Letter

Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle

Karine G. Le Roch,[1,2,6] Jeffrey R. Johnson,[1,6] Laurence Florens,[1,6,7] Yingyao Zhou,[2] Andrey Santrosyan,[2] Munira Grainger,[3] S. Frank Yan,[2] Kim C. Williamson,[4] Anthony A. Holder,[3] Daniel J. Carucci,[5] John R. Yates III,[1] and Elizabeth A. Winzeler[1,2,8]

[1]Department of Cell Biology ICND202, the Scripps Research Institute, La Jolla, California 92037, USA; [2]Genomics Institute of the Novartis Research Foundation, San Diego, California 92121, USA; [3]Division of Parasitology, National Institute for Medical Research, London NW7 1AA, United Kingdom; [4]Department of Biology, Loyola University Chicago, Chicago, Illinois 60626, USA; [5]Naval Medical Research Center, Malaria Program (IDD), Silver Spring, Maryland 20910, USA

To investigate the role of post-transcriptional controls in the regulation of protein expression for the malaria parasite, *Plasmodium falciparum*, we have compared mRNA transcript and protein abundance levels for seven different stages of the parasite life cycle. A moderately high positive relationship between mRNA and protein abundance was observed for these stages; the most common discrepancy was a delay between mRNA and protein accumulation. Potentially post-transcriptionally regulated genes are identified, and families of functionally related genes were observed to share similar patterns of mRNA and protein accumulation.

[Supplemental material is available online at www.genome.org.]

Despite the identification of few genes encoding transcriptional regulators in the genome of *Plasmodium falciparum* (Gardner et al. 2002), analysis of gene expression in the malaria parasite has shown that transcription is generally monocistronic and developmentally regulated (Lanzer et al. 1992, 1994; Alano et al. 1996; Horrocks et al. 1996; Cheesman et al. 1998; Scherf et al. 1998). The identification of promoter regions regulating transcription has also remained elusive, yet large-scale analyses of the parasite's transcriptome have shown that there appears to be relatively good correspondence between the timing of gene expression and when its product is required by the cell (Bozdech et al. 2003; Le Roch et al. 2003). High-throughput transcript and protein-expression profiling techniques have emerged as powerful tools to study the biology of an organism in a systems-wide manner, and the union of these data sets can help to discern the role of post-transcriptional regulation by comparing mRNA and protein abundance levels.

Several putative RNA-binding proteins have been identified in the *P. falciparum* genome, including two members of the Puf family of RNA-binding proteins known to regulate translation and RNA stability (Cui et al. 2002), as well as proteins with sequence similarity to UBA2, a promoter-independent mRNA-stabilizing protein originally identified in *Arabadopsis thaliana* (Lambermon et al. 2002). In addition, a bioinformatic survey of transcription-associated proteins (TAPs) present in the *P. falciparum* genome indicated that TAPs are much less abundant in *P. falciparum* compared with other eukaryotic organisms, but proteins modulating mRNA decay and translation were the most abundant TAPs detected in the genome (Coulson et al. 2004). The presence of these RNA-binding proteins indicates that

mechanisms of post-transcriptional control should be expected in *P. falciparum*, and post-transcriptional regulation has been described for genes involved in sexual differentiation (Vervenne et al. 1994; Dechering et al. 1997), mitochondrial RNA processing (Rehkopf et al. 2000), and the stability of RNAs encoding surface antigens (Lanzer et al. 1993; Levitt et al. 1993).

Quantitative analyses of mRNA levels measured at nine time-points of the *P. falciparum* life cycle (six asexual intraerythrocytic, merozoite, late gametocyte, and salivary gland sporozoite) using a short oligonucleotide array (Le Roch et al. 2003), and semiquantitative analyses of protein levels measured at seven stages (ring, trophozoite, schizont, merozoite, gametocyte, gamete, and salivary gland sporozoite) detected by multidimensional protein identification technology (MudPIT) (Florens et al. 2002; J.R. Johnson, L. Florens, M. Grainger, Y. Wu, A.A. Holder, D.J. Carucci, and J.R. Yates, in prep.) provide abundance profiles for thousands of parasite transcripts and proteins, with abundance correlated to specific stages and time points in the parasite's life cycle. Here, we investigate the correspondence between mRNA transcript and protein abundance, and consider the effects of mRNA stability and/or post-transcriptional regulation on the timing of mRNA and protein detection. Using a systems-wide approach, families of genes sharing similar patterns of post-transcriptional regulation are identified, as well as putative sequence motifs that may confer regulatory effects.

Results

Overlap between mRNA and protein detection

Our expectation was that there would be good overlap between the set of transcripts and proteins detected for different stages. Overall, 4294 transcripts were detected in at least one of the six stages examined, whereas 2904 proteins were detected in at least one of the seven stages (Supplemental Tables 1 and 2). For all stages, 2584 genes and their cognate proteins were found in both

analyses, representing 89% of the identified proteins and 60% of the detected transcripts (Table 1). Of the 319 proteins whose corresponding transcripts were not detected in any of the mRNA time-points, 38 were represented by a small number of probes on the microarray (≤6), and therefore, were assigned low P-values and filtered out of the transcriptome data set. The majority of these 319 proteins were detected at very low levels in the pro-teomic data set (60% were single peptide hits) and 81% were hypothetical proteins. However, several had virtually no mRNA expression in the transcriptome data set and were detected at significant levels in the asexual stages, such as glutathione per-oxidase (PFL0595c) and glyoxalase I (PF11_0145).

On the other hand, proteins were not found for 2167 (50%) transcripts detected during the microararray analysis. Clearly, the microarray approach provided a greater depth of detection with, on average, 2.7 times as many transcripts as proteins detected at a particular stage, except for the merozoite stage, where relatively fewer transcripts were detected (1474 transcripts vs. 1077 pro-teins). It is certainly possible that the transcriptional activity of the merozoite stage is lower than any other stage of the malaria parasite, but the low number of merozoite transcripts observed could have been the result of the merozoite extraction protocol (see Methods), which involves long incubations that may lead to significant RNA degradation. On the basis of the mRNA intensity distributions of genes whose protein products were detected or not, the transcripts with a corresponding protein detected in the asexual erythrocyte cycle were, on average, 2.5 times more abun-dant than transcripts for which no proteins were detected (Fig. 1), demonstrating the tendency for proteomics methodologies to preferentially detect proteins of higher abundance. The majority of these transcripts were of moderate abundance (61% had mRNA expression values <100). However, this was not true for the sporozoite and gametocyte stages, where cumulative mRNA intensity distributions were not significantly different for genes where protein products were or were not detected. The most abundant asexual stage transcripts that lacked detection of their corresponding protein were mostly annotated as hypothetical, but notable exceptions were the circumsporozoite protein (CSP, PFC0210c) and the chloroquine resistance transporter (MAL7P1.27). The CSP is a sporozoite-specific protein, but the detection of untranslated CSP mRNA transcripts in the asexual stages has been described (Levitt et al. 1993).

In addition, 770 of 2167 genes (36%) for which no protein products were detected appeared to code for proteins predicted to have an N-terminal signal sequence (566, 26%)—that is, proteins secreted or targeted to subcellular compartments—or with trans-membrane domains (464, 21%) or glycosylated phosphatidyl-
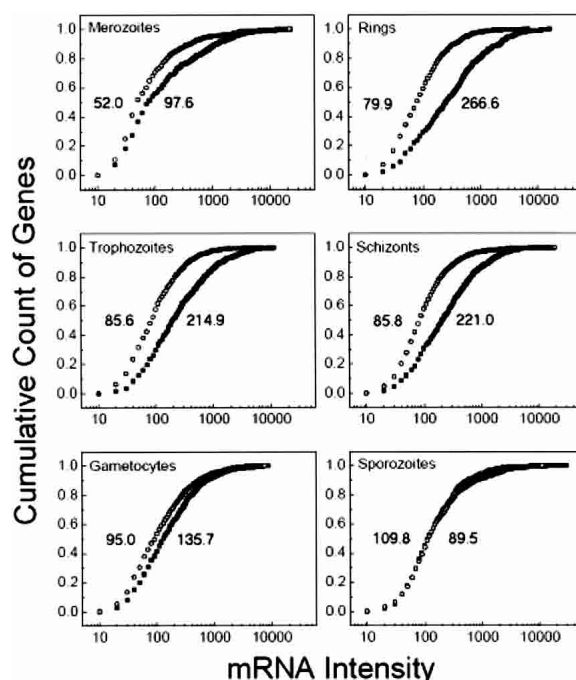
Figure 1. Cumulative distributions of mRNA intensity. Cumulative in-tensity distributions of mRNA intensity are plotted for each stage where both transcriptome and proteome data was available. Intensity distribu-tions are represented for the cognate proteins that were not detected (open squares), or were detected (closed circles) in the same stage by the proteomic analysis, with median intensities indicated to the *left* (proteins not detected) and *right* (proteins detected) of the curves.

