

Oxidative Stress Genes in *Plasmodium falciparum* as Indicated by Temporal Gene Expression

R. Guerra^{1*}, C. Shaw³, B. Christian¹, G. Fox¹, J. Noyola-Martinez¹, M. Stevens^{1,2}, N. Garg⁴, M. Gustin²

¹Department of Statistics, ²Department of Biochemistry and Cell Biology, Rice University

³Department of Molecular and Human Genetics, Baylor College of Medicine

⁴Department of Microbiology, University of Texas Medical Branch

* Corresponding author: rguerra@rice.edu

ABSTRACT

Entry of *Plasmodium falciparum* into human red blood cells is a stressful event for both the host and the parasite. Conversion of hemoglobin into usable food by *P. falciparum* is accompanied by the production of chemically reactive and toxic molecules called oxidants. Examination of the temporal sequence of gene expression during the intraerythrocytic development cycle (IDC) [4] can help elucidate how *Plasmodium* responds to these self-generated harmful chemicals while proceeding through its normal developmental program. Our study has three parts: identification of temporally-defined sets of co-regulated oxidative stress response genes in this parasite; comparison of the temporal patterns of the oxidative stress response to that of co-regulated gene sets involved in other processes; and identification of putative transcription factor binding sites by finding DNA motifs unique to the upstream regions of co-regulated oxidative stress response genes.

Keywords: Malaria; oxidative stress; microarray; time-course; clustering; motif.

1. INTRODUCTION

Plasmodium falciparum is a virulent pathogen that is the major cause of the human malaria epidemic seen in developing tropical countries. *P. falciparum* utilizes mosquitoes as a vector to enter the host body and invade red blood cells. Once inside the erythrocytes, *Plasmodium* will replicate, utilizing the host's hemoglobin as a source of amino acids. After replication, the red blood cells are lysed, releasing parasites, toxic heme groups and parasitic proteins into the host's blood plasma causing the characteristic fever. The parasites will then re-enter new red blood cells and the cycle will continue, or they will be picked up by feeding mosquitoes, and will go on to infect another host.

Because of its importance as a possible Achilles heel, the *P. falciparum* oxidative stress response (OSR) has been heavily studied and many of the key proteins and enzymes have been identified.

After attachment and entry into the red blood cell, the parasite

uses its food vacuole to engulf the concentrated hemoglobin and break the protein down into usable amino acids and leftover heme. Most of the heme forms an inert pigmented polymer inside the vacuole. However, a small amount of free heme becomes a major iron-based catalyst for formation of superoxide and other oxidants such as hydrogen peroxide. These highly reactive oxidants form covalent bonds with proteins, nucleic acids, and lipids, thereby impairing their function. The parasite defends itself against these oxidants with enzymes that convert the oxidants into less reactive chemicals and with enzymes that repair the damaged cell molecules. Without the function of these enzymes, the genome and lipid membranes of *P. falciparum* are vulnerable to devastating oxidative damage. Although oxidative stress proteins have been used as a drug target for many years, the expression of the genes for these proteins has not been studied. Understanding the transcriptional activity of genes that respond to oxidative stress may be crucial to developing new drugs and gaining a better understanding of how current drugs function.

In microbes such as yeast or *E. coli*, exposure to oxidants produces a well-characterized whole genome transcriptional response [6,7,8,10,11]. Genes induced by oxidants include those that encode oxidant-scavenging proteins and enzymes that repair oxidant-damaged proteins, DNA, and lipid. But most of the yeast genes modulated by oxidants such as hydrogen peroxide are also regulated in the same fashion by other stressful conditions such as heat, starvation, or an increase in osmolarity caused by high sugar concentrations. This common response, termed the environmental stress response or ESR, is characterized by increase in mRNA for stress response genes and reduction in mRNA for genes involved in nucleic acid and protein synthesis, i.e., cell growth and division. Another striking feature of the oxidative stress response is that the increase in ESR gene mRNA is transient, falling back to control levels in less than one hour after adding hydrogen peroxide to cells. This decrease in mRNA occurs even though the stimulus, the hydrogen peroxide concentration, stays constant for a 2-hour period (Gasch, Spellman et al. 2000).

A potentially important difference between *Plasmodium* and other microbes concerns the timing of oxidative stress in the normal life cycle. For yeast and bacteria, a low level of oxidants is produced during normal oxygen-dependent metabolism. Immune cells and even plants attack invading microbes by producing large quantities of superoxide and hydrogen peroxide. However, the timing of this oxidative assault is highly variable and genome responses are triggered by the oxidant per se. On the other hand, for an invading *Plasmodium*, production of large amounts of oxidants (from the breakdown of hemoglobin) would occur at a fairly predictable interval after entry of the parasite.

