±0.05)	Not detected[e]
		t_3	0.62 (±0.03)	0.35 (±0.04)
PFI1090w	AdoMet synthetase	t_1	0.42	0.53 (±0.07)

a. Treated sample of maximum increase/decrease in transcript abundance

b. Data of transcripts with multiple oligonucleotides ($p<0.05$) were averaged, the standard deviation of the mean in parentheses.

c. Real-time PCR was performed in triplicate with the standard deviation of the mean indicated in parentheses.

d. Relatively low abundance transcripts

e. Not detected in sample of maximum decrease in abundance (t_1) and repeated for t_3

3.4 DISCUSSION

The ability to measure the expression of thousands of genes in a single experiment simultaneously assured the rapid adaptation of DNA microarray into most biomedical research fields [232]. However, one of the biggest challenges to microarray studies is deciphering the significance of the wealth of information obtained from high-quality raw data [136]. In this study, cautious experimental design was used to ensure that the data obtained would be maximally informative regarding the effects of the perturbation [61]. A reference design microarray experiment was performed to enable easy comparison between samples and to simplify data analysis. Secondly, synchronous parasites were treated (with DFMO and MDL73811) to limit the background noise of various parasite stages to enable the detection of transcriptional responses above the basal transcriptional level in the IDC. Thirdly, two biological and two technical replicates were included to enable statistical analysis to determine the sifnificance of the observations. Fourthly, a time course study was performed to assess the effects of the perturbation at the highest possible resolution i.e. samples were harvested over three time points. Initial assessment of the quality of the data included visual inspection of the arrays and data flagging, followed by a series of diagnostic analyses to ensure appropriate data correction and normalisation. These analyses indicated that local background subtraction and robust spline normalisation had to be performed within each array to correct for spatial effects, whereas Gquantile between array normalisation enabled consolidation of replicates and cross-sample comparison for differential abundance analysis with LIMMA. The statistical significance of differential transcript abundance was calculated with moderated t-statistics [206] and only transcripts with at least a 1.7-fold change ($\log_2$-ratio ≥ 0.75 or ≤ -0.75) and $p<0.05$ were regarded as differentially affected. These thorough analyses confirmed data quality, which indirectly determined the reproducibility and reliability of the derived gene lists (e.g. the LIMMA dataset) and the validity of the resulting biological conclusions [232]. The raw microarray data and detailed information on the

experimental methodology and analysis (MIAME) were subsequently submittted to GEO, as required for publication purposes and to enable independent verification and meta-analyses.

Transcriptional profiling of cells treated with cytostatic drugs has previously been performed in cancer [233, 234] but global transcriptome studies with cytostatic drugs in multistage organisms such as *P. falciparum* have not yet been reported. The cytostatic effect of DFMO or MDL73811 inhibition on PfAdoMetDC/ODC in *P. falciparum* is well established [153, 154], but the exact mechanism by which polyamine depletion (induced by these drugs) results in growth inhibition, was not yet elucidated [235]. To aid in a better understanding of this process, the physiological response of *P. falciparum* during polyamine depletion induced cytostasis was evaluated on a transcriptional level. Transcriptional arrest was observed with three different analyses (phase ordering, Pearson correlation with the 3D7 IDC transcriptome [91] and within the PfAdoMetDC/ODC co-inhibition data). The transcriptional arrest preceding and resulting in cytostatic growth arrest due to polyamine depletion was demonstrated here for the first time to our knowledge in any organism. *P. falciparum* is a multistage organism and the transcriptional arrest of treated and normal transcriptional progression of untreated parasites was clearly visible when the data were ordered according to peak expression times within the IDC (Fig. 3.11A). Pearson correlation calculations indicated that the approximate time of transcriptional arrest occured in the late ring/early trophozoite stage, which correlates to the time of *PfAdoMetDC/ODC* transcription and the subsequent availability of the protein for drug inhibition (Fig. 3.11B). Complete enzyme inhibition occured soon after protein expression (Fig. 3.3) and underscored the enzyme-specific inhibitory effects of DFMO and MDL73811 [153, 154]. The exact mechanism by which polyamine depletion results in transcriptional arrest is not clear, but the importance of polyamines in macromolecular synthesis (including RNA and proteins, e.g. transcription factors) [88], optimal ribosome function [236] and the association of the main fraction of polyamines with RNA [64] has been demonstrated. In this study, the increased transcript abundance of several putative transcription factors and ribosomal components were detected (Table 3.3), which could indicate an attempt to induce transcription and translation (as opposed to DNA replication) in order to overcome the transcriptional arrest caused by the perturbation.