inosital (GPI) anchors (38, 2%). The proportion of such proteins in the whole *P. falciparum* genome is 37% (1993 of 5276 genes). This trend was also observed in the gametocyte and sporozoite analyses, with 38% and 40% of the missing proteins predicted to be nonsoluble, respectively. This observation illustrates a bias in the proteomic analysis of whole-cell lysates, in that such meth-ods may fail to detect secreted or membrane proteins present in low abundance. Nine genes uniquely detected here at the tran-scriptome level were recently highlighted as potential secreted proteins (pPIESPs) in the targeted proteomic analysis of erythro-cyte surface proteins (Florens et al. 2004). Two other genes (PFC1080c and PF11_0014) encode Maurer's cleft transmem-brane proteins, (annotated as PfMC_2TMs), which were detected in the proteomic analysis of coimmunoprecipitated complexes (Sam-Yellowe et al. 2004).

Table 1. Overlap between microarray and proteomic data sets

		Protein stage							
		Mero.	Ring	Troph.	Schiz.	Gameto.	Gamete	Sporo.	Total mRNA
mRNA time-point	Ring	766	551	834	666	758	535	658	2533
	Troph.	884	614	941	766	916	651	842	3295
	Schiz.	904	623	910	778	909	635	834	3217
	Mero.	493	397	513	415	458	340	385	1474
	Gameto.	851	574	861	690	949	657	853	3363
	Sporo.	594	417	585	476	618	421	576	2111
Total protein		1077	717	1118	861	1197	763	1206	

The number of genes/proteins to be compared excluded mitochondrially encoded genes, ribosomal RNA transcripts, and proteins without complimentary probes on the microarray.

Correlation of mRNA and protein abundance levels

It has been suggested that there is little correlation between mRNA and protein abundance in yeast as well as higher eukaryotes (Ander-son and Seilhamer 1997; Futcher et al. 1999; Gygi et al. 1999). To test this in *P. falciparum*, we compared mRNA and protein abundance lev-els using the Spearman rank corre-lation. mRNA abundance was rep-resented by the expression levels as
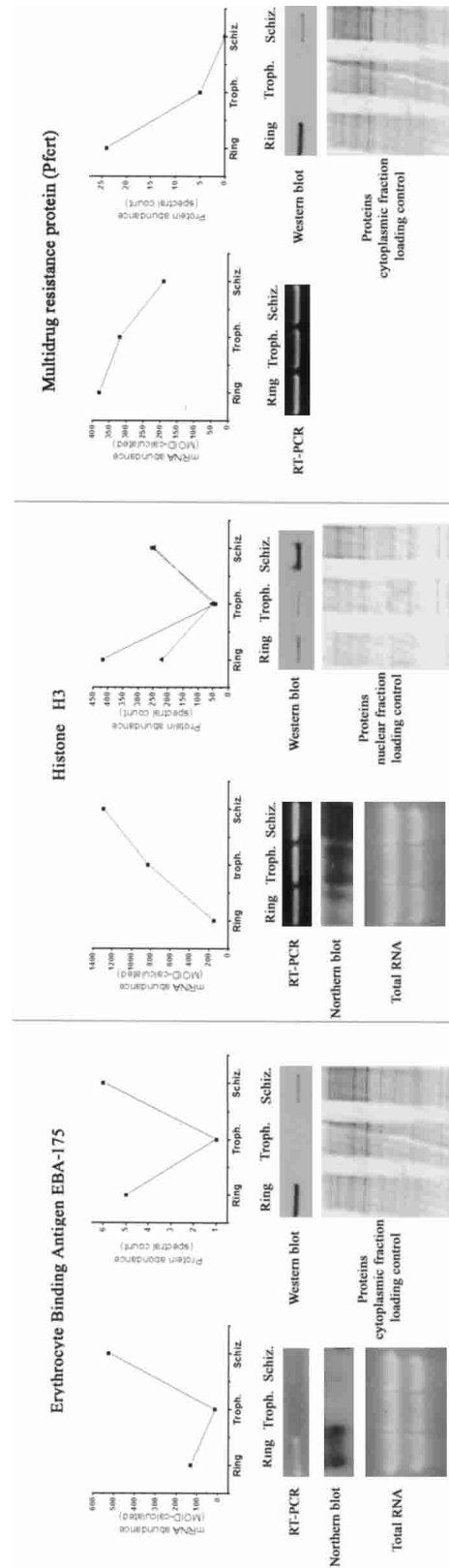
Figure 2. Validation of mRNA and protein abundance using a single parasite population. mRNA and protein expression levels obtained by oligonucleotide array analysis and MudPIT, respectively, were validated for the erythrocyte-binding antigen 175 (*EBA-175*, PF07_0128, closed circles), the histone H3 (MAL6P1.106, closed circles and MAL6P1.248, closed triangles), and the chloroquine-resistance transporter (*Pfcrt*, PFE1150W) using RT–PCR, Northern blot (for PF07_0128 and MAL6P1.248), and Western blot analyses as indicated. MOID-calculated mRNA abundance and spectral count-estimated protein abundance are indicated in the *top* graphs. Probes specific for the MAL6P1.248 gene were not represented on the oligonucleotide array. Discrepancies between the wet experiments (RT–PCR, Northern blot, and Western blot) and the whole-genome experiments should be expected, as they were performed on different parasite populations extracted at different times.

calculated by the MOID algorithm, and protein abundance was estimated by the number of MS/MS spectra identified per protein (spectral count). Similar semiquantitative parameters have been shown recently to provide a reliable estimate of protein abundance when more precise quantitative approaches are unavailable or unfeasible (Florens et al. 2002; Gao et al. 2003; Durr et al. 2004; Liu et al. 2004). Because proteins with equal spectral count values are assigned identical ranks, a minimum spectral count range of 10 was enforced to avoid conflicts arising from proteins assigned equal ranks, which is particularly common for proteins of low abundance. To validate the applicability of using transcript levels obtained by microarray analysis and protein levels obtained by MudPIT analysis, RT–PCR, Northern, and Western blots were compared with microarray and MudPIT abundance measurements for three genes observed in this data set (Fig. 2). MOID-calculated levels of mRNA abundance and spectral count estimates of protein abundance sufficiently represented those values obtained by conventional, gene-by-gene approaches.

Each proteome stage observed a moderately high Spearman correlation with its corresponding transcriptome stage (self-correlation, Table 2) in the range of those reported in the literature for quantitative proteomic analyses, where three independent comparisons of mRNA and protein levels in *Saccharomyces cerevisiae* produced partial positive Spearman-rank correlation coefficients of 0.21 (Griffin et al. 2002), 0.45 (Washburn et al. 2003), and 0.57 (Ghaemmaghami et al. 2003). Although comparable to these other studies, the Sr values obtained here indicate a higher positive correlation (up to 0.59) between mRNA and protein data sets, which are clearer in scatterplots of mRNA and protein abundance (Fig. 3). Rather than focusing on absolute Sr values, we compared the relative differences between Sr values for each transcriptome with all proteome data sets to determine which stage's proteome correlated best with the transcriptome of interest (Table 2). For example, the trophozoite-stage transcriptome correlated better with the trophozoite proteome (Sr = 0.583) than with any other stage's proteome. This is in contrast to comparisons between unrelated stages, such as that for the comparison between the schizont transcripts and the gametocyte proteins, where the Spearman correlations are very low (Sr = 0.230). Self-correlations were also highest for the schizont and sporozoite data sets. In the case of the gametocyte data sets, the Sr values were low (average Sr = 0.31) compared with other correlations. Interestingly, the gametocyte transcriptome correlated best with the proteome of the following stage, the gamete proteome (Sr = 0.447). This shift was also observed for other stages, where Sr values were highest for the comparison between a stage's transcriptome and the proteome of a later stage. The merozoite stage transcriptome, for example, correlated best with

the ring-stage proteome (Sr = 0.731), whereas the ring-stage transcriptome correlated best with the trophozoite proteome (Sr = 0.655). This phenomenon suggests a regulatory effect at the level of mRNA stability and/or translation.