Leave this section blank for copyright statements, distribution, etc.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

Conference '04, Month 1-2, 2004, City, State, Country.

Copyright 2004 ACM 1-58113-000-0/00/0004...\$5.00.

Thus, there could be two strategies in mounting an effective defense. Either the parasite is like yeast in using the oxidant as the direct inducer of oxidative stress responses of the genome. Or, the protozoan would induce oxidative stress response genes as part of its developmental cycle.

From our initial analysis, we propose here a novel biological mechanism for dealing with stress. Our model is that *P. falciparum* does not initiate a general stress response during oxidative stress, but rather their stress reaction is specific and anticipatory. That is, the parasite will initiate transcription of oxidative stress response genes such as glutathione transferase and peroxiredoxin prior to the onset of that stress. This model is more probable for *P. falciparum* than it is for yeast since *Plasmodium* makes use of host hemoglobin proteins, even at the expense of having to dispose of self-generated toxic heme groups. As this metabolic process is an inherent part of the *P. falciparum* life cycle, the strategy of anticipating the accompanying oxidative stress may have a selective advantage over reacting to stress. One prediction of this model is that the increase in oxidative stress response gene expression would be co-regulated with genes involved in normal developmental processes such as making protein or RNA.

Whatever the strategy employed, the physiological situation in parasite-infected red blood cells appears very different from yeast exposed to oxidants. Oxidants cause yeast to stop making protein and RNA, to stop growing and dividing while it adapts to stress. *Plasmodium* is initiating large-scale protein and nucleic acid synthesis – using red blood cell nutrients – during oxidative stress. Expression of oxidative stress response genes continues for many hours, rather than falling as they do in yeast. This difference in timing suggests very different control mechanisms in the parasite, which may be critical to drug development.

Analytical Objective: Identifying genes co-regulated with known oxidative stress response (OSR) genes, as well as DNA sequence elements common to the upstream regions of OSR genes will be an important step forward in defining the pathways and signals that control this response. To this end, we report on a statistical analysis that identifies several promising genes and motifs that may underlie the oxidative stress mechanism in *P. falciparum*. A novel proposal is made for (1) defining selected features of time-course profiles for classification and (2) defining control groups for subsequent motif discovery.

2. Methods

2.1 Study Design and Data

A large-scale culture of *P. falciparum* (HB3 strain) was grown for RNA sample isolation, cDNA synthesis, labeling, and DNA hybridization with a long-oligonucleotide microarray [4]. Samples for 48 individual (hourly) time points (Cy5) corresponding to the intraerythrocytic development cycle were hybridized against a reference pool (Cy3) comprised of RNA samples representing all developmental stages of *P. falciparum*. In the present analysis we used the Quality Control (QC) dataset that included the set of oligonucleotides that passed all quality control filters as specified by [xxx] and were normalized by a linear scalar (global normalization).

2.2 Statistical Methods

Gene profiles represented by multiple oligos were summarized by averaging pointwise the individual oligos. The QC dataset was missing time-points 23 and 29hr. Imputation for these time points, as well as for randomly missing individual points, was by averaging flanking time points. The two major analytical steps were (1) finding expression profiles similar to the OSR genes and (2) finding sequence motifs in the upstream region of all ORF's belonging to a cluster set.

Step 1: Classifying profiles by OSR genes. The objective of this step is to find genes that are potentially co-regulated with OSR genes. Having done so allows further consideration of common function and/or motifs. Two approaches were followed to find profiles similar to OSR genes. The first uses the whole profiles in determining similarity between OSR expression profiles and other genes. The second uses segments of entire profiles to capture certain features that may be unique to OSR genes. In the latter case, interest centers on “early expressing” OSR profiles that show an increasing trend at the start (hours 1-10) of the profile (Figures 1, glutathione transferase and peroxiredoxin). This will allow us to focus on the expression profiles that are similar to an OSR gene over a particular part of the IDC. We anticipate that this will increase our power to detect useful motifs (below) operating in a relatively small window of the IDC. Indeed, the IDC itself is divided into three stages: Ring, Trophozoite, Schizont. The first 10 hours of the time-course profile are within the Ring stage. Given a reference profile (e.g., one of the OSR genes) $\mathbf{x}=(x_1, x_2, \dots, x_T)$ we classify all other profiles $\mathbf{y}=(y_1, y_2, \dots, y_T)$ as “close” to \mathbf{x} if the correlation between \mathbf{x} and \mathbf{y} is at least 0.9. The number of points entering the correlation calculation depends on whether we use entire profiles or partial profiles. Alternative distance measures, such as Euclidean distance, are possible.