Despite the generalised transcriptional arrest, 538 transcripts with fold changes ranging between maximum 3.2-fold up and 5-fold down (Appendix A) were shown to be differentially affected with LIMMA analysis. The range of fold change detected is in agreement with other transcriptome reports of perturbed Plasmodia where relatively small amplitude transcriptional responses were detected [62], especially in the increased abundance datasets [189, 196]. These changes were quantitated compared to a relative t_0, which was used as reference point for quantitative analysis throughout the whole functional genomics investigation. Due to the transcriptional arrest of the treated and normal progression of untreated parasites, the standard parallel time point comparison approach of treated versus untreated would have indicated stage and life cycle differences and not the perturbation-specific effects of polyamine depletion. A similar relative t_0 strategy and fold change

cut-off were used in the analysis of the artesunate-perturbation of *P. falciparum* [198]. Furthermore, Temez and colleagues also compared trophozoites that were arrested by the inhibition of sphingomyelin biosynthesis with trophozoite controls rather than ring controls, to limit the detection of stage-specific transcriptional differences as opposed to perturbation-specific differences [199]. Both these studies used synchronised parasites [198, 199], as well as a relative t_0 strategy, and both detected perturbation-specific transcriptional responses similar to the study reported here. In comparison, the reported perturbations of Plasmodia that failed to detect programmed transcriptional responses generally employed asynchronous cultures [194, 196]. The transcriptional arrest demonstrated after PfAdoMetDC/ODC co-inhibition would have been masked if asynchronous cultures had been used and defining a reference point for quantitative analysis, i.e. relative t_0, would have been difficult. The use of synchronised cultures also enabled comparison with the IDC transcriptome and transcripts with treated profiles that deviated from their IDC profiles further corroborated the findings of the differential abundance analysis. It appears as if the parasite attempts to respond to environmental stress on the transcriptional level with a specific though small amplitude response, removed from the normal transcriptional control and these non-random changes may potentially be missed if asynchronous cultures are used [128].

The 538 transcripts include eight transcripts from polyamine and methionine metabolism that were differentially affected (Fig. 3.25). The presence of these eight transcripts in the LIMMA dataset suggested the potential enrichment of the data for transcripts from proteins functionally connected to this pathway or to PfAdometDC/ODC. Therefore, the LIMMA dataset was compared with the *in silico* predicted interactome data of PfAdoMetDC/ODC [113], which revealed 12/20 of the top 20 (highest probability) scored binding partners of PfAdoMetDC/ODC within the dataset compared to only 2/20 of the top 20 binding partners of an unrelated bifunctional protein, DHPS/PPPK, and corroborated the enrichment for transcripts from polyamine and methionine metabolism (Table 3.6).

The LIMMA dataset was furthermore classified into 14 functional groups (Fig. 3.19) using GO terms obtained from DAVID and PlasmoDB. In many cases, the co-inhibition caused both an increase and a decrease in abundance of transcripts representing the same biological process, e.g. proteolysis. These paradoxical effects of polyamines or depletion thereof are not uncommon, e.g. in mammalian cells polyamine depletion increased the half life of long-lived proteins, but decreased the half life of short-lived proteins [237]. In the current investigation, the highest percentages of increased transcripts were related to RNA metabolism, translation and host/parasite interaction, whereas most of the decreased transcripts represented DNA and primary metabolism (including carbohydrate, lipid and energy metabolism). The increase in abundance of transcripts associated with host/parasite interaction (including surface antigens) is regarded as a general stress response [189, 196]. More than half of the dataset (51%) represents hypothetical proteins with unknown biological function at the current time, which limits data interpretation. Recently, a database for GO annotation prediction

of *P. falciparum* was released, named PlasmoDraft [238]. Instead of sequence homology, PlasmoDraft uses post-genomic (transcriptomic, proteomic and interactomic) data for "guilt by association", based on the similarity of gene expression profiles. This database could suggest identities and the biological significance for many of these hypothetical proteins. The dataset of 538 was validated with real-time PCR of three increased and three decreased transcripts and their differential abundance was confirmed (Table 3.7). Relatively low abundance transcripts within the dataset were validated by the inclusion of LDC and DHFR/TS in the real-time PCR strategy.