## Individual expression-profile correlation and gene-by-gene analysis

The asexual erythrocytic stages of the parasite, consisting of the merozoite, ring, trophozoite, and schizont stages, provide a full cycle with which to analyze stage transitions of the parasite and the corresponding fold changes in mRNA and protein abundance over these transitions. Scatterplots were constructed to evaluate the correspondence between mRNA and protein fold changes during each transition in the parasite's asexual erythrocytic cycle (Fig. 3). The scatterplots indicate a weak correlation between the mRNA and protein data sets, which is supported by the suboptimal Spearman-rank values obtained above. The scatterplots are divided into four quadrants, where genes observing complementary changes in mRNA and protein abundance would fall into quadrants I and III, whereas genes with discrepancies would fall into quadrants II and IV. On average, 55% of data-points fell into quadrants I or III, indicating that most mRNA/protein fold changes follow the same trend. These are genes for which regulation of expression is mostly achieved at the transcriptional level. For the data-points that fell into quadrants II or IV, replotting the same mRNA fold change points against protein fold change measured for the following transition fixed the majority of these points, relocating them into quadrants I or III 74% of the time (Fig. 3, insets). Thus, for most genes that do not observe a response in protein levels complementary to the changes in mRNA levels, the corresponding protein changes are usually observed in the next transition. This indicates post-transcriptional mechanisms for controlling protein levels.

The Spearman-rank correlation and scatterplot analyses suggested a delay between mRNA and protein accumulation. To investigate specifically which genes or families of genes observed this trend, the correlation coefficient between mRNA and protein-expression profiles for each gene considered significantly regulated within the asexual erythrocytic stages (spectral count range ≥10) was calculated (Supplemental Table 3). To consider the possibility of a time-shift, a forward-shifted correlation coefficient was also computed, in which the order of the transcript stages was shifted forward, such that the proteome of each stage was correlated with the transcriptome of the preceding stage. As a control, and to establish a false-positive rate for this analysis, a reverse-shifted correlation coefficient was computed, in which the order of the transcript stages was shifted in reverse, so that the proteome of each stage was correlated with the transcriptome of the following stage. The correlation coefficient between mRNA and protein-expression profiles improved by 0.5 for 171 of 459 genes (37%) when the forward time-shift was considered, whereas the reverse-shifted transcript profiles improved the correlation coefficient by 0.5 for only 72 genes (16%). Interestingly, almost all genes involved in glycolysis observed significantly improved correlations when the transcript stages were for-

Table 2. Spearman rank correlation between mRNA and protein abundance

		Protein stage						
		Mero.	Ring	Troph.	Schiz.	Gameto.	Gamete	Sporo.
mRNA stage	Mero	0.527	**0.731**	0.646	0.532	0.443	0.482	0.577
	Ring	0.434	0.598	**0.655**	0.450	0.293	0.352	0.294
	Troph.	0.434	0.515	**0.583**	0.484	0.279	0.362	0.247
	Schiz.	0.502	0.533	0.412	**0.511**	0.177	0.287	0.242
	Gameto.	0.258	0.319	0.263	0.230	0.370	**0.447**	0.278
	Sporo.	0.423	0.559	0.510	0.435	0.449	0.461	**0.551**

Self-correlation coefficients are boxed. Proteomic datasets best correlating with mRNA levels are bolded.
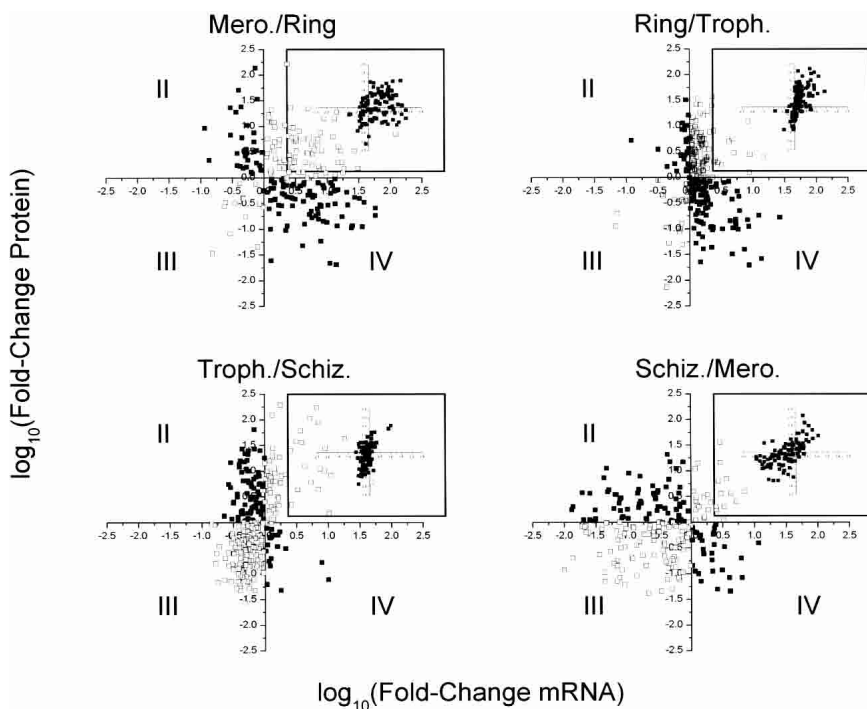
Figure 3. Scatterplots of mRNA and protein fold-changes during developmental transitions. The fold-change in mRNA abundance vs. the fold-change in protein abundance is plotted for each transition in the asexual erythrocytic cycle. Quadrants are designated with roman numerals. Data points falling into quadrants I and III were given open circles, data points falling into quadrants II and IV were given closed circles. Data points that fell into quadrants II and IV were replotted (*inset*) as the mRNA fold change of the transition indicated vs. the protein fold change of the following transition (i.e., the *inset* of Mero./Ring is a plot of the fold change in mRNA abundance for the merozoite-to-ring transition vs. the fold change in protein abundance for the ring-to-trophozoite transition).

ward shifted, including hexokinase (MAL6P1.189), glucose-6-phosphate isomerase (PF14_0341), phosphofructokinase (PFI0755c), triose phosphate isomerase (PF14_0378), glyceraldehyde-3-phosphate dehydrogenase (PF14_0598), phosphoglycerate kinase (PFI1105w), phosphoglycerate mutase (PF11_0208), enolase (PF10_0155), and pyruvate kinase (MAL6P1.160). The only enzyme involved in glycolysis that was not apparently time shifted was aldolase, which is known to have a role in cell invasion (Jewett and Sibley 2003). Several proteins localized to the rhoptry organelle involved in cell invasion shared similar time-shift patterns within their respective families, but the two families of rhoptry proteins were quite different. None of the high molecular-weight rhoptry complex (RhopH) proteins (PFC0120w, PFC0110w, PFI0265c, PFI1445w) appeared to be time shifted, observing unshifted correlation coefficients ranging from 0.81 to 0.99. In contrast, the low molecular-weight rhoptry complex proteins (PFE0075c, PF14_0102, PFE0080c) all exhibited significant time-shifts, with their mRNA abundances peaking in the schizont stage and their protein abundances peaking in the trophozoite stage of the following cycle. Unshifted correlation values were high for virtually all genes encoding 40S and 60S ribosomal proteins, and only six of 57 (Supplemental Table 3) had their correlation values improved by 0.5 or greater when considering a forward time-shift.

In addition to the asexual erythrocytic cycle, mRNA and protein data were available for sporozoites, gametocytes (both mRNA and protein data), and gametes (protein data only). Com-

paring the gametocyte mRNA transcripts without cognate proteins in the gametocyte proteome with the proteins detected at the gamete stage identified a set of genes involved in ubiquitin-dependent degradation. These included six ubiquitin-conjugating E2 enzymes and three ubiquitin hydrolases, all of which had protein products detected in the gamete-stage proteome, but were completely absent in the gametocyte proteome, where their mRNA transcripts were detected. The ubiquitin-activating enzyme E1, however, was detected more abundantly in the gametocyte proteome than in the gamete proteome, and its mRNA was detected in the gametocyte transcriptome.