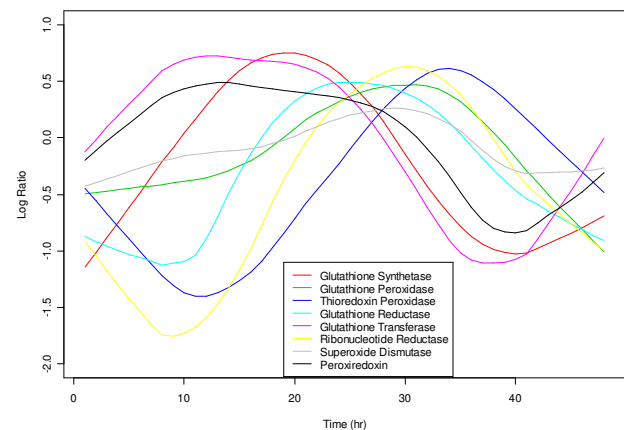


Figure 1: Oxidative stress genes.

Step 2: Motif searching within clusters. Given gene clusters defined by the OSR genes we next search for common motifs within each cluster. The software program MEME [1] was used to discover motifs within each group. We restricted the MEME motif search to 6-8 nt. To assess the biological significance of motif findings we selected comparison groups in which we would use MetaMEME [9] to determine the occurrence frequency of the discovered motifs. If the comparison group showed similar

matches we discarded the discovered motifs. Comparisons were conducted based on the following groups of genes:

Early-Entire: $n=147$ genes matched ($r > 0.9$) to the entire average profile of the two early OSR genes (Figure 1, glutathione transferase and peroxiredoxin).

Late-Entire: $n=289$ genes matched ($r > 0.9$) to the entire average profile of the two late OSR genes (Figure 1, ribonucleotide reductase and thioredoxin peroxidase).

Early-10: $n=62$ genes matched ($r > 0.95$) to the first 10 hours of the average profile of the two early OSR genes (Figure 1, glutathione transferase and peroxiredoxin).

Late-10: $n=95$ genes matched ($r > 0.95$) to the first 10 hours of the average profile of the two late OSR genes (Figure 1, ribonucleotide reductase and thioredoxin peroxidase). Note that over the first 10 hours of the IDC the Early genes monotonically increase in relative expression, while the Late genes decrease in relative expression. Therefore, we surmised, that different motifs would be acting on these genes.

To address the issue of whether the parasite responds or anticipates oxidative stress we considered expression patterns of genes involved with hemoglobin (Hb) degradation, hemozoin sequestration/degradation and FP-binding proteins. Indeed, since Hb is the major source of oxidative stress, the expression pattern of Hb degradation-related genes in relation to antioxidant stress response genes would help confirm if the antioxidants are up-regulated in response to stress or it is an anticipated response. If the OSR genes are up-regulated before the Hb digestion related genes that would indicate anticipation of oxidative stress; up-regulation after expression of the Hb genes would signify a response to the oxidative stress.

3. Results

3.1 Expression of Oxidative Stress Genes

P. falciparum protects itself against oxidants [2] by releasing (1) oxidant-consuming enzymes, including superoxide dismutase, peroxiredoxin, glutathione S-transferase and (2) enzymes involved in the repair of oxidized proteins and lipids reductant synthesis, including thioredoxin peroxidase reductase, glutathione reductase, and glutathione synthetase. Figure 1 shows the time-course expression profiles for these eight oxidative stress response OSR genes. The same data have also recently been used by [Bozdech and Ginsburg] to examine the expression profiles of these and other genes in antioxidant defense in *P. falciparum*. Using the three stages of the IDC the genes may be classified according to where they (approximately peak): *Ring*: glutathione S-transferase, peroxiredoxin; *Ring/Trophozoite*: glutathione synthetase; *Trophozoite*: glutathione reductase, superoxide dismutase; *Trophozoite/Schizont*: ribonucleotide reductase, glutathione peroxidase; *Schizont*: thioredoxin peroxidase

The parasite begins significant digestion of the host hemoglobin (Hb) at the boundary between the Ring and Trophozoite stages, roughly between 12 and 18 hours into the 48-hour IDC.