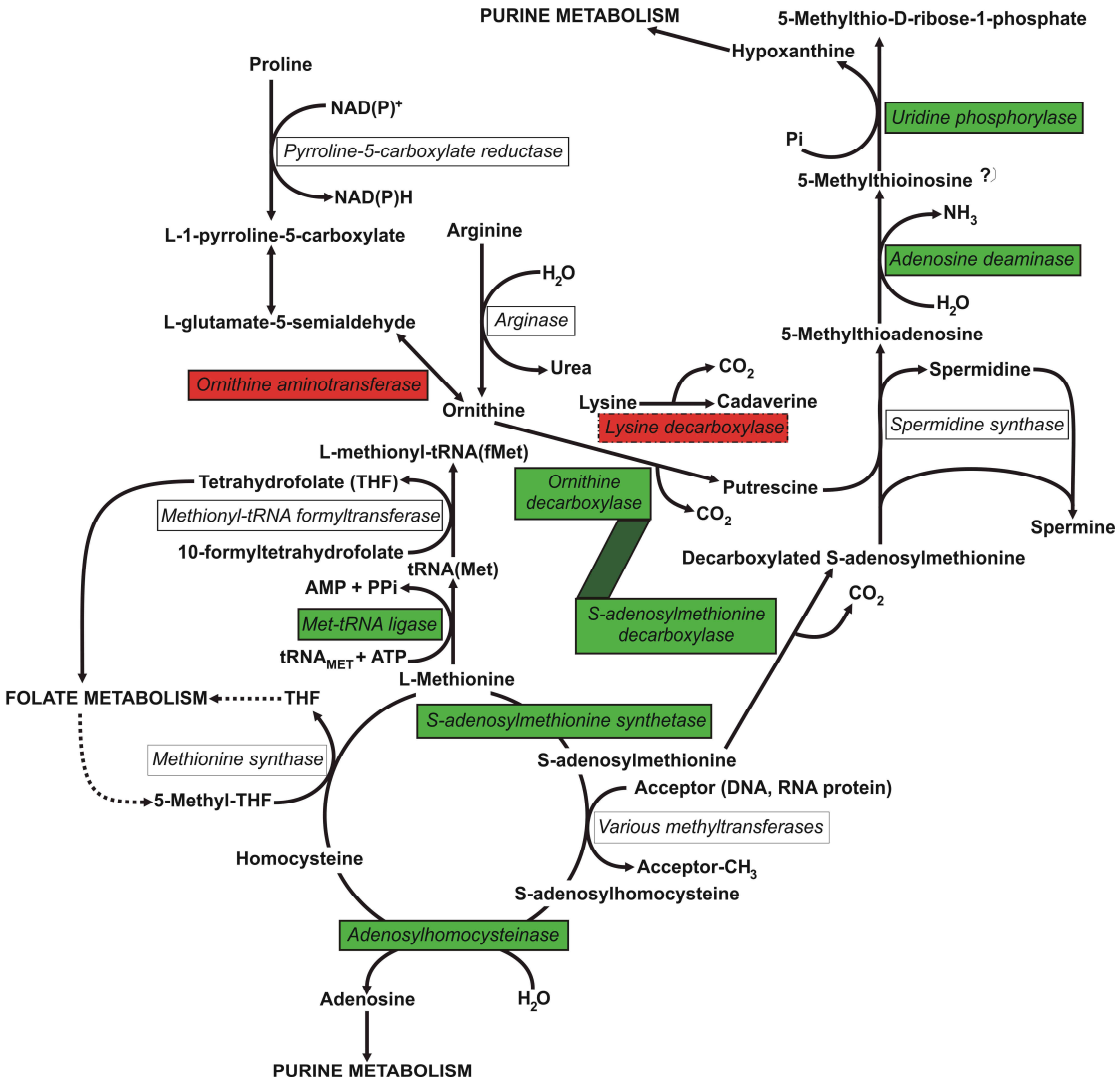


Fig. 3.25 Polyamine and methionine metabolism (adapted from MPMP at http://sites.huji.ac.il/malarial/). Plasmodial spermine synthesis is currently believed to be catalysed by spermidine synthase as indicated [85]. Enzymes of which the transcript abundance was significantly increased due to PfAdoMetDC/ODC co-inhibition are indicated in red and those significantly decreased are indicated in green.