Sequence analysis of untranslated regions

To identify putative regulatory sequence elements for genes sharing common mRNA/protein accumulation patterns, we attempted to align sequences in the 5′- and 3′-untranslated regions (UTRs) of post-transcriptionally regulated genes identified by the individual mRNA/protein correlations of significantly regulated genes in the asexual erythrocytic cycle (Supplemental Table 3). Consensus motifs identified by MEME analysis (Bailey and Elkan 1994) were often highly repetitive (i.e., poly-G or poly-C sequences), many were very short sequences that were not specific to the group of genes being aligned, and most motifs were not aligned with considerable confidence for a significant number of genes in the group (data not shown).

Alternatively, known RNA-binding motifs were analyzed for their presence in the 3′-UTRs of significantly regulated *P. falciparum* transcripts. One of the most well-characterized mRNA-destabilizing elements is the AU-rich element (ARE) found in the 3′-UTR of many unstable mammalian mRNAs (Chen et al. 1995). However, the A/T-rich nature of the *P. falciparum* genome makes the identification of a functional ARE nearly impossible. Similarly, an RNA consensus motif in *Plasmodium* that was observed to enhance the expression of a sexual-stage protein (a combination of a polyadenylation signal and a U-rich region (Golightly et al. 2000) could not be correlated with the post-transcriptional events identified here due to the A/T composition of this motif.

Recently, two RNA-binding proteins belonging to the Puf family of translation and mRNA stability factors were identified in the *P. falciparum* genome, one of which (PfPuf1) was shown by a yeast three-hybrid system to bind in vivo specifically to the Nanos-responsive elements (NREs) found in the 3′-UTR of the *Drosophila hunchback* (*hb*) mRNA (Zamore et al. 1997; Cui et al. 2002). Furthermore, the mRNA expression of both Puf proteins was limited to the gametocyte stage, and the PfPuf1 protein was maximally detected in the gametocyte stage proteome. It has not yet been determined whether NRE elements in *P. falciparum* play a role in post-transcriptional regulation. Only 95 genes in the *P. falciparum* genome contained an NRE-like sequence (see Methods). The *P. falciparum* genes that contained the NRE sequence in

their 3′-UTRs and were considered to be significantly regulated at the protein level (spectral count range ≥10) are shown in Table 3. Of the seven genes with transcripts detected in the gametocyte stage, six had disproportionately low protein expression (spectral count <10) in the gametocyte proteome (PFL2225w, MAL13P1.214, PFC0435w, PFI1565w, and PFE1250w); the latter two showing a sharp increase in protein levels in gametes. Similarly, for six of seven genes with mRNA expression in the sporozoite stage, we either did not observe protein expression in the sporozoite proteome (PFL2225w, PFE1250w, PFC0435w, PF14_0697, MAL7P1.222), or the protein was detected at disproportionately low abundance (PFI1565w). The three proteins expressed during the erythrocytic cycle displayed the delayed translation phenomenon, where the protein products maximally accumulated one stage after the transcripts' levels peaked. This suggests that the delayed translations observed in both asexual cycle and sexual stages are likely to involve similar mRNA-binding proteins; however, the low number of *P. falciparum* genes containing an NRE-like sequence indicates that other mechanisms likely facilitate this delay.

Discussion

Differential mRNA/protein detection, Spearman-rank correlation coefficients, and individual mRNA/protein expression profile correlations indicate significant discrepancies between mRNA and protein abundance in *P. falciparum*. The most frequent discrepancy observed was a delay between the maximum detection of an mRNA transcript and that of its cognate protein. Figure 2 illustrates this phenomenon; the comparison of mRNA and protein abundance by RT–PCR, Northern blot, and Western blot experiments shows that, whereas the EBA-175 gene does not exhibit a time shift, the histone H3 gene does. It is important to note that these mRNA and protein extractions for Northern blot, Western blot, and RT–PCR analyses were carried out using the same pool of parasites, yielding results similar to the genome-wide analyses, with minor discrepancies resulting from using different parasite populations. Although this display was a widespread phenomenon, it did not appear to be a nonspecific effect, considering that families of functionally related genes shared similar patterns of mRNA and protein accumulation. Similarly,

the mRNA intensity distribution for detected and undetected proteins (Fig. 1) indicated that the nonoverlapping transcripts in the asexual erythrocytic stages were commonly missed by proteomic analysis, because they were low-abundance transcripts, whereas those for the sporozoite and gametocyte stages were of equal abundance to the overlapping transcripts, possibly indicating a more prevalent occurrence of post-transcriptional events during these stages of development. On the basis of the time-shift hypothesis, transcripts/proteins that are not shared between the two analyses could be most interesting to study, as these should allow us a glimpse of the physiology of the preceding or following stages. For example, in the case of sporozoites, transcripts for which no proteins were detected are very likely to code for proteins involved in the physiology of one of the most cryptic cellular forms of the life cycle, the liver schizont. Genes implicated in cell cycle regulation and cell division, possibly involved in the schizogony of the liver stage, were often detected at the transcriptional level in sporozoites. This is especially true for the cdc2-homolog kinase (MAL6P1.27) or the cell cycle regulator (PFL1330c), as well as their associated cyclins (PFL2280w and PFE092c). A large number of DNA polymerases (e.g., PF10_0165), DNA replication licensing factors (e.g., PF07_0023), or helicases (e.g., PF14_0437) were also detected. These genes are undoubtedly involved in the massive cell divisions that occur in the liver stage, and represent the parasite's ability to pre-adapt to the human host. Similarly, for the gametocyte stage, where mRNA and protein abundance appear least correlated, the prefabrication of transcripts in the gametocyte stage could allow for very rapid production of proteins that would be required for gametogenesis, one of the most rapid cell-division processes observed in any organism.

These three independent bioinformatic interpretations of experimental data imply mechanisms of post-transcriptional control, either involving interplay between mRNA stability and degradation, gene-specific control of mRNA translation, or a combination of both. The prefabrication and latent storage of untranslated mRNA transcripts has been previously demonstrated for select genes in *Plasmodium*. *Pbs21*, a *P. berghei* sexual-stage gene whose mRNA is detected in female gametocytes, does not maximally accumulate its cognate protein until the ookinete stage (Paton et al. 1993; Vervenne et al. 1994). Similarly, tran-

Table 3. mRNA and protein expression profiles for genes with UTRs containing the NRE binding domain

Locus	Description	Protein abundance[a]							mRNA abundance[a]					
		M	R	T	S	G	Ga	Sp	M	R	T	S	G	Sp
PFL2225w	Myosin A tail domain interacting protein MTIP, putative	9	0	0	22	0	0	0	476	154.5	282	843.9	2576	844.5
PFE1250w	Long-chain fatty acid CoA ligase, putative	25	42	21	21	3	20	0	2390	1012	1180	1488	1210	346.6
PFI1565w	Conserved protein	2	8	0	17	1	36	5	115	115	879	1405	938.5	931.2
PFC0435w	Hypothetical protein	3	0	5	21	0	2	0	0	14.2	405	453.7	231.4	59.4
PF14_0327	Methionine aminopeptidase, type II, putative	0	0	10	0	5	0	2	93.9	284.3	405	145.4	190.4	106
PF14_0697	Dihydroorotase, putative	0	0	9	2	17	0	0	37.8	82.2	101	79.4	162.4	68.2
MAL13P1.214	Phosphoethanolamine N-methyltransferase, putative	229	92	4	63	0	0	15	0	128.6	286	342.7	49.6	0
PF11_0454	Ribosomal protein, 40S subunit, putative	11	2	17	4	4	6	0	0	467.2	355	256	0	0
MAL7P1.122	Conserved GTP-binding protein, putative	14	4	1	16	13	0	0	46.4	88.2	158	83.6	0	17.1
PF10_0325	Hypothetical protein, conserved	3	4	1	14	0	0	0	0	98	101	101.9	0	0

[a](M) merozoite; (R) ring; (T) trophozoite; (S) schizont; (G) gametocyte; (Ga) gamete; (Sp) sporozoite.

scripts encoding for the circumsporozoite protein (CSP) are detected in the asexual erythrocytic stages, but the protein is not maximally detected until the mature sporozoite stage (Ruiz i Altaba et al. 1987; Levitt et al. 1993). CSP utilizes heterologous polyadenylation sites, but a majority of CSP transcripts lack poly(A) tails (Ruvolo et al. 1993), suggesting the possibility that CSP transcripts are stored in a deadenylated form until their protein product is required. Comparative genomic analysis of transcriptional controls in *P. falciparum* revealed the presence of a disproportionately high number of proteins involved in mRNA stability and deadenylase activity encoded in the genome, suggesting that post-transcriptional controls may dominate transcriptional controls in regulating gene expression throughout development (Coulson et al. 2004). Post-transcriptional control at the level of mRNA translation is not unlikely in *P. falciparum*, considering the presence of structurally distinct ribosome populations during different stages of development and differentiation (Waters et al. 1989; McCutchan et al. 1995; Li et al. 1997). A-type ribosomes are predominant in the asexual erythrocytic stages of the *P. falciparum*, whereas a transition from the A- to C-type ribosome occurs during sexual differentiation in the mosquito midgut (Waters et al. 1989). The relatively constant population of the A-type ribosome in the asexual erythrocytic stages of the parasite indicates that this type of post-transcriptional control is likely most prevalent during sexual differentiation, while the parasite undergoes a switch from one ribosome type to another.