Therefore, relative to the digestion of Hb it may be conjectured that glutathione S-transferase and peroxiredoxin, involved in H_2O_2 dismutation, are anticipating the endogenous generation of oxidants. Glutathione synthase, involved in reduced glutathione (GSH) synthesis, may be activated in response to utilization of GSH in initial response. Glutathione reductase, ribonucleotide

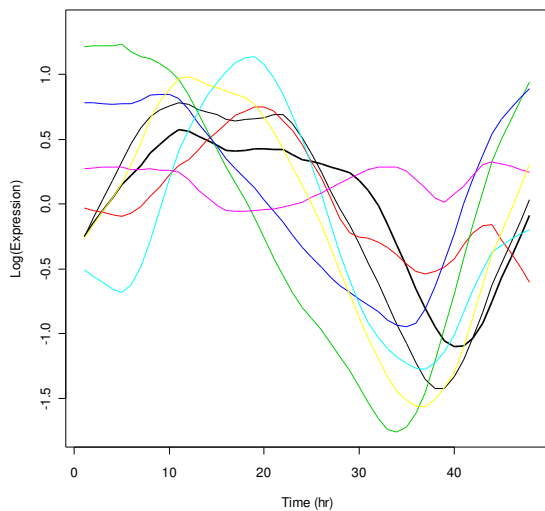
reductase, and thioredoxin reductase react to the stress of accumulation of oxidized glutathione (GSSG) and thioredoxin ($TrxS_2$) and thus glutathione reductase may be reacting to, instead of anticipating, the oxidative stress as indicated by their sharp increase in expression coincidental with intense hemoglobin digestion. Glutathione peroxidase and catalase (both dismutate H_2O_2) may then constitute the secondary defense mechanism against oxidative stress. The remaining genes are a bit more ambiguous.

Additional information pertaining to the anticipation/reaction conjecture may be obtained by considering the expression profiles of Hb degradation-related genes. Indeed, since Hb degradation is the major source of oxidative stress the expression pattern of OSR genes in relation to those of Hb-related genes will help confirm if the antioxidants are up-regulated in response to stress or if it is an anticipated response. Here we consider vacuolar proteins which maintain the acidic environment within the food vacuole and proteases which break down hemoglobin. Figures 2 and 3 show vacuolar proteins and proteases, respectively, in relation to the Early OSR genes. The vacuolar genes are generally increasing over the Ring Stage but at a lower level than the two Early OSR genes, again supporting the hypothesis that they are anticipatory. Three of the proteases (falcipain 2 precursor, falcipain 2 precursor putative, plasmepsin 2 precursor) increase in expression over the Ring stage but the curves appear to follow the OSR genes, which also have higher levels of expression. Two proteases (plasmepsin putative, plasmepsin 1 precursor) show no change in expression and then begin to decrease at about the time that the OSR genes peak, which again seems to support a readiness by the OSR genes. All but one (falcipain 3) protease drops precipitously beginning at ~18hr through ~32hr. The remaining six OSR genes peak after both the vacuolar and protease genes indicating a reaction to the oxidative stress. On balance, the data support that glutathione S-transferase and peroxiredoxin anticipate oxidative stress as they peak before HB digestion and they peak or express themselves at higher levels relative to Hb-related genes.

3.2 Motif Discovery

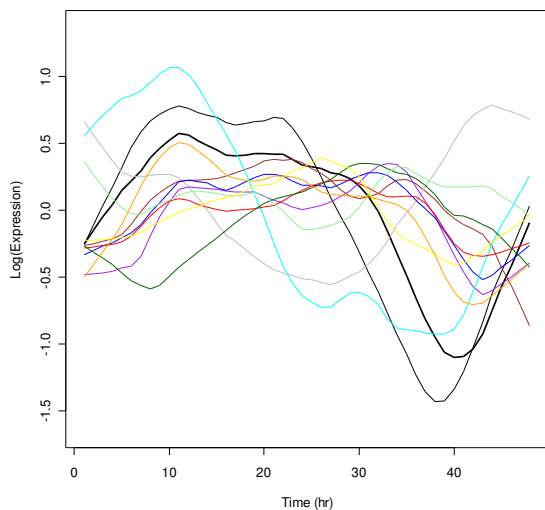
To further characterize the oxidative stress response genes we conducted a motif search of the 1000bp upstream sequence of gene groups defined by OSR genes. The "Early" group was

defined by the $n=157$ genes that were correlated at a level of



Protease and Early Oxidative Stress Genes

Proteases - Plasmepsin 2 precursor = red; Plasmepsin 1 precursor = green; Plasmepsin putative = dark blue; Falcipain 2 precursor = light blue; Falcipain 3 = pink; Falcipain 2 precursor putative = yellow; **OSR** - Glutathione Transferase = light black; Peroxiredoxin = heavy black