The abundance of the majority (~70%) of the 538 transcripts was decreased with PfAdoMetDC/ODC co-inhibition. However, in two investigations of DFMO-induced polyamine depletion (Loikkanen, 2005, PhD thesis, University Oulu)[219] the majority of transcripts were increased. The latter study used twice the DFMO dose and similar exposure times compared to the co-inhibition study reported here. However, MDL73811-treatment

of *T. brucei brucei* resulted in a 20-fold increase of AdoMet [157], which is the principal biological methyl donor in trans-methylation of e.g. DNA, RNA, proteins and phospholipids [239]. The accumulation of AdoMet, rather than polyamine depletion, was proposed to be the antitrypanosomal mechanism of MDL73811 [157]. Therefore, the general transcriptional suppression observed in the current investigation could be due to transcriptional silencing as a result of AdoMet accumulation causing hypermethylation of e.g. histones [125] or 2-deoxycytosine bases within gDNA. However, there is contradicting evidence as to the latter as an epigenetic mechanism in Plasmodia [124, 126, 127]. In addition to polyamine depletion, co-inhibition of PfAdoMetDC/ODC may potentially alter the methylation status of *P. falciparum* due to increased levels of AdoMet. The decrease of the transcript for AdoMet synthetase (PFI1090w, Table 3.3) may act as a compensatory strategy induced to maintain AdoMet levels. The exact mechanism behind this regulation needs to be elucidated, but AdoMet concentration in MDL73811-treated mammalian cells was effectively regulated through substrate feedback inhibition of AdoMet synthetase activity [240]. This is not the case in *P. falciparum* as it was demonstrated that AdoMet does not allosterically regulate AdoMet synthetase [79], but regulation of AdoMet synthetase by AdoMet at the transcriptional level needs to be investigated. In contrast, the trypanosomal AdoMet synthetase is apparently poorly regulated, resulting in the substantial accumulation of AdoMet with AdoMetDC inhibition [157]. S-adenosylhomocysteine (AdoHcy) is the major by-product of AdoMet-dependent trans-methylation and a competitive inhibitor of trans-methylation reactions [239, 241]. The AdoMet/AdoHcy ratio is an indicator of cellular methylation status and a decrease in this ratio is associated with reduced methylation potential [241]. Therefore, the decrease in abundance of the transcript for adenosylhomocysteinase (PFE1050w), which catalyses the hydrolysis of AdoHcy [241], may be an attempt to restore the AdoMet:AdoHcy ratio and cellular methylation status. Transcriptome analysis of individually inhibited *P. falciparum* AdoMetDC and ODC is currently under way to distinguish between the effects of polyamine depletion with and without the proposed AdoMet accumulation. AdoMet levels and the role of gDNA methylation in the mode of action of MDL73811 in *Plasmodium* are discussed further in Chapter 4.

The transcript for PfAdoMetDC/ODC was decreased ~2-fold with DFMO/MDL73811-treatment (Fig. 3.25, Table 3.3). In accordance, the transcripts for the proteins DHFR/TS and DHPS/PPPK, targeted by pyrimethamine and sulphadoxine in the study of antifolate-treated *P. falciparum*, were also decreased [197]. DHFR/TS expression is regulated by the binding of the protein to its own mRNA, thus acting as a negative feedback to control its own translation [242]. A similar mechanism could apply to PfAdoMetDC/ODC, but this needs to be elucidated. The decrease in the transcripts for a putative adenosine deaminase (PF10_0289) and uridine phosphorylase (PFE0660c) was postulated to be due to decreased spermidine synthesis (resulting from reduced levels of putrescine and dcAdoMet) and thus less production of 5-methylthioadenosine and subsequently 5-methylthioinosine. An extensive metabolomics investigation was subsequently performed to resolve these hypotheses, as is discussed in Chapter 4.