Gene expression in parasitic protozoa has shown unique mechanisms of regulation. For instance, the parasitic protozoa *Leishmania*, and other members of the trypanosomatidae, have polycistronic transcription with maturation of their mRNA by transplicing events (Agabian 1990). Nuclear run analyses in *Leishmania* or *Trypanosoma* species have shown that the steady-state level of mRNA is essentially regulated by post-transcriptional regulation and/or mRNA stability (Beetham et al. 1997, 2003; Burchmore and Landfear 1998; Martinez-Calvillo et al. 2003; Soto et al. 2003;). Although polycistronic transcription is common in bacteria and Archaea, it has been assumed for many years that this phenomenon does not occur in eukaryotic cells. However, recent works have shown that polycistronic transcription exists for 15% of the genes in the nematode *Caenorhabditis elegans* (Blumenthal et al. 2002). Whereas trypanosomatidea species show a transcriptional regulation closely related to bacteria or Archaea cells, transcriptional regulation of *P. falciparum* genes appears more closely related to eukaryotic cells. Transcription in *P. falciparum* is monocistronic; however, the genome contains a dearth of genes encoding transcription factors compared with other eukaryotic genomes (Gardner et al. 2002; Coulson et al. 2004). Post-transcriptional mechanisms may impart advantages for the parasite to quickly adapt to environmental changes, and could explain the common usage of such mechanisms throughout Apicomplexa.

Discrepancies between mRNA and protein abundance also have implications in malaria biology for those genes that have been studied primarily at the level of mRNA. The histones of *P. falciparum* are a prime example of this case, because they have been studied extensively using Northern blot analysis and are determined to be developmentally regulated throughout the stages in the asexual erythrocytic cycle (Lobo and Kumar 1999; Przyborski et al. 2003). Transcription profiles indicated that histone transcript abundance was low or absent in the ring stage, accumulating to high levels in the trophozoite and schizont stages. As such, cell cycle models and assumptions regarding

DNA replication and the timing of mitotic events were based on these transcriptional profiles. In light of the proteome data, however, it appears that histone mRNA transcripts are not immediately transcribed, and histone protein abundance in the ring stage is actually quite high for histone genes H2A (PFC0920w), H2B (PF07_0054), H3 (MAL6P1.106), and H4 (PF11_0061). This was confirmed by Western blotting, Northern blot, and RT–PCR analyses (Fig. 2). The decrease in histone abundance observed during the ring-to-trophozoite stage is confounding, as the onset of DNA synthesis occurs during the early trophozoite (Inselburg and Banyal 1984; Graeser et al. 1996). The discordance between histone abundance and DNA content in the ring stage may indicate that the histones possess a gene-regulatory function in the asexual erythrocytic cycle, although further biochemical evidence will be required to validate this hypothesis. The histones in the ring stage could function in a general regulatory manner by repressing overall transcription in areas of the genome that are packed into nucleosomes. Although most models of nucleosome remodeling involve modifications of the histone tails (Rice and Allis 2001), nucleosomes have been observed to unfold completely at transcriptionally active promoters (Boeger et al. 2003) and to repress transcription on a genome-wide scale in yeast (Sabet et al. 2003), suggesting that a down-regulation of histone protein abundance could similarly be associated with increased overall transcription in trophozoites.

It remains unclear what signals are required to cause a significant delay between mRNA and protein accumulation. The in silico identification of a consensus sequence conferring RNA stability in the UTRs of *Plasmodium* transcripts is difficult without functional characterization of the UTR sequences. RNA-binding motifs are frequently A/T-rich, but the abnormally high A/T-richness of *Plasmodium* genome makes distinguishing functional motifs from random intergenic sequence nearly impossible. The NRE-binding sequence recognized by the PfPuf family was identified in several genes observing abnormal correlations between their mRNA and protein-expression profiles. It has been implicated that differential polyadenylation of transcripts in *P. falciparum* could confer some regulatory effects, but it remains to be determined whether putative mRNA stabilization sequences like the NRE sequence modulate stabilization by recruiting polyadenylation factors. To this end, we are currently investigating the polyadenylation of transcripts for genes with and without the NRE sequence.

By comparing transcriptome and proteome data throughout the *P. falciparum* life cycle, we show a slightly positive correlation, demonstrating the importance of transcriptional regulation in the malaria parasite. In addition, we show that most discrepancies observed between mRNA and protein abundance are due to a time shift observed between the detection of the transcript and its cognate protein, an effect observed within specific gene families.

Methods

Methods for microarray analyses

Parasite preparation and microarray analysis were performed as previously described (Le Roch et al. 2003).

Parasite material for proteomic analyses

Sporozoites, gametocytes, trophozoites, and merozoites were prepared as described in Florens et al. (2002). Additional stages were prepared as follows.

Ring and schizont stages

P. falciparum line 3D7 was maintained in vitro using either 10% human serum or 0.5% (w/v) ALBUMAX I in RPMI 1640 medium (Guevara Patino et al. 1997). Highly synchronous developmental stages were obtained using a MACS type-D depletion column with a SuperMACS II magnetic separator (Miltenyi Biotec GmbH). Schizonts and merozoites were purified as described previously (Florens et al. 2002; Taylor et al. 2002). To produce young ring stages, purified schizonts were mixed with uninfected erythrocytes and allowing merozoites to invade erythrocytes for 1 h. The remaining schizonts were removed using the magnet. Schizont-free ring stages were treated with either saponin or streptolysin O to release erythrocyte proteins.

Pfs230⁻ gametes

The gene for Pfs230 was disrupted at 1356 bp by targeted integration of pDT.Tg23.230-D1.356 into *P. falciparum* (strain 3D7) parasites (Eksi et al. 2002). Pfs230⁻ clone D1.356a was maintained in culture in the presence of 500 ng·mL^{-1} pyrimethamine and gametocytogenesis was induced (Ifediba and Vanderberg 1981). N-acetyl-D-glucosamine (50 mM) was added to the gametocyte cultures on days 9–12 to eliminate asexual parasites, and on day 14, gamete/zygote production was induced by incubating the gametocytes in emergence medium (10 µM xanthurenic acid, 1.67 mg·mL^{-1} glucose, 8 mg·mL^{-1} NaCl, 8 mM tris(hydroxymethyl)aminomethane (Tris)-HCl at pH 8.2) at RT. One hour later, emerged gametes were isolated at the 6%–11% interface of a discontinuous Accudenz gradient (Accurate Chemical and Scientific Corp.).

Preparation and digestion of protein fractions

Whole-cell proteome analyses of *P. falciparum* sporozoites (five independent preparations), gametocytes (three independent preparations), trophozoites (five independent preparations), and merozoites (four independent preparations) were described previously (Florens et al. 2002). In addition, two more merozoite (free of other stages), three schizont (95%–100% pure), seven schizont-free ring-stage, and three gamete preparations were lysed by osmotic shock in 10 mM Tris-HCl (pH 8.5) for 1 h on ice. The pellet fraction was separated from the supernatant by centrifugation for 30 min at 18,000g at 4°C. The membrane pellet was solubilized in 0.1 M sodium carbonate (pH 11.5). After 1 h at 4°C, the supernatant and sodium carbonate-extracted membrane pellet were separated by centrifugation. The pellet was resuspended again in 0.1 M sodium carbonate (pH 11.5). The three protein fractions (first supernatant, second supernatant, and membrane pellet) were (1) denatured in 8 M urea; (2) reduced in 5 mM Tris(2-Carboxyethyl)phosphine hydrochloride (TCEP, Roche); (3) alkylated by 20 mM iodoacetamide (IAM); and (4) digested with proteinase K (Roche) for 4 h at 37°C, in 0.1 M Sodium Carbonate (pH 11.5), as described in Wu et al. (2003).