Vacuolar and Early Oxidative Stress Genes

Vacuolar - V. ATP Synthase = brown; V. Proton Translocation ATPase = gray; V. ATP Synthase Subunit F = purple; V. ATP Synthase Catalytic Subunit A = blue; V. ATP Synthase Subunit H = light blue; V. ATP Synthase Subunit D = red; V. ATP Synthase Subunit B = dark green; V. ATP Synthetase = light green; V. ATP Synthase Subunit E = yellow; V. ATP Synthase Subunit G = orange; **OSR** - Glutathione Transferase = light black; Peroxiredoxin = heavy black

Pearson's $r > 0.9$ with the average profile of the two Early OSR genes, glutathione S-transferase, peroxiredoxin. A "Late" group of $n=289$ genes was similarly defined for the two Late OSR genes, thioredoxin peroxidase and ribonucleotide reductase. Since no *a priori* information concerning expected motifs was available, the

two groups served as each other's control in evaluating the biological significance of a putative motif. Thus, the E-value from MEME was used as the initial screen for putative motifs and the control group was used as a confirmation that a given motif was specific to a set of genes with similar expression profiles. The confirmation in the control group was carried out by MetaMEME.

Table 1: Top 10 consensus sequences found by MEME in 157 genes correlated with early¹ expression OSR genes. Comparison group comprised of 289 genes correlated with late¹ expression OSR genes. Entry is percent of group members that match the motif.

Consensus Motif	E-value Score	MEME Early	MetaMEME	
			Early	Late
CCCCCA	9.6E-39	30%	29%	26%
GGGGGA	2.9E-37	32	37	35
CCCTTT	8.6E-10	25	29	29
CCTTCC	2.4E-7	29	38	31
AGCACA	5.7E-5	32	38	34
CGAGGAGG ₂	5.6E-3	11	14	6
TCCTTC	1.5E-1	16	17	14
TCAAACCC	5.4	6	8	4
AAAAGGAA	4.9E-2	25	22	28
CAAGAAGC	1	13	31	25

¹See Figure 1 for Early and Late expression OSR genes.

²p-value = 0.005 for proportion difference between early and late groups.

Table 2: Top 10 consensus sequences found by MEME in 289 genes correlated with late¹ expression OSR genes. Comparison group comprised of 157 genes correlated with early¹ expression OSR genes. Entry is percent of group members that match the motif.

Consensus Motif	E-value Score	MetaMEME	
		Late	Early
CCCCC	9.6E-47	23%	21%
GTGTGC ²	1.5E-24	16	4
GGGGAA	2.6E-20	12	14
ACACAC	1.7E-09	27	3
ACATACAC ₂	8.2E-14	49	18
CCCTTTT	1.3E-19	2	1
GGAGGG	6.2E-06	18	15
GTGTGT ²	2.4E-05	18	1
CACACA ²	5.9E-04	29	6
AGAAAAGG	1.0E-07	39	27

¹See Figure 1 for Early and Late expression OSR genes.

²p-value < 1.0E-05 for proportion difference between early and late groups.

Table 1 shows the best ten motifs in the Early group as defined by the E-value. To assess the match scoring by MetaMEME we applied it first to the Early group to assess how well MetaMEME would find the motifs that MEME claimed to have found; this was

our positive control. A comparison of columns 3 and 4 in Table 1 shows that MetaMEME is adequately scoring the motifs. Column 5 shows the percentage of $n=289$ Late genes that included the given motif in Column 1. All except motif **CGAGGAGG** had similar frequencies in the Early and Late groups, indicating that most motifs were not biologically significant in that they were not (fairly) unique to the early expression of these two OSR genes. This significant motif is associated with cell growth in eukaryotes. Table 2 shows analogous results for the Late group motifs. Here we find four of ten motifs that are statistically different in their frequencies between Early and Late groups; they all have p-values less than 10^{-5} . Current efforts are in progress to assess the biological significance of these motifs. Similarly, we find (data not shown) statistically significant differences in motif frequencies between the Early and Late groups restricted to the profiles in the Ring stage (1-10hr).