Polyamines interact with DNA, RNA and proteins [63, 64] and it is, therefore, difficult to discriminate where their main regulatory effects are. In the present investigation, many of the differentially affected transcripts translate proteins that are known to require polyamines for optimal functioning in other organisms (Table 3.3). These include the transcript for DNA topoisomerase II (PF14_0316), which was decreased with PfAdoMetDC/ODC co-inhibition and it was shown in mammalian cells that the enzyme activity also decreased upon polyamine depletion [224]. Moreover, polyamines inhibit calcium/calmodulin-dependent protein kinase 2 [227] and stimulate casein kinases [243]. In accordance, the transcript for a putative calcium/calmodulin-dependent protein kinase 2 (PFL1885c) was increased and that of casein kinase I (PF11_0377) was slightly decreased by DFMO/MDL73811-treatment in *P. falciparum*. Polyamines protect macromolecules from peroxidation reactions [65, 66] and increase the expression of oxidative stress defence genes (including glutathione reductase) in *Escherichia coli* [225]. In the absence of polyamines in the current investigation, the transcripts for the antioxidant protein 1-cys-peroxiredoxin (PF08_0131), as well as glutathione reductase (PF14_0192) and a putative glutathione S-transferase (PF14_0187), were all decreased. Polyamines were also shown to increase the expression of genes related to energy metabolism and iron/zinc transport in *E. coli* [226]. In the current investigation, seven transcripts associated with oxidative phosphorylation (CoI, CoxI, CoxIII_2, PF11_0412, MAL7P1.75, PFE0970w, PF13_0121); five concerned with glycolysis (PF10_0155, PF13_0141, PF14_0378, PF14_0598, PFF1300w) and the transcript for a putative zinc transporter (PF07_0065) were significantly decreased. Although there are differences between polyamine metabolism of different organisms [82], these analogies may indicate that polyamines have a regulatory or stabilising effect on the transcripts of these proteins in *P. falciparum.*

Polyamines can also act as transcriptional repressors, which may be indirectly via the regulation of specific transcription factors [226]. Expression of the transcription factor c-Myc was shown to increase upon polyamine depletion [228]. PfAdoMetDC/ODC co-inhibition caused increased transcript abundance of several transcription factors (Table 3.3), including that of a hypothetical protein (PF11_0241) with Myb-like domains (SANT, Homeodomain-like and Myb domains, http://www.plasmodb.org). The Myb proteins, e.g. PfMyb1, are sequence-specific transcription factors that regulate the expression of genes implicated in growth and cell cycle regulation [244]. Furthermore, there was an increase in abundance of the transcript for a C2H2-type zinc-finger transcription factor, krox1 (PFL0465c) and two other putative transcription factors, namely a putative CCCH-type zinc-finger protein (PFE1245w) and a hypothetical protein (PFD0560w) with a TATA-box binding protein-like domain (http://www.plasmodb.org). Interestingly, a TATA-binding protein-associated factor was shown to be important in the regulation of mammalian polyamine transport for maintenance of basal polyamine levels [88] [245]. In addition, the transcript abundance of a putative transcription factor iib (PFE0415w) was decreased.

Polyamines are required for the degradation of cyclin B1 mRNA in the G1-phase of the classic eukaryotic cell cycle to commit cells to enter the S-phase for completion of the cycle [68]. On the other hand, polyamines stabilise cyclin D1, which upon their depletion result in the decrease of this cyclin and cell cycle arrest [63, 68]. Growth arrest in the trophozoite stage of *P. falciparum* due to ODC and/or AdoMetDC inhibition resembles the late-G1 of the eukaryotic cell cycle [88]. The homologues of these specific cyclins are not yet known within the *P. falciparum* genome but three cyclin-associated transcripts were differentially affected as a result of the perturbation (Table 3.3). The transcripts for proliferating cell nuclear antigen (PF13_0328) and a hypothetical protein (PF14_0604) with cyclin homology (http://david.abcc.ncifcrf.gov/) were both decreased. In addition, the transcript for a hypothetical protein (PFL1330c), implicated in cell cycle regulation within the liver stage [246], was increased. The transcript levels of none of the protein kinases associated with cyclins in *P. falciparum* were differentially affected.

Several clusters of adjacently located genes with a co-ordinated decrease in transcription were identified (Table 3.4), including an eleven gene cluster from chromosome 10 (PF10_0014 to PF10_0025, Fig. 3.20). However, the genes within these clusters are not normally all co-expressed according to the 3D7 IDC transcriptome [91], making the decrease/dysfunction of a common transcription factor unlikely. Moreover, nuclear expression of contiguous genes are rarely co-regulated in *P. falciparum* [29]. A more logical explanation may therefore be partial DNA unwinding due to polyamine depletion [230], thereby causing genes previously buried within the nucleosome particle to become more exposed to e.g. oxidative damage. The observed differential co-regulation of these transcripts after polyamine depletion reiterates a perturbation-specific effect on the transcriptome as a result of polyamine depletion.

The most dramatic perturbation-specific transcriptional response observed was the 2.8-fold increase in abundance of the transcript for LDC and ~2-fold increase of that for OAT (Table 3.3, Fig. 3.17). Both of these proteins are intricately involved with polyamine metabolism and their differential abundance indicates a transcriptional mechanism or potential compensatory feedback to overcome the perturbation (Fig. 3.25). The increase in the transcript for OAT may be a compensatory mechanism or buffering effect to prevent toxic ornithine accumulation [247] as a result of ODC inhibition by DFMO. OAT regulates ornithine metabolism and catalyses both synthesis of ornithine from glutamate-5-semialdehyde when levels are low and its degradation to proline and glutamate when present in excess [248]. Lysine decarboxylation produces cadaverine, a diamine and structural analogue of putrescine with one additional methylene group (Fig. 1.4) [69]. LDC activity and cadaverine accumulation have been reported to alleviate ethylene inhibition of arginine decarboxylase and AdoMetDC in pea seedlings [74] and 0.4 mM cadaverine reversed DFMO-induced growth arrest to some extent in *P. falciparum* [154]. Otherwise there is little known about the biological role of LDC or cadaverine in *P. falciparum*, but the Plasmodial recombinant enzyme was successfully expressed and lysine decarboxylation could be demonstrated [79]. The transcripts for LDC and PfAdoMetDC/ODC are expressed at approximately

the same time in *P. falciparum* (25 hpi and 24 hpi respectively), which is expected, should LDC serve as a compensatory mechanism for polyamine/diamine biosynthesis. The transcript for LDC was increased and that of PfAdoMetDC/ODC was decreased, again illustrating the differential effect of the co-inhibition on the abundance of specific transcripts despite the generalised transcriptional arrest. The increase in the transcript for LDC may indicate a potential resistance mechanism, should PfAdoMetDC/ODC be clinically targeted in future. Interestingly, a lysine decaboxylase-like protein was also among the LIMMA dataset, but it was 2.5-fold decreased. This protein is much smaller than LDC (39 kDa versus 280 kDa) and it also contains a putative LDC domain (PlasmoDB 5.4), but activity has not yet been demonstrated. An arginine decarboxylase (associated with putrescine biosynthesis in microorganisms and higher plants, [81, 249]), has not yet been identified within the *P. falciparum* genome (C Wrenger, personal communication).

The programmed compensatory mechanisms (as opposed to a random transcriptional response), which resulted in the highly specific increased abundance of the OAT and LDC transcripts upon polyamine depletion, are also supported by other transcriptome studies of polyamine-depleted *P. falciparum.* Increased transcript levels of OAT were reported after treatment with DFMO alone [219], but were not affected when the downstream enzyme, spermidine synthase (which should not result in ornithine accumulation), was inhibited (J. Becker, unpublished data). However, an increase in LDC transcript abundance was recently also detected with both ODC (K. Clark, unpublished data) and spermidine synthase inhibition (J. Becker, unpublished data), to compensate for the resulting polyamine depletion. In contrast, the transcripts for neither OAT nor LDC were increased after exposure of *P. falciparum* to a variety of perturbations, including a series of antimalarial drugs and environmental stressors (M. Llinás, unpublished data) [189, 196, 198]. It therefore appears as if the increased abundance of LDC and OAT are transcriptional responses specific to the perturbation of polyamine metabolism in *P. falciparum.*

In this chapter the differential abundance of specific transcripts involved in polyamine and methionine metabolism as compensatory responses, as well as the co-regulation of clusters of genes as result of the perturbation, provides evidence of a drug-specific response. Taken together, these results, as well as evidence from other perturbations [179, 189, 190, 198, 199], provide support for the ability of the parasite to react to environmental pressure at the transcriptional level.

In the following chapter, the functional genomics investigation will proceed with proteomics and metabolomics analyses of PfAdoMetDC/ODC co-inhibited *P. falciparum* to confirm the conclusions of the transcriptomics investigation. In addition, specific hypotheses will be investigated with biochemical analyses.

3.5 RAW DATA AND SUPPLEMENTARY WEBSITE

The microarray raw data can be downloaded from PUMAdb (http://puma.princeton.edu/cgi-bin/publication/viewPublication.pl?pub_no=523) and additional data can be obtained from the supplementary website (http://genomics-pubs.princeton.edu/PfAdoMetDC_ODC) supporting the publication that resulted from this work (*van Brummelen et al. J. Biol. Chem, 10.1074/jbc.M807085200)*. The microarray data can also be accessed at the NCBI's GEO [250], accession GSE13578, according to MIAME recommendations.