MudPIT

As described previously (Washburn et al. 2001), peptide mixtures were concentrated and buffer exchanged to 5% Acetonitrile (ACN), 0.5% Acetic Acid on SPEC-PLUS PTC18 Cartridges (Ansys). They were then loaded onto a 100 µm inner-diameter fused-silica microcapillary column (Polymicro Technologies) with a 5 µm tip (Sutter Instruments P-2000 laser puller), packed first with 5 µm C$_{18}$ reverse phase (Aqua, Phenomenex), followed by 5 µm strong cation exchange material (Partisphere SCX, Whatman). Loaded microcapillary columns were installed in-line with a quaternary Agilent 1100 series HPLC pump. Fully automated six- or 12-step chromatography runs were carried out. The flow rate was set to 200–300 nL·min^{-1}. Three different elution buffers were used: Buffer A (5% ACN, 0.1% Formic Acid), Buffer B (80% ACN, 0.1% Formic Acid), and Buffer C (500 mM Ammonium Acetate, 5% ACN, 0.1% Formic Acid). The application of a 2.5 kV distal voltage electrosprayed the eluting peptides directly into a LCQ-Deca ion trap mass spectrometer equipped with a nano-LC electrospray ionization source (ThermoFinnigan). Full MS spectra were recorded on the peptides over a 400–1600 m/z range, followed by three tandem mass (MS/MS) events sequentially generated in a data-dependent manner on the first, second, and third most intense ions selected from the full MS spectrum (at 35% collision energy). Mass spectrometer scan functions and HPLC solvent gradients were controlled by the Xcalibur data system (ThermoFinnigan).

MS/MS data sets acquired for trophozoites, merozoites, schizonts, rings, gametocytes, and gametes were searched against a database combining host proteins (human, mouse, and rat sequences from NCBI RefSeq, www.ncbi.nlm.nih.gov/RefSeq/) with the latest release of the *Plasmodium falciparum* genome (Gardner et al. 2002) complemented with missing sequences of known *Plasmodium* proteins. Sporozoite data sets were searched against the same *P. falciparum* database combined with protein sequences from the *Anopheles gambiae* genome (Holt et al. 2002). The PEP_PROBE algorithm (Sadygov and Yates III 2003), a modified version of SEQUEST (Eng et al. 1994), was used to match MS/MS spectra to peptides. The outputs were parsed and filtered using DTASelect (Tabb et al. 2002). Spectra/peptide matches were retained only if they had a minimum cross-correlation score (XCorr) of 1.8 for singly charged, 2.5 for doubly charged, and 3.5 for triply charged spectra, and a normalized difference in correlation score (DeltCn) of at least 0.08. In addition, the confidence for the matches to be nonrandom had to be at least 85% as defined by PEP_PROBE.

Note, we only required the presence of a Lys or Arg residue at either end of the peptides to account for the fact that about half of the data set had been generated from insoluble fractions, in which case, the proteins had been first chemically cleaved by cyanogen bromide in formic acid treatment. In addition, it has been widely demonstrated that proteolytic digestion with trypsin often generates nontryptic peptides, due to chymotryptic impurities, chymotryptic activity of pseudotrypsin (generated by trypsin self-cleavage), in-source fragmentation within the mass spectrometer, or the wide diversity of proteases represented in the lysates of parasite extracts. Finally, no specific peptide ends were required for data sets generated from proteinase K-digested samples.

RT–PCR, and Northern blots

Total RNA was extracted from synchronized parasites at 0, 24, and 30 h after merozoite invasion.

RT–PCR was performed as previously described (Le Roch et al. 2002) using the following primers: 5′-GATAATATAAATGT TACTGAACAAGG-3′ and 5′-GGAACACTAATTTCGTTTTGTAC-3′ (EBA175); 5′-ATGGCAAGAACTAAACAAACAGC-3′ and 5′-TTATGATCTTTCTCCACGGATACG-3′ (histone H3); 5′-CAA ACCAATCTGGATCTGCAGAAG-3′ and 5′-CCATCTTGTGCTGA TAATAATTCATC-3′ (Pfcrt).

Northern analysis was performed according to the manufacturer's protocol (NorthernMax—Ambion). Briefly, 20 µg of total RNA was subjected to gel electrophoresis and blotted onto a Brightstar-plus Nylon membrane (Ambion). The blot was then probed with gene-specific probe PCR products obtained using the cDNA template and primer pairs described above. DNA probes were radiolabeled with [α-^{32}P]dATP, using the Prime-a

Gene labeling system (Promega). Unincorporated labeled nucleotides were removed by exclusion chromatography using sephadex G-50 columns (Amersham).

Western blots

Synchronized parasite cultures (0, 24, and 30 h postinvasion) were first centrifuged and washed with PBS (Invitrogen) before being homogenized in 0.15% saponin. Parasites were incubated on ice for 15 min and collected by centrifugation after several washes in PBS. Whole-cell lysates were obtained by suspension of the cell pellet in cytoplasmic lysis buffer (0.6% NP40, 0.15M NaCl, 10 mM Tris at pH 7.9, 1 mM EDTA) containing antiprotease cocktail (Complete Boehringer Mannheim), incubated on ice for 5 min, and centrifuged. The supernatant containing the cytoplasmic fraction was removed and pellets were resuspended in nuclear extraction buffer (10 nM HEPES at pH 7.9, 0.1 mM EGTA, 0.1 mM EDTA, 1.5 mM MgCl$_2$, 420 mM NaCl, 0.5 mM DTT, and 25% Glycerol) containing antiprotease cocktail (Complete Boehringer Mannheim), incubated on ice 20 min, and centrifuged. The supernatant containing the nuclear fraction was removed. Total protein concentrations of the cytoplasmic and nuclear fractions for each time point (ring, trophozoite, and schizont) were measured using the Redi Micro BCA protein assay system (Pierce). Equal protein amounts for each time point (~10 µg) were loaded onto a Novex pre-cast 10%–12% Bis-tris gel (Invitrogen) and electrophored in MES buffer (1 M MES, 1 M tris bas, 69.3 mM SDS, 20.5 mM EDTA) for 1 h at 110V. Cytoplasmic fractions were loaded for the EBA-175 and Pfcrt Western blots, nuclear fractions were loaded for Histone H3. Gels were transferred to a nitrocellulose membrane (Invitrogen). For EBA-175 and Pfcrt blots, membranes were incubated with rabbit antiserum (MRA-2) raised against purified recombinant GST fusion protein of *P. falciparum* EBA-175 (regionVI) and rabbit antiserum MRA 308, respectively, kindly provided by MR4 as primary antibody. For the histone H3 blot, membranes were incubated with rabbit anti-acetylated histone H3 (Upstate). Secondary antibody (goat anti-rabbit IgG, Sigma) binding was visualized using the PicoWest ECL Plus chemiluminescence detection system (Amersham).

mRNA abundance calculations

Transcriptome and proteome analyses were performed using *P. falciparum* clone 3D7 synchronized as previously described (Lambros and Vanderberg 1979); however, the analyses were carried out separately on different parasite populations. The number of genes/proteins to be compared excluded mitochondrially encoded genes, ribosomal RNA transcripts, and proteins without complimentary probes on the microarray (Table 1). The abundance of mRNA transcripts was calculated by applying the MOID algorithm for high-density oligonucleotide array analysis, which provides a P-value for each measurement, and thus, a metric to evaluate the confidence of each data point (Zhou and Abagyan 2002). Transcripts were considered to be present when their expression levels were greater than 10 and the log of the P-value (logP) associated with their MOID-calculated expression level was less than −0.5.