4. Discussion

When *P. falciparum* invades red blood cells it is subject to constant oxidative stress largely stemming from its digestion of host hemoglobin, as well as from reactive oxygen and nitrogen species arising from the host immune system. Using the time-course microarray data of [3] the work reported in this paper examined and characterized the gene expression profiles of eight oxidative stress response genes, the detection of other genes possibly co-regulated with these genes, and discovery of sequence motifs that may play a role in their regulation.

Our proposed model is that control of expression of oxidative stress response genes in *P. falciparum* is mediated not by an induced response to oxidants, but as a part of the normal developmental sequence important for asexual reproduction of the parasite in red blood cells. The analysis so far supports this model. The two OSR genes for peroxiredoxin and glutathione transferase increase in expression relatively early in the infection cycle, before expression of many of the enzymes needed for hemoglobin digestion. These two genes are co-expressed with genes needed for mRNA synthesis and translation. Thus, the shared function of this gene set is clearly related to the products of Hb breakdown: amino acids and oxidants. One wrinkle is that not all of the OSR genes are expressed in this early set. For example, mRNAs for glutathione peroxidase and glutathione reductase appear later in the infection cycle. One possible explanation for different waves of OSR gene expression is that different oxidants may appear at different times. The early appearance of both peroxiredoxin and glutathione transferase mRNAs is likely because these two enzymes are the major route for elimination of hydrogen peroxide, the most abundant oxidant appearing during Hb digestion in the Plasmodium food vacuole.

Our initial analysis of the upstream regions of the Early and Late OSR gene sets has revealed several closely related MOTIFs that are significantly unique to the latter set, e.g., CACACA, ACACAC, ACATACAC, all including CA pairs. The biological significance is under study, but could be a binding site for a repressor that prevents early gene expression or an activator that positively regulates gene expression at later times in the infection cycle. Our motif analysis is preliminary, using a standard set of search parameters and only one program (MEME). Additional motif analysis using MEME and programs based on different search algorithms will be required for a comprehensive identification of the motifs or combinations of motifs that are

uniquely correlated with the different OSR gene sets. The results of our study provide novel insights into the biology of *P. falciparum*. We present an approach based on biological and statistical reasoning that together lead to promising areas of inquiry. Key to motif assessment in this and similar applications are (1) careful selection of classification features from the time-course profiles (e.g., Ring stage segment), which in turn should lead to more meaningful motif discovery by also using the selected feature for (2) defining control groups.

5. REFERENCES

- [1] Bailey, T.L and Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, Menlo Park, California, 1994, 28-36.
- [2] Becker K., Tilley L., Vennerstrom J.L., Roberts D., Rogerson S., Ginsburg H. Oxidative stress in malaria parasite-infected erythrocytes: host-parasite interactions. *International Journal for Parasitology*, 34 (2004), 163-89.
- [3] Bozdech, Z. and Ginsburg, H. Antioxidant defence in Plasmodium falciparum: data mining of the transcriptome. *Malaria Journal*, 3:23 (9 July 2004).
- [4] Bozdech, Z., Llinás, M., Pulliam, B.L., Wong, E.D., Zhu, J., DeRisi, J.L. The transcriptome of the intraerythrocytic development cycle of *Plasmodium falciparum*. *PLoS Biology*, 1, 1(2003), 1-16.
- [5] Carmel-Harel O. and Storz G. Roles of the glutathione- and thioredoxin-dependent reduction systems in Escherichia coli and Saccharomyces cerevisiae responses to oxidative stress. *Annu. Rev. Microbiol.* 54 (2000), 439-61.
- [6] Causton, H. C., B. Ren, et al. Remodeling of yeast genome expression in response to environmental changes. *Mol Biol Cell*, 12, 2, (2001), 323-37
- [7] Chen, D., W. M. Toone, et al. Global transcriptional responses of fission yeast to environmental stress. *Mol Biol Cell* 14, 1(2003), 214-29.
- [8] Gasch A.P., Spellman P.T., Kao C.M., Carmel-Harel O., Eisen M.B., Storz G, Botstein, D, Brown PO. Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*, 11, (2000), 4241-57.
- [9] William N. Grundy, Timothy L. Bailey, Charles P. Elkan and Michael E. Baker. Meta-MEME: Motif-based hidden Markov models of protein families. *Computer Applications in the Biosciences*, 13, 4 (1997), 397-406.
- [10] Pomposiello, P. J. and B. Dimple. Global adjustment of microbial physiology during free radical stress. *Adv Microb Physiol*, 46, (2002), 319-41
- [11] Zheng, M., X. Wang, et al. DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide, *J Bacteriol*, 183, 15(2001), 4562-70.