Statistical analysis and Spearman-rank correlation

The Spearman-rank correlation has been commonly used for comparisons between mRNA and protein data sets, as it provides a statistical method to compare mRNA and protein abundance for an entire data set while circumventing issues arising from the different scales by which mRNA and protein abundance are measured. To calculate the Spearman-rank correlation between a transcriptome and proteome data set, each data set was sorted by mRNA or protein abundance, and each value in the data set assigned a rank. Only genes having a minimum spectral count range of 10 across the seven proteome stages analyzed were included for Spearman analysis. The Spearman rank correlation coefficients were calculated using the following equation:

$$S_r = 1 - \frac{6\sum_{i=1}^{N}D_i^2}{N(N^2-1)}$$

where D is the difference between the ranks of corresponding mRNA and protein expression values for the ith gene, and N is the number of mRNA–protein value pairs (Lehmann 1975).

Scatterplot analysis

Scatterplots were constructed from a set of genes whose spectral count range was at least 10. This value was determined by assessing spectral count variation from previous MudPIT experiments, and is likely a conservative threshold, as spectral count differences below 10 have been shown to be significant (Liu et al. 2004). Genes were excluded if the log could not be calculated for either fold change in mRNA or protein abundance (i.e., if the fold change of mRNA or protein abundance was zero), or if the fold change in mRNA or protein abundance could not be calculated (i.e., the mRNA or protein abundance of the first stage in the transition was zero).

Gene-by-gene correlation analysis

The correlation coefficient (ρ) of each gene's mRNA and protein expression profile was calculated using the following equations:

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sigma_x \cdot \sigma_y}$$

and:

$$Cov(X,Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)$$

where x is the array of transcriptome values for each stage, y is the array of proteome values for each stage, σ_x is the standard deviation of x, σ_y is the standard deviation of y, n is the number of stages to be compared, μ_x is the mean of x, and μ_y is the mean of y. The value of ρ can have a range $-1 \le \rho \le 1$.

Sequence analysis of untranslated regions

For initial sequence alignments, sets of 20 genes were extracted from Supplemental Table 3, where their proteins were maximally detected in the same stage. The 1500 nucleotides downstream of the stop codon for each gene was extracted from the *P. falciparum* genome and each set submitted to the MEME sequence alignment algorithms (Bailey and Elkan 1994), with the top 30 motifs reported. The same downstream 1500 nucleotides were used to search for AU-rich elements (AUUUA), polyadenylation signals (AATAAA/ATTAAA) (Ruvolo et al. 1993), U-rich elements, or an NRE-like sequence (defined as GTTGTNNNATTGT, where the middle N sequence could range from 3 to 8 nucleotides).

Acknowledgments

References

Agabian, N. 1990. Trans splicing of nuclear pre-mRNAs. *Cell* **61:** 1157–1160.

Alano, P., Silvestrini, F., and Roca, L. 1996. Structure and polymorphism of the upstream region of the pfg27/25 gene, transcriptionally regulated in gametocytogenesis of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **79:** 207–217.

Anderson, L. and Seilhamer, J. 1997. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18:** 533–537.

Bailey, T.L. and Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2:** 28–36.

Beetham, J.K., Myung, K.S., McCoy, J.J., Wilson, M.E., and Donelson, J.E. 1997. Glycoprotein 46 mRNA abundance is post-transcriptionally regulated during development of *Leishmania chagasi* promastigotes to an infectious form. *J. Biol. Chem.* **272:** 17360–17366.

Beetham, J.K., Donelson, J.E., and Dahlin, R.R. 2003. Surface glycoprotein PSA (GP46) expression during short- and long-term culture of *Leishmania chagasi*. *Mol. Biochem. Parasitol.* **131:** 109–117.

Blumenthal, T., Evans, D., Link, C.D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W.L., Duke, K., Kiraly, M., et al. 2002. A global analysis of *Caenorhabditis elegans* operons. *Nature* **417:** 851–854.

Boeger, H., Griesenbeck, J., Strattan, J.S., and Kornberg, R.D. 2003. Nucleosomes unfold completely at a transcriptionally active promoter. *Mol. Cell* **11:** 1587–1598.

Bozdech, Z., Llinas, M., Pulliam, B.L., Wong, E.D., Zhu, J., and DeRisi. J.L., 2003. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1:** E5.

Burchmore, R.J. and Landfear, S.M. 1998. Differential regulation of multiple glucose transporter genes in *Leishmania mexicana*. *J. Biol. Chem.* **273:** 29118–29126.

Cheesman, S., Horrocks, P., Tosh, K., and Kilbey, B. 1998. Intraerythrocytic expression of topoisomerase II from *Plasmodium falciparum* is developmentally regulated. *Mol. Biochem. Parasitol.* **92:** 39–46.

Chen, C.Y., Xu, N., and Shyu, A.B. 1995. mRNA decay mediated by two distinct AU-rich elements from c-fos and granulocyte-macrophage colony-stimulating factor transcripts: Different deadenylation kinetics and uncoupling from translation. *Mol. Cell. Biol.* **15:** 5777–5788.

Coulson, R.M., Hall, N., and Ouzounis, C.A. 2004. Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*. *Genome Res.* **14:** 1548–1554.

Cui, L., Fan, Q., and Li, J. 2002. The malaria parasite *Plasmodium falciparum* encodes members of the Puf RNA-binding protein family with conserved RNA binding activity. *Nucleic Acids Res.* **30:** 4607–4617.

Dechering, K.J., Thompson, J., Dodemont, H.J., Eling, W., and Konings, R.N. 1997. Developmentally regulated expression of pfs16, a marker for sexual differentiation of the human malaria parasite *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **89:** 235–244.

Durr, E., Yu, J., Krasinska, K.M., Carver, L.A., Yates, J.R., Testa, J.E., Oh, P., and Schnitzer, J.E. 2004. Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. *Nat. Biotechnol.* **22:** 985–992.

Eksi, S., Stump, A., Fanning, S.L., Shenouda, M.I., Fujioka, H., and Williamson, K.C. 2002. Targeting and sequestration of truncated Pfs230 in an intraerythrocytic compartment during *Plasmodium falciparum* gametocytogenesis. *Mol. Microbiol.* **44:** 1507–1516.

Eng, J.K., McCormack, A.L., and Yates III, J.R. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass. Spectrom.* **5:** 976–989.

Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L., et al. 2002. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419:** 520–526.

Florens, L., Liu, X., Wang, Y., Yang, S., Schwartz, O., Peglar, M., Carucci, D.J., Yates, J.R., and Wu, Y. 2004. Proteomics approach reveals novel

proteins on the surface of malaria-infected erythrocytes. *Mol. Biochem. Parasitol.* **135:** 1–11.

Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S., and Garrels, J.I. 1999. A sampling of the yeast proteome. *Mol. Cell. Biol.* **19:** 7357–7368.

Gao, J., Opiteck, G.J., Friedrichs, M.S., Dongre, A.R., and Hefta, S.A. 2003. Changes in the protein expression of yeast as a function of carbon source. *J. Proteome Res.* **2:** 643–649.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., et al. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419:** 498–511.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. 2003. Global analysis of protein expression in yeast. *Nature* **425:** 737–741.

Golightly, L.M., Mbacham, W., Daily, J., and Wirth, D.F. 2000. 3′ UTR elements enhance expression of Pgs28, an ookinete protein of *Plasmodium gallinaceum*. *Mol. Biochem. Parasitol.* **105:** 61–70.

Graeser, R., Wernli, B., Franklin, R.M., and Kappes, B. 1996. *Plasmodium falciparum* protein kinase 5 and the malarial nuclear division cycles. *Mol. Biochem. Parasitol.* **82:** 37–49.

Griffin, T.J., Gygi, S.P., Ideker, T., Rist, B., Eng, J., Hood, L., and Aebersold, R. 2002. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* **1:** 323–333.

Guevara Patino, J.A., Holder, A.A., McBride, J.S., and Blackman, M.J. 1997. Antibodies that inhibit malaria merozoite surface protein-1 processing and erythrocyte invasion are blocked by naturally acquired human antibodies. *J. Exp. Med.* **186:** 1689–1699.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. 1999. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19:** 1720–1730.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., et al. 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298:** 129–149.

Horrocks, P., Jackson, M., Cheesman, S., White, J.H., and Kilbey, B.J. 1996. Stage specific expression of proliferating cell nuclear antigen and DNA polymerase delta from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **79:** 177–182.

Ifediba, T. and Vanderberg, J.P. 1981. Complete in vitro maturation of *Plasmodium falciparum* gametocytes. *Nature* **294:** 364–366.

Inselburg, J. and Banyal, H.S. 1984. Synthesis of DNA during the asexual cycle of *Plasmodium falciparum* in culture. *Mol. Biochem. Parasitol.* **10:** 79–87.

Jewett, T.J. and Sibley, L.D. 2003. Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. *Mol. Cell* **11:** 885–894.

Lambermon, M.H., Fu, Y., Wieczorek Kirk, D.A., Dupasquier, M., Filipowicz, W., and Lorkovic, Z.J. 2002. UBA1 and UBA2, two proteins that interact with UBP1, a multifunctional effector of pre-mRNA maturation in plants. *Mol. Cell. Biol.* **22:** 4346–4357.

Lambros, C. and Vanderberg, J.P. 1979. Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J. Parasitol.* **65:** 418–420.

Lanzer, M., de Bruin, D., and Ravetch, J.V. 1992. Transcription mapping of a 100 kb locus of *Plasmodium falciparum* identifies an intergenic region in which transcription terminates and reinitiates. *EMBO J.* **11:** 1949–1955.

Lanzer, M., Wertheimer, S.P., de Bruin, D., and Ravetch, J.V. 1993. *Plasmodium*: Control of gene expression in malaria parasites. *Exp. Parasitol.* **77:** 121–128.

Lanzer, M., de Bruin, D., Wertheimer, S.P., and Ravetch, J.V. 1994. Transcriptional and nucleosomal characterization of a subtelomeric gene cluster flanking a site of chromosomal rearrangements in *Plasmodium falciparum*. *Nucleic Acids Res.* **22:** 4176–4182.

Le Roch, K.G., Zhou, Y., Batalov, S., and Winzeler, E.A. 2002. Monitoring the chromosome 2 intraerythrocytic transcriptome of *Plasmodium falciparum* using oligonucleotide arrays. *Am. J. Trop. Med. Hyg.* **67:** 233–243.

Le Roch, K.G., Zhou, Y., Blair, P.L., Grainger, M., Moch, J.K., Haynes, J.D., De La Vega, P., Holder, A.A., Batalov, S., Carucci, D.J., et al. 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301:** 1503–1508.

Lehmann, E.L. 1975. *Nonparametrics: Statistical methods based on ranks*. Holden-Day, San Francisco, CA.

Levitt, A., Dimayuga, F.O., and Ruvolo, V.R. 1993. Analysis of malarial transcripts using cDNA-directed polymerase chain reaction. *J. Parasitol.* **79:** 653–662.

Li, J., Gutell, R.R., Damberger, S.H., Wirtz, R.A., Kissinger, J.C., Rogers, M.J., Sattabongkot, J., and McCutchan, T.F. 1997. Regulation and trafficking of three distinct 18 S ribosomal RNAs during

development of the malaria parasite. *J. Mol. Biol.* **269:** 203–213.

Liu, H., Sadygov, R.G., and Yates III, J.R. 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76:** 4193–4201.

Lobo, C.A. and Kumar, N. 1999. Differential transcription of histone genes in asexual and sexual stages of *Plasmodium falciparum. Int. J. Parasitol.* **29:** 1447–1449.

Martinez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P.J. 2003. Transcription of *Leishmania* major Friedlin chromosome 1 initiates in both directions within a single region. *Mol. Cell* **11:** 1291–1299.

McCutchan, T.F., Li, J., McConkey, G.A., Rogers, M., and Waters, A.P. 1995. The cytoplasmic RNAs of *Plasmodium spp. Parasitol. Today* **11:** 134–138.

Paton, M.G., Barker, G.C., Matsuoka, H., Ramesar, J., Janse, C.J., Waters, A.P., and Sinden, R.E. 1993. Structure and expression of a post-transcriptionally regulated malaria gene encoding a surface protein from the sexual stages of *Plasmodium berghei. Mol. Biochem. Parasitol.* **59:** 263–275.

Przyborski, J.M., Bartels, K., Lanzer, M., and Andrews, K.T. 2003. The histone H4 gene of *Plasmodium falciparum* is developmentally transcribed in asexual parasites. *Parasitol. Res.* **90:** 387–389.

Rehkopf, D.H., Gillespie, D.E., Harrell, M.I., and Feagin, J.E. 2000. Transcriptional mapping and RNA processing of the *Plasmodium falciparum* mitochondrial mRNAs. *Mol. Biochem. Parasitol.* **105:** 91–103.

Rice, J.C. and Allis, C.D. 2001. Histone methylation versus histone acetylation: New insights into epigenetic regulation. *Curr. Opin. Cell. Biol.* **13:** 263–273.

Ruiz i Altaba, A., Ozaki, L.S., Gwadz, R.W., and Godson, G.N. 1987. Organization and expression of the *Plasmodium knowlesi* circumsporozoite antigen gene. *Mol. Biochem. Parasitol.* **23:** 233–245.

Ruvolo, V., Altszuler, R., and Levitt, A. 1993. The transcript encoding the circumsporozoite antigen of *Plasmodium berghei* utilizes heterogeneous polyadenylation sites. *Mol. Biochem. Parasitol.* **57:** 137–150.

Sabet, N., Tong, F., Madigan, J.P., Volo, S., Smith, M.M., and Morse, R.H. 2003. Global and specific transcriptional repression by the histone H3 amino terminus in yeast. *Proc. Natl. Acad. Sci.* **100:** 4084–4089.

Sadygov, R.G. and Yates III, J.R. 2003. A hypergeometric probability model for protein identification and validation using tandem mass spectral data and protein sequence. *Anal. Chem.* **75:** 3792–3798.

Sam-Yellowe, T., Florens, L., Johnson, J.R., Wang, T., Drazba, J.A., Le Roch, K.G., Zhou, Y., Batalov, S., Carucci, D.J., Winzeler, E.A., et al. 2004. A *Plasmodium* gene family encoding Maurer's cleft membrane proteins: Structural properties and expression profiling. *Genome Res.* **14:** 1052–1059.

Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C.,

Pouvelle, B., Gysin, J., and Lanzer, M. 1998. Antigenic variation in malaria: In situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in *Plasmodium falciparum. EMBO J.* **17:** 5418–5426.

Soto, M., Quijada, L., Larreta, R., Iborra, S., Alonso, C., and Requena, J.M. 2003. Leishmania infantum possesses a complex family of histone H2A genes: Structural characterization and analysis of expression. *Parasitology* **127:** 95–105.

Tabb, D.L., McDonald, W.H., and Yates III, J.R. 2002. DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1:** 21–26.

Taylor, H.M., Grainger, M., and Holder, A.A. 2002. Variation in the expression of a *Plasmodium falciparum* protein family implicated in erythrocyte invasion. *Infect. Immun.* **70:** 5779–5789.

Vervenne, R.A., Dirks, R.W., Ramesar, J., Waters, A.P., and Janse, C.J. 1994. Differential expression in blood stages of the gene coding for the 21-kilodalton surface protein of ookinetes of *Plasmodium berghei* as detected by RNA in situ hybridisation. *Mol. Biochem. Parasitol.* **68:** 259–266.

Washburn, M.P., Wolters, D., and Yates III, J.R. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19:** 242–247.

Washburn, M.P., Koller, A., Oshiro, G., Ulaszek, R.R., Plouffe, D., Deciu, C., Winzeler, E., and Yates III, J.R. 2003. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae. Proc. Natl. Acad. Sci.* **100:** 3107–3112.

Waters, A.P., Syin, C., and McCutchan, T.F. 1989. Developmental regulation of stage-specific ribosome populations in *Plasmodium. Nature* **345:** 438–440.

Wu, C.C., MacCoss, M.J., Howell, K.E., and Yates, J.R. 2003. A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.* **21:** 532–538.

Zamore, P.D., Williamson, J.R., and Lehmann, R. 1997. The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. *RNA* **3:** 1421–1433.

Zhou, Y. and Abagyan, R. 2002. Match-only integral distribution (MOID) algorithm for high-density oligonucleotide array analysis. *BMC Bioinform.* **3:** 3.

Web site references

www.ncbi.nlm.nih.gov/RefSeq/; NCBI RefSeq.

Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle

Karine G. Le Roch, Jeffrey R. Johnson, Laurence Florens, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2004/10/06/14.11.2308.DC1
References	This article cites 67 articles, 18 of which can be accessed free at: http://genome.cshlp.org/content/14/11/2308.full.html#ref-list-1
License	
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions