Metabolomics, peptidomics and glycoproteomics studies on human schistosomiasis mansoni

Balog, C.I.A.

METABONOMIC INVESTIGATION OF HUMAN *SCHISTOSOMA MANSONI* INFECTION

4

ABSTRACT

Schistosomiasis is a parasitic infection that is endemic in many developing countries in the tropics and subtropics afflicting more than 200 million people primarily in rural areas. After malaria, it is the second most important parasitic infection in terms of socio-economic and public health importance. Investigation of the host-parasite interaction at the molecular level and identification of biomarkers of infection and infection-related morbidity would be of value for improved strategies for treatment and morbidity control. To this end, we conducted a Nuclear Magnetic Resonance-based metabonomics study involving a well-characterized cohort of 447 individuals from a rural area in Uganda near Lake Victoria with a high prevalence of *Schistosoma mansoni*, a species predominantly occurring in Africa, Madagascar and parts of South America. From the individuals included in the cohort urine samples were collected at five time-points, before and after (one or two times) chemotherapy with praziquantel. Using a combination of unsupervised and supervised multivariate statistical analysis of the 1-D NMR spectra generated, we were able to discriminate infected from uninfected individuals in two age groups (children and adults) based on differences in their urinary profiles. The underlying molecular markers of *S. mansoni* infection were found to be primarily linked to changes in gut microflora, energy metabolism and liver function. These findings are in agreement with data from earlier studies on *S. mansoni* infection in experimental animals and thus provide corroborating evidence for the existence of metabolic response specific for this infection.

Crina I.A. Balog[1], Axel Meissner[1], Sibel Göraler[1], Marco R. Bladergroen[1], Birgitte J. Vennervald[2], Oleg A. Mayboroda[1] and André M. Deelder[1]

[1] Biomolecular Mass Spectrometry Unit, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, The Netherlands [2] DBL Centre for Health Research and Development, Faculty of Life Sciences, University of Copenhagen, Denmark

Submitted

INTRODUCTION

Schistosomiasis is, after malaria, the second-most socio-economically devastating parasitic disease. An estimated 600 million people in 74 countries are at risk of infection, with over 200 million chronically infected (World Health Organization, www.who.int). Individuals with persistent or heavy infections may show activity of severe clinical manifestations such as hypertension, hepatomegaly, splenomegaly, ascites, haematemesis, oedema and per-umbilical varices (1-3). Traditionally, schistosomiasis is diagnosed by detecting parasite eggs in faeces and urine (4;5). This microscopy-based method is very specific, simple and cheap but due to the uneven distribution of *Schistosoma mansoni* eggs in solid excreta and the considerable day-to-day fluctuation in egg output, infections, and especially low infections, are easily missed (6). Alternatively, detection of antibodies is a highly sensitive and specific method to diagnose schistosomiasis. Serology mostly gives straightforward answers for patients tested within months after their first exposure, but data are difficult to interpret for those who have a history of previous infection. Detection of circulating antigens by immunological means is an effective strategy to distinguish active from past infections. Two major intestinal polysaccharide antigens have been described: a negatively charged circulating anodic antigen (CAA) and a positively charged circulating cathodic antigen (CCA) (7;8). Sensitive and specific immunoassays using monoclonal antibodies have been developed for the detection and quantification of these two antigens in serum and urine of patients (9-12). Although these methods work well in endemic areas with moderate and high prevalence of the disease, they are less sensitive in low infection areas (11;12). These shortcomings indicate the need for new and improved tools that can be utilized for sensitive and specific diagnosis of schistosomiasis. However, there is also an obvious need for development better tools for monitoring morbidity, understanding the host-parasite interaction, getting new insights into immunity and treatment and control. Metabonomics, as one of the well established "omics" disciplines ,is being applied increasingly to investigate diseases and mechanisms of diseases in humans. In the clinical field, the exploitation of the Nuclear Magnetic Resonance-generated metabolic data sets in combination with multivariate statistical analysis allows sample classification, effective interpretation of the biochemical pathways and has demonstrated potential for identifying candidate biomarkers. Several encouraging results have been obtained using metabolite profiling, including studies on heart disease (13), type 2 diabetes (14), cancer (15), and nervous system diseases such as schizophrenia (16), amyotrophic lateral sclerosis (17), Huntington's disease (18) and Parkinson's disease (19).

Since in all parasitic infections there is a significant metabolic interaction between pathogen and host, systematic NMR metabolic fingerprints have been applied to investigate the metabolic consequences of parasitic worm infections in rodent models

(20-25). The metabolic analysis (by NMR) in mice infected with *S. mansoni* determined a metabolic signature of the infection that indicated the feasibility of using an "omics"-based approach for accurate diagnosis (24). Likewise, using capillary electrophoresis (CE) a characteristic metabolite fingerprint of *S. mansoni* infection in mouse urine was identified (26). Recently, four parasite-rodent models, namely *Plasmodium berghei*-mouse, *Trypanosoma brucei brucei*-mouse, *S. mansoni*-mouse and *Fasciola hepatica*-rat models were used to investigate a multiplex panel of cytokines and metabolites fingerprints and the authors showed that each parasite generated a unique metabolic signature in the host (27).

Although, altered metabolic activities due to parasitic infections have been observed in rodent models, little is known about metabolic biomarkers for infections in human. In comparison with animal studies, our study, based on urine samples collected in an area endemic for *S. mansoni* infection is more intricate due to variable infection intensities and differences in the time course of the infection. Moreover, the inter- and intra-individual biological variability, different food patterns, co-morbidities with other infections, differences in age and gender are features influencing classification procedures. Nevertheless, using a MALDI-ToF peptidomics profiling based approach for the same cohort, high recognition rates for the detection of *S. mansoni* infection in children were obtained(28).

In this paper we continue this line of research and using an NMR platform as a first step towards understanding the metabolic consequences of *S. mansoni* infection in human, we investigated the metabolic effects of this parasite infection in a large cohort from an rural area in Uganda near lake Victoria endemic for *S. mansoni* and made an attempt to give a physiological interpretation of our findings.

MATERIAL AND METHODS

Study design and sample collection

This study took place in Musoli village, Mayuge district, Uganda, an area situated on the shore of Lake Victoria with perennial transmission of both *S. mansoni* and malaria(28). Briefly, a cohort of 446 volunteers was selected using a stratified random selection balanced for sex and age. Only individuals of 7 years and older were included in the cohort. From each member of the cohort, three stool samples were collected on consecutive days and these were examined for the presence of *S. mansoni* ova. The presence of each infection was recorded as eggs per gram faeces (EPG) using two Kato-Katz thick smears per stool sample (4). The study cohort received two single doses of praziquantel (PZQ) 40 mg/kg of body weight with a two weeks interval. Urine samples, collected at five time points, were used in the current study: before treatment (time point A), one day after the first dose of PZQ (time point B), 2 weeks after the first

dose of PZQ (time point C), one day after the second dose of PZQ (time point D) and 8 weeks after the first dose of PZQ (time point E). Fifty percent of the participants received an additional dose of PZQ at time point C (Fig. S1, Supplement Material). The urine samples were kept cold (4°C) immediately after collection and stored at -20°C when the day's field activities were over. For long-term storage, samples were kept at -80°C. Urine samples were analyzed using CCA strips, as previously described(11). The test results thus obtained were given scores of 0, 0.5, 1, 2 and 3, respectively.

Ethical considerations

Ethical clearance was obtained from the Uganda National Council of Science and Technology and the study was presented to the Danish National Committee on Biomedical Research Ethics in Denmark. Consent forms were developed in the local language. Although most of the potential study participants could read the consent forms themselves, the purpose and contents of the study were explained in detail to the community in the local language. They were informed that the decision to participate in the survey was voluntary and any one who wished to withdraw was free without any reprimand. Informed consent was obtained from individual adult participants; for children the parents or guardians consented on their behalf. Thereafter, each individual signed a consent form before commencement of any activity. All information obtained from participants was kept confidential.

Preparation of urine samples for ^{1}H NMR spectroscopy

After thawing, urine samples were centrifuged for 10 min at 4000 rpm at 4°C for the removal of cellular components. For sample preparation 600 µL urine were added to 300 µL of pH 7.4 phosphate buffer (0.2 M) in 10% D_2O containing 1.2 mM sodium 3-trimethylsilyl-tetradeuteriopropionate (TSP) and 0.1% (w/v) NaN_3 in a 96 deep-well plate using an eight-channel Hamilton Microlab STAR robotic workstation and immediately stored at -80°C. Well plates were thawed completely and centrifuged at 3000 g for 10 min prior to analysis in order to remove any precipitate(29).

NMR data acquisition and processing

^{1}H NMR data was obtained using a Bruker 600 MHz AVANCE II spectrometer equipped with a 5 mm TCI cryo probe and a z-gradient system; a Bruker BEST (Bruker Efficient Sample Transfer) system was used in combination with a 120 µL cryofit flow insert for sample transfer. One-dimensional ^{1}H NMR spectra were recorded at 300 K using the first increment of a NOESY pulse sequence (30) with presaturation (γB1 = 50 Hz) during a relaxation delay of 4 sec and a mixing time of 10 msec for efficient water suppression (31). Eight scans of 65,536 points covering 12,335 Hz were recorded and zero filled to 65,536 complex points prior to Fourier transformation, an exponential window function was applied with a line-broadening factor of 1.0 Hz. The spectra were

manually phase and baseline corrected and automatically referenced to the internal standard (TSP = 0.0 ppm). Phase offset artifacts of the residual water resonance were manually corrected using a polynomial of degree 5 least square fit filtering of the free induction decay (FID) (32). In order to monitor proper filling of the NMR flow cell and for quality control one-dimensional gradient profiles (33) along the z-axis were recorded for each sample prior and post data acquisition. Duration of 90 degree pulses were automatically calibrated for each individual sample using a homonuclear-gated nutation experiment (34) on the locked and shimmed samples after automatic tuning and matching of the probe head.

Multivariate data analysis of NMR spectroscopic data

A bucket table with a bucket size of 0.04 ppm was generated for the regions 9.8 – 6.0 and 4.5 – 0.2 ppm, respectively, using an AMIX (version 3.5; Bruker Biospin, Germany) command line based in-house automation routine. Buckets between 2.74 and 2.66 (both included) were merged to take account for the pH-related shift of the citrate region and all buckets were normalized to a total area of $1.0 * 10^4$ followed by log transformation of absolute values. Variables were centered, but not scaled prior to statistical analysis using the SIMCA-P+ (version 12.0, Umetrics, Sweden) software package. For initial analysis and outlier detection principal component analysis (PCA) was performed using 7 components. Outliers were identified based on scores and distance to model-X (DModX) values and visual inspection of the individual NMR spectra. Data with insufficient spectra quality due to poor water suppression, low signal-to-noise ratio (high diluted samples) and high ethanol levels (alcohol abuse) were excluded from the analysis. For partial least squares-discriminant analysis (PLS-DA), orthogonal projection to latent structure discriminant analysis (OPLS-DA) and orthogonal signal correction (OSC) filtering samples were categorized based on egg-count in feces (Kato-Katz smear) (4) and an ELISA based dip-stick assay (11). Subjects were considered as non-infected when egg counts in feces were zero and dry CCA strip test scores below 1. For classification as heavily infected subject an arbitrary cut-off of 500 EPG was chosen. All other samples were classified as low-medium infected ($1 \leq$ EPG < 500 and CCA strip test ≥ 1) or ambiguous (egg count = 0 and CCA strip test ≥ 1). Individual models for children aged 7 to 15 years and adults between 20 to 40 years were built. Data from the group between 15 and 20 years old and from subjects older than 40 years were not selected for further analysis.

RESULTS

Since the complete study cohort is very diverse including male and female subjects of all age groups, different socio-economic background and professions as well as showing differences in prevalence of a variety of co-morbidities and infections the initial

analysis was focused on school-aged children of age ranging from 7 to 15 years only. This selection reduces age-related differences and to some degree avoids variability due to life-style. In addition, acquired immunity and organ damage as result of chronic re-infection with *S. mansoni* should be less pronounced in younger individuals. We also expected to find the metabolic discriminators to be more prominent in this group, since younger individual usually show higher infection intensities compared to older subjects (35). Initial PCA analysis of NMR spectra from 593 samples of children using a model of 7 components explaining 69% of the variability shows a very homogenous distribution in the scores for all 7 components with no clustering according to *S. mansoni* infections as exemplified in the scores plot for the first two principal components in Fig. 1. This lack of initial clustering indicates that the differences in urinary profiles due to *S. mansoni* or any other co-morbidity are rather subtle and the inter-individual and day-to-day variations clearly dominate the model. This is in contrast to animal models where usually clustering due to *S. mansoni* infection can be found within the first 3-4 principal components (PCs) of the PCA scores plot. This better discrimination can most likely be attributed to the controlled environment and the high similarity between the individual animals, as well to the much higher parasite load in the animal models.

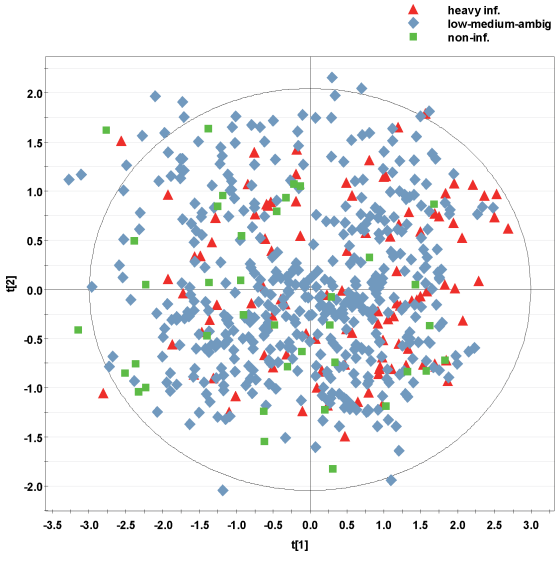


Figure 1. Scores plot of first two principal components from PCA model built using 593 NMR spectra of urine from children aged 7 to 15 years. Data points are colored according to level of infection indicating non-infected individuals with zero *S. mansoni* egg count and CCA < 1 (■); low-to-medium infected individuals with 0 < *S. mansoni* egg count < 500 EPG or no available egg count data and 1 <= CCA < 3 (◆); heavily infected individuals with *S. mansoni* egg count > 500 EPG (▲).

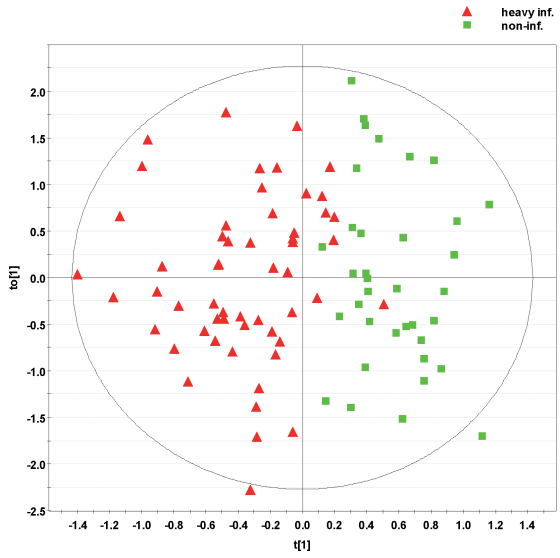


Figure 2. Scores plot of first and orthogonal component from OPLS/O2PLS-DA two class model built using 91 NMR spectra of urine from children aged 7 to 15 years. Data points are colored according to non-infected individuals (■) and heavily infected individuals (▲), respectively.

In order to find molecular discriminators for *S. mansoni* infection and morbidities in humans an OPLS/O2PLS-DA two class model was build using non-infected *vs.* heavy infection classification as response variable. The scores plot in Fig. 2 of the first and orthogonal component of an OPLS-DA model built using 91 urine spectra shows clear discrimination between heavy and non-infected subjects. Based on coefficients and loadings from this model the responsible metabolites for the discrimination of the two groups were identified as summarized in Table 1. Apart from the endogenous metabolites listed in Table 1 we also found a compound of non-mammalian origin as one of the potential discriminators. This compound with resonances in the buckets 1.14, 3.46, 3.50, 3.58, 3.62, 3.66, 3.70, 3.86 and 3.82 ppm drew our attention already at visual inspection of the spectra during data processing due to its high abundance in some of the samples as shown in Fig 3. A full NMR-based structure elucidation (see supplement material) revealed the presence of 2-methylbutane-1,2,3,4-tetrol. The stereochemistry was not determined independently, but based on comparison with literature data (36-38) the compound was annotated as the 2-C-Methylerythritol stereoisomer. Initially an apparent decrease of this non-endogenous metabolite in the presence of *S. mansoni* infection was observed as determined by multivariate data analysis. However, in depth analysis based on relative quantification of the compound in urine (creatinine normalized data) revealed correlation with time-point and negative correlation with age but no link to *S. mansoni* infection (Fig. 4) or any

Table 1. Changes of metabolites observed in urine of *S. mansoni*-infected subjects compared to control. (↑, increase; ↓, decrease; ●, no change in concentration).

Metabolite	Change in Children	Change in Adults	Chemical shift [ppm]
2-Oxo isovalerate	↓	↓	1.12
2-Oxo isocaproate	↓	↓	0.94
3-Hydroxy butyrate	↓	●	1.20
2-Oxo glutarate	↓	↓	3.01
Acetate	●	●	1.92
Acetone	●	●	2.24
Alanine	●	●	1.48
cis Aconitate	●	●	3.13
Citrate	↓	↓	2.67, 2.54
Creatine	↑	●	3.93, 3.04
Dimethylamine	↓	●	2.72
Formate	●	●	8.46
Fumarate	↓	↓	6.52
Guanidino acetate	↑	●	3.80
Hippurate	↓	↓	7.84, 7.64, 7.56, 3.97
Methylguanidine	↑	↑	2.83
Phenylacetylglutamine	↑	↑	7.43, 7.36
Pyruvate	●	●	2.38
Succinate	↓	↓	2.41
TMAO	↑	●	3.27
trans Aconitate	↓	↓	6.59, 3.45
Trimethylamine	↑	↑	2.88
Trigonelline	↓	↓	9.13, 8.84, 8.09, 4.44
3.34 ppm	↑	↑	3.34
3.30 ppm	↑	↑	3.30
2.35 ppm	↑	↑	2.35

of the investigated co-morbidities was seen. The apparent decrease of this metabolite in the OPLS-DA model with infection is therefore an artifact and can most likely be attributed to a bias between time point and infection status in the children model since most heavily infected individuals can be found at time point A whereas most non-infected controls are from the time point E group.

In order to exclude any seasonal bias from the metabolic discriminators of *S. mansoni* infection, a second model was built based on data of urine collected exclusively at time point A. For this model, samples from adults age 20 to 40 were used given that

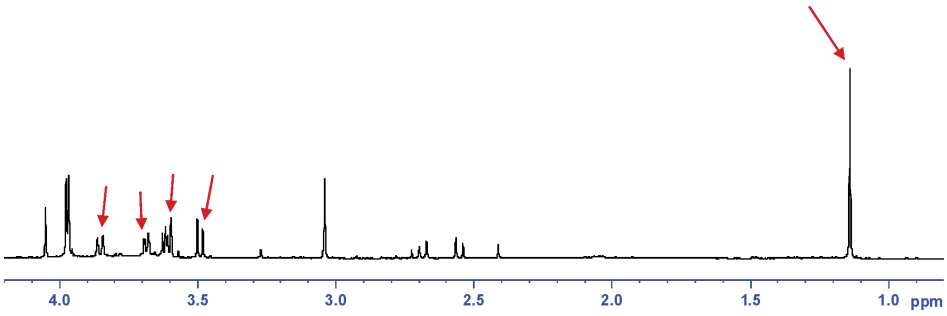


Figure 3. Excerpt of typical 600 MHz 1D-NOESY ^{1}H-NMR spectrum of urine from non-*S. mansoni* infected subject with high content of 2-C-Methylerythritol (resonances as indicated by arrows).

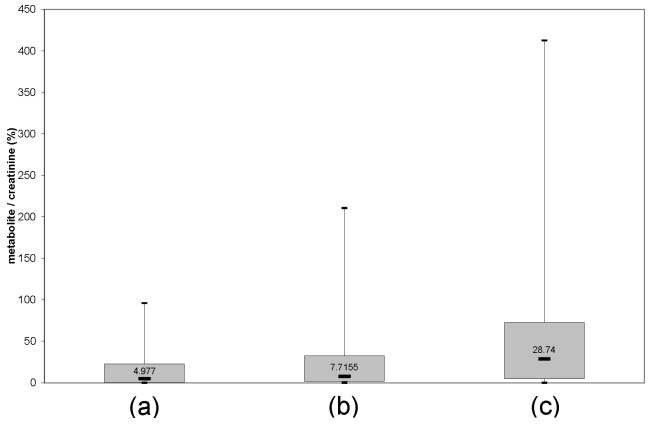


Figure 4. Distribution of normalized 2-C-Methylerythritol content in urine of 20-40 year-old subjects. (a) Samples collected at time point A – first visit - from individuals with no *S. mansoni* infection (egg count = 0); (b) samples collected at time point A from heavily *S. mansoni* infected individuals (egg count > 500 EPG); (c) samples collected at time point E – 8 weeks follow up - from individuals with no *S. mansoni* infection (egg count = 0).

the number of available samples from non-infected children for this time point was not sufficient. We also excluded the buckets (1.10, 1.14, 1.18, 3.46, 3.50, 3.58, 3.62, 3.66, 3.70, 3.86 and 3.82 ppm), representing 2-C-Methylerythritol from the analysis prior to normalization of the data, since the results from relative quantification as described above did not show any correlation with *S. mansoni* infection for this metabolite. In Fig. 5 the scores plot for the first and orthogonal component of a two-class (non *vs.* heavily infected) OPLS-DA model after OSC-filtering using 5 components with a remaining sum of squares of 34.2 % and an Eigenvalue of 2.96 is shown. Spectra from heavily infected and non-infected individuals are clearly separated in the scores plot.

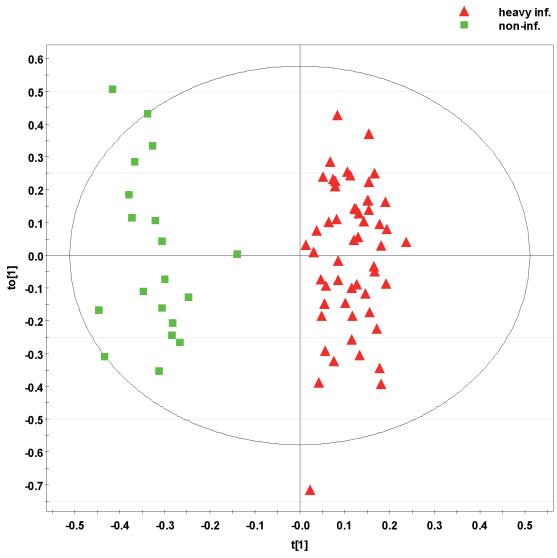


Figure 5. Scores plot of first and orthogonal component from OPLS/O2PLS-DA two class model built after removal of 2-C-Methylerythritol buckets and OSC filtering of data using 73 NMR spectra of urine from adults aged 20 to 40 years. Data points are colored according to non-infected individuals (■) and heavily infected individuals (▲), respectively.

The identified molecular discriminators from this model and their relative changes are largely in agreement with the metabolites identified in the children models. It can therefore be concluded that neither seasonal nor time point bias can be seen for the identified endogenous metabolites and the observed changes of these metabolites are indeed a result of *S. mansoni* infection and *S. mansoni* infection-related morbidities in human.

Since the initial children-based model was relatively limited in the number of spectra that were included, a separate model was built based on 209 urine spectra from adult individuals of age 20 to 40 years. In addition to spectra from time points A and E also spectra from time point B were included with the assumption that the state of infection / egg count is not significantly changing within the 24h time interval between time points A and B. Only urine samples from heavily infected (egg count > 500) and non-infected (egg count = 0) individuals were used to built the model. The scores plot of an OPLS/O2PLS-DA two class model based on this data in Fig. 6 shows a clear trend between the two groups, but as expected the discrimination is not as distinct as for the equivalent children-based model. The identified metabolites responsible for the discrimination between the two classes and their relative changes are summarized in column 2 of Table 1. Direct comparison between the metabolic signatures of *S. mansoni* infection in adults and children show a high degree of concordance. Differences in relative changes were only observed for 3-Hydroxy butyrate, Creatine, Dimethylamine,

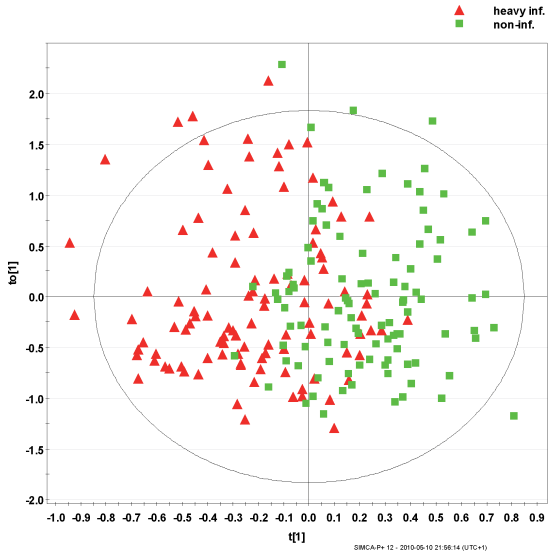


Figure 6. Scores plot of first and orthogonal component from OPLS/O2PLS-DA two class model built using 209 NMR spectra of urine from adults aged 20 to 40 years. Data points are colored according to non-infected individuals (■) with egg count = 0 and heavily infected individuals (▲) with egg count > 500 EPG. For samples collected at time point b) no egg count data were available and classification of infection was extrapolated based on egg count from time point a).

Guanidino acetate and TMAO. The much larger discrimination between the infected and non-infected group in the single time point model in Fig. 5 compared with the multiple time point models in Fig. 2 and 6, respectively, can be attributed to the strong OSC filtering approach, but is also due to less "noise" introduced by seasonal influences (e.g., time point E coincided with the end of the raining season) and absence of any treatment effect (e.g., systemic response to PZQ treatment, see Fig. S8 – supplement material).

DISCUSSION

The current study is part of a multidisciplinary project to investigate the influence of the treatment strategies on *Schistosoma* re-infection rates and *Schistosoma*-related pathology. The overall study design involves the comparison of two treatment strategies and includes samples from 460 individuals collected at five different time points (Fig. S1 – supplement material). However, in the present study we have concentrated on the global examination of urinary metabolites in the context of *S. mansoni* infection aiming at the identification of metabolic signatures of this infection in humans. In fact, so far only one report on metabonomics of *Schistosoma* infection in a human population has been published (39). The main focus of that publication was on the

potential of ^{1}H NMR-based metabonomics in large population studies. Even though molecular discriminators for *Schistosoma* infection in animal models have been reported, no discriminators were provided for human *S. mansoni* infection. However, the complexity of such an analysis due to parasitic co-infections was mentioned. The earlier respective study was carried out in the region of Man, western Ivory Cost – an area with the highest prevalence of schistosomiasis in the country. The total study group consisted of 500 participants (a number comparable to our cohort of 460 participants) aged between 1 and 91 years. Despite many factors (geographical location, differences in nutritional habits, co-infections), making a direct comparison of the results difficult, the authors made several important observations which proved to be of great value for interpretation of our data. For example, they showed a clear age-dependent grouping in PCA analysis. The age factor dominated the first three principal components of the model concealing effects of gender, nutritional status and most importantly infection status and parasitic load. Initial analysis of our cohort also identified age as a dominant factor (supplement material: Age-related effects). Moreover, it has been shown by Naus *et al.* (35), that the intensity of *Schistosoma* infection itself has a convex relationship with age, peaking in childhood and decreasing thereafter. Taken together, these facts support our decision to carry out a separate analysis on each of two age sub-groups: children of 7-15 years and adults of 20-40 years.

The initial PCA analysis of urinary metabolic profiles of children showed no clear infection-related grouping. This is in contrast to studies on animal models where the clustering due to *S. mansoni* infection was found within the first three PC's (24). The lack of clustering in our PCA model reflects, in a way, the complexity of the data set where the combination of such factors as general biological variability, gender differences, nutritional status, co-infections and different infection intensities may dominate the model. In addition, the systemic response in the described animal models is expected to be more pronounced due to the relatively much higher parasite load in the animals. Working in an endemic area we also have to anticipate the fact that multiple co-infections and associated morbidity are "the norm rather than the exception" (39). Together with variations due to social and nutritional status these conceal metabolic changes characteristic for *Schistosoma* infection. To unmask changes relevant to our study we applied supervised modelling building a two-class OPLS/O2PLS-DA model where samples of non-infected and heavily infected children (more than 500 EPG) were used as classes. To avoid a misbalance in the model due to the unequal number of samples in the classes we had to combine baseline time point A, where most of the cases represented heavy infection, with the 8 weeks follow-up time point E where the majority of non-infected samples are present. In the model those samples were treated as independent observations, which would appear to be an acceptable compromise considering the general variability and complexity of the data set. The score plot of the first and orthogonal component shows a clear discrimination between the heavily and

non-infected subjects. However, 2-C-methylerythritol, a compound of non-mammalian origin was identified as a major discriminator between the classes. An analysis based on relative quantification of this compound in urine revealed a positive correlation with the time point of sampling and a negative correlation with age but no correlation with *S. mansoni* infection (Fig. 4) or any of the investigated co-morbidities. Therefore, the buckets corresponding to 2-C-methylerythritol metabolite were debarred from further analysis. This exclusion of 2-C-methylerythritol had minimal influence on the model.

Analysis of the adult group provided the opportunity to build a non-infected *vs.* heavily infected model using only the baseline samples, and excluding any seasonal bias from the metabolic discriminators of *S. mansoni* infection. The resulting model demonstrated clear differences between the classes. Further, to get a model with a more balanced number of spectra per class, we expanded the model by including the urine samples collected at other time points (B and E). The respectively two class OPLS/O2PLS-DA model also showed a clear trend between infected and non-infected adults.

Analysis of the metabolic discriminators most significantly contributing to the models revealed a similarity in the metabolic signatures of infected children and infected adults (Table 1). Furthermore, interestingly there exists a significant overlap with the metabolic signatures of *S. mansoni* infection reported for the animal models (24;25;39). Indeed, animal model data and our data indicate involvement of the same classes of the metabolites, with gut microflora associated compounds and energy metabolism components taking the largest share. At a first glance, it looks like the conventional set of the "usual suspects", but in the context of *S. mansoni* infection, the involvement of these metabolites might have a physiological explanation. For example, the impact of gut microflora on the host metabolism needs no additional "justification", nor the fact that parasitic infections directly affect the symbiotic mammalian-microbial equilibrium (40). Abdominal pain, diarrhea, and blood in the stool are common symptoms of intestinal schistosomiasis. Under such circumstances changes in the concentration of "mammalian-microbial co-metabolites" (41), usually reflected in alteration of urinary aromatic compounds such as hippurate or/and phenylacetylglutamine (PAG), are to be expected. Indeed, in the present study, we observed a reduced concentration of hippurate and elevated levels of PAG. This decrease in hippurate levels is in agreement with the data reported for *S. mansoni*-infected mice, as shown with two independent analytical approaches: NMR-based metabolic profiling and capillary electrophoresis based metabolic fingerprinting (24;26). Moreover, this effect appears to be common to all helminth infections studied to date (22-25;27;42). Elevated levels of PAG, a human metabolite corresponding to mouse metabolite phenylacetylglycine, are also in agreement with data reported for the *Schistosoma*-infected rodents (24;25).

Alterations in the compounds related to the energy metabolism may also be considered as a consequence of the infection. *S. mansoni* resides in the inferior mesenteric vein having direct access to the flow of nutrients. The energy metabolism

of *S. mansoni* solely depends on glucose. The parasite uses large quantities of the host glucose – an amount equal to the dry weight the worm is utilized every 5 hours (43). Being a so-called homolactic fermenter, the worm releases an approximately equal quantity of lactate into the blood stream (44;45). The consequences of glucose losses and the local elevation of lactate levels in the blood stream will depend on the severity of the infection and nutritional status of the host. The energy metabolism of mammalians is undoubtedly more flexible than that of worms: it can use alternative pathways to maintain balance but such adaptive changes will be instantly reflected in the metabolic composition of urine. The available data provides us with little evidence to decide whether TCA intermediates are "tapped off" from the TCA cycle, or released to urine through the anaplerotic pathways. However, it would appear to be a plausible hypothesis to consider an alteration of the energy metabolism related compounds in the urine as part of the host adaptive reaction. In this context also the changes observed in the concentration of 2-oxoacids (2-oxo isovalerate, 2-oxo isocaproate and 2-oxo glutarate) could be interpreted as part of the same response.

Further, rather than building a "mechanistic" pathway-based interpretation we consider the metabolic perturbations observed in the urine as a signature of the response of the host on an infection. This, in turn, raises two questions: (i) does the signature represent the morbidity or the infection status of the host, and (ii) how specific is this signature for *S. mansoni*. This study by no means pretends to give a final answer to these questions. However, as mentioned above our cohort has a longitudinal design and on several occasions we had to include post-treatment samples of by then uninfected individuals into the "non-infected" group. Considering the most common morbidity and pathology characteristics reported for intestinal schistosomiasis such as periportal fibrosis, portal hypertension, hepatosplenomegaly, diarrhea, abdominal pain and blood in the stool, it would appear unlikely that a complete reversal of many of these characteristics would be possible within a two-week period (Fig. S1 – supplement material). Consequently, the observed metabolic signature probably reflects more the infection status rather than a particular state of morbidity. On the other hand, such morbidity symptoms as intestinal bleeding and diarrhea can reverse within the given time-period leading to, at least partial, recovery of gut microflora equilibrium. The specificity of our findings is difficult to evaluate on basis of the available data. It is evident that no single metabolite presented in Table 1 can be a specific marker for schistosomiasis. On the contrary, the metabolic signature itself could, at least in theory, pass the "sensitivity test". To address this hypothesis, an additional study, preferably in another endemic area to allow extrapolation to a different endemic setting, would be necessary.

In conclusion, we have presented for the first time the metabolic signatures of human-*S. mansoni* infection and provided a possible physiological interpretation. The underlying molecular markers of *S. mansoni* infection were found to be primarily

linked to changes in gut microflora, energy metabolism and liver function. These findings are in agreement with data from earlier studies on *S. mansoni* infection in experimental animals and thus provide corroborating evidence for the existence of metabolic response specific for this infection. However the specificity of the presented metabolic signatures and the impact of such factors as, for example, geographical location, differences in nutritional habits, co-infections still remains to be evaluated.

ACKNOWLEDGEMENTS

We thank Claudia de Dood for assistance with sample preparation. The study was supported by The European Commission, Sixth framework programme, Specific Measures in support of International Cooperation (INCO), Contract no.: 517733

REFERENCE LIST

1. Boisier P, Serieye J, Ravaoalimalala VE, Roux J, Esterre P. Ultrasonographical Assessment of Morbidity in Schistosomiasis-Mansoni in Madagascar - A Community-Based Study in A Rural-Population. Trans R Soc Trop Med Hyg 1995;89:208-12.
2. Gryseels B, Nkulikyinka L. The morbidity of schistosomiasis mansoni in the highland focus of Lake Cohoha, Burundi. Trans R Soc Trop Med Hyg 1990;84:542-7.
3. Gryseels B. Morbidity Due to Infection with *Schistosoma-Mansoni* - An Update. Tropical and Geographical Medicine 1992;44:189-200.
4. Katz N, Coelho PM, Pellegrino J. Evaluation of Kato's quantitative method through the recovery of *Schistosoma mansoni* eggs added to human feces. J Parasitol 1970;56:1032-3.
5. Peters PA, El Alamy M, Warren KS, Mahmoud AA. Quick Kato smear for field quantification of *Schistosoma mansoni* eggs. Am J Trop Med Hyg 1980;29:217-9.
6. Kongs A, Marks G, Verle P, Van der Stuyft P. The unreliability of the Kato-Katz technique limits its usefulness for evaluating *S. mansoni* infections. Trop Med Int Health 2001;6:163-9.
7. Deelder AM, Klappe HTM, Vandenaardweg GJMJ, Vanmeerbeke EHEM. *Schistosoma-Mansoni* - Demonstration of 2 Circulating Antigens in Infected Hamsters. Experimental Parasitology 1976;40:189-97.
8. Deelder AM, Dejonge N, Boerman OC, Fillie YE, Hilberath GW, Rotmans JP *et al*. Sensitive Determination of Circulating Anodic Antigen in *Schistosoma-Mansoni* Infected Individuals by An Enzyme-Linked Immunosorbent-Assay Using Monoclonal-Antibodies. American Journal of Tropical Medicine and Hygiene 1989;40:268-72. Deelder AM, Kornelis D, Vanmarck EAE, Eveleigh PC, Vanegmond JG. *Schistosoma-Mansoni* - Characterization of 2 Circulating Polysaccharide Antigens and the Immunological Response to These Antigens in Mouse, Hamster, and Human Infections. Experimental Parasitology 1980;50:16-32.
9. Stothard JR, Kabatereine NB, Tukahebwa EM, Kazibwe F, Rollinson D, Mathieson W *et al*. Use of circulating cathodic antigen (CCA) dipsticks for detection of intestinal and urinary schistosomiasis. Acta Tropica 2006;97:219-28.
10. van Dam GJ, Wichers JH, Ferreira TMF, Ghati D, van Amerongen A, Deelder AM. Diagnosis of schistosomiasis by reagent strip test for detection of circulating cathodic antigen. Journal of Clinical Microbiology 2004;42:5458-61.

11. van Lieshout L, Polderman AM, Deelder AM. Immunodiagnosis of schistosomiasis by determination of the circulating antigens CAA and CCA, in particular in individuals with recent or light infections. Acta Tropica 2000;77:69-80.
12. Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HWL *et al*. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using H-1-NMR-based metabonomics. Nature Medicine 2002;8:1439-44.
13. Griffin JL, Nicholls AW. Metabolomics as a functional genomic tool for understanding lipid dysfunction in diabetes, obesity and related disorders. Pharmacogenomics 2006;7:1095-107.
14. Di Leo A, Claudino W, Colangiuli D, Bessi S, Pestrin M, Biganzoli L. New strategies to identify molecular markers predicting chemotherapy activity and toxicity in breast cancer. Annals of Oncology 2007;18:8-14.
15. Holmes E, Tsang TM, Huang JTJ, Leweke FM, Koethe D, Gerth CW *et al*. Metabolic profiling of CSF: Evidence that early intervention may impact on disease progression and outcome in schizophrenia. Plos Medicine 2006;3:1420-+.
16. Rozen S, Cudkowicz ME, Bogdanov M, Matson WR, Kristal BS, Beecher C *et al*. Metabolomic analysis and signatures in motor neuron disease. Metabolomics 2005;1:101-8.
17. Underwood BR, Broadhurst D, Dunn WB, Ellis DI, Michell AW, Vacher C *et al*. Huntington disease patients and transgenic mice have similar pro-catabolic serum metabolite profiles. Brain 2006;129:877-86.
18. Bogdanov M, Matson WR, Wang L, Matson T, Saunders-Pullman R, Bressman SS, Beal MF. Metabolomic profiling to develop blood biomarkers for Parkinson's disease. Brain 2008;131:389-96.
19. Garcia-Perez I, Couto AA, Angulo S, Li JV, Utzinger J, Ebbels TM *et al*. Bidirectional correlation of NMR and capillary electrophoresis fingerprints: a new approach to investigating *Schistosoma mansoni* infection in a mouse model. Anal Chem 2010;82:203-10.
20. Li JV, Holmes E, Saric J, Keiser J, Dirnhofer S, Utzinger J, Wang YL. Metabolic profiling of a *Schistosoma mansoni* infection in mouse tissues using magic angle spinning-nuclear magnetic resonance spectroscopy. International Journal for Parasitology 2009;39:547-58.
21. Saric J, Li JV, Wang Y, Keiser J, Bundy JG, Holmes E, Utzinger J. Metabolic profiling of an Echinostoma caproni infection in the mouse for biomarker discovery. PLoS Negl Trop Dis 2008;2:e254.
22. Wang Y, Xiao SH, Xue J, Singer BH, Utzinger J, Holmes E. Systems metabolic effects of a necator americanus infection in Syrian hamster. J Proteome Res 2009;8:5442-50.
23. Wang YL, Holmes E, Nicholson JK, Cloarec O, Chollet J, Tanner M *et al*. Metabonomic investigations in mice infected with *Schistosoma mansoni*: An approach for biomarker identification. Proceedings of the National Academy of Sciences of the United States of America 2004;101:12676-81.
24. Wang YL, Utzinger J, Xiao SH, Xue J, Nicholson JK, Tanner M *et al*. System level metabolic effects of a *Schistosoma japonicum* infection in the Syrian hamster. Molecular and Biochemical Parasitology 2006;146:1-9.
25. Garcia-Perez I, Whitfield P, Bartlett A, Angulo S, Legido-Quigley C, Hanna-Brown M, Barbas C. Metabolic fingerprinting of *Schistosoma mansoni* infection in mice urine with capillary electrophoresis. Electrophoresis 2008;29:3201-6.
26. Saric J, Li JV, Swann J, Utzinger J, Calvert G, Nicholson JK *et al*. Integrated Cytokine and Metabolic Analysis of Pathological Responses to Parasite Exposure in Rodents. J Proteome Res 2010;9:2255-64.
27. Balog, C. I. A., Alexandrov, T., Dercks, R. J., Hensbergen, P. J., van Dam, G. J., Tukahebwa, E. M., Kabatereine, N. B., Thiele, H., Vennervald, B. J., Mayboroda, O. A., and Deelder, A. M. The feasibility of mass spectrometry and advanced data processing for monitoring *Schistosoma mansoni* infection. Prot.Clin.Applic. 2010.
28. Ref Type: In Press

29. Keun HC, Ebbels TM, Antti H, Bollard ME, Beckonert O, Schlotterbeck G *et al*. Analytical reproducibility in (1)H NMR-based metabonomic urinalysis. Chem Res Toxicol 2002;15:1380-6.
30. Kumar A, Ernst RR, Wuthrich K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem Biophys Res Commun 1980;95:1-6.
31. Price WS. Water Signal Suppression in NMR Spectroscopy. Annu Rep NMR Spectrosc 1999;38:289-354.
32. Coron A, Vanhamme L, Antoine JP, Van HP, Van HS. The filtering approach to solvent peak suppression in MRS: a critical review. J Magn Reson 2001;152:26-40.
33. Vanzijl PCM, Sukumar S, Johnson MO, Webb P, Hurd RE. Optimized Shimming for High-Resolution NMR Using Three-Dimensional Image-Based Field Mapping. J Magn Reson A 1994;111:203-7.
34. Wu PS, Otting G. Rapid pulse length determination in high-resolution NMR. J Magn Reson 2005;176:115-9.
35. Naus CW, Booth M, Jones FM, Kemijumbi J, Vennervald BJ, Kariuki CH *et al*. The relationship between age, sex, egg-count and specific antibody responses against *Schistosoma mansoni* antigens in a Ugandan fishing community. Trop Med Int Health 2003;8:561-8.
36. Fontana A, Messina R, Spinella A, Ciminо G. Simple and versatile synthesis of branched polyols: (+)-2-*C*-methylerythritol and (+)-2-*C*-methylthreitol. Tetrahedron Lett 2000;41:7559-62.
37. Giner J-L, Ferris WV, Mullins JJ. Synthesis of 2-Methyl-D-erythritol via Epoxy Ester-Orthoester Rearrangement. J Org Chem 2002;67:4856-9.
38. Anthonsen T, Hagen S, Kazi MA, Shah SW, Tagar S. 2-C-Methyl-erythritol, a New Branched Alditol from Convolvuluc glomeratus. Acta Chem Scand B 1976;30:91-3.
39. Singer BH, Utzinger J, Ryff CD, Wang Y, Holmes E. Exploiting the Potential of Metabonomics in Large Population Studies: Three Venues. In: Lindon JC, Nicholson JK, Holmes E, eds. The Handbook of Metabonomics and Metabolomics. Oxford: Elsevier, 2007:289-325.
40. Nobre V, Serufo JC, Carvalho OS, Mendonca CL, Santos SG, Mota EM *et al*. Alteration in the endogenous intestinal flora of Swiss Webster mice by experimental Angiostrongylus costaricensis infection. Mem Inst Oswaldo Cruz 2004;99:717-20.
41. Yap IKS, Angley M, Veselkov KA, olmes E, indon JC, icholson JK. Urinary Metabolic Phenotyping Differentiates Children with Autism from Their Unaffected Siblings and Age-Matched Controls. J Proteome Res 2010;9:2996-3004.
42. Wu JF, Holmes E, Xue J, Xiao SH, Singer BH, Tang HR *et al*. Metabolic alterations in the hamster co-infected with *Schistosoma japonicum* and Necator americanus. Int J Parasitol 2010;40:695-703.
43. Bueding E. Carbohydrate metabolism of *schistosoma mansoni*. J Gen Physiol 1950;33:475-95.
44. Bueding E, Fisher J. Metabolic requirements of Schistosomes. J Parasitol 1982;68:208-12.
45. Schiller EL, Bueding E, Turner VM, Fisher J. Aerobic and anaerobic carbohydrate-metabolism and egg-production of *Schistosoma-mansoni* invitro. J Parasitol 1975;61:385-9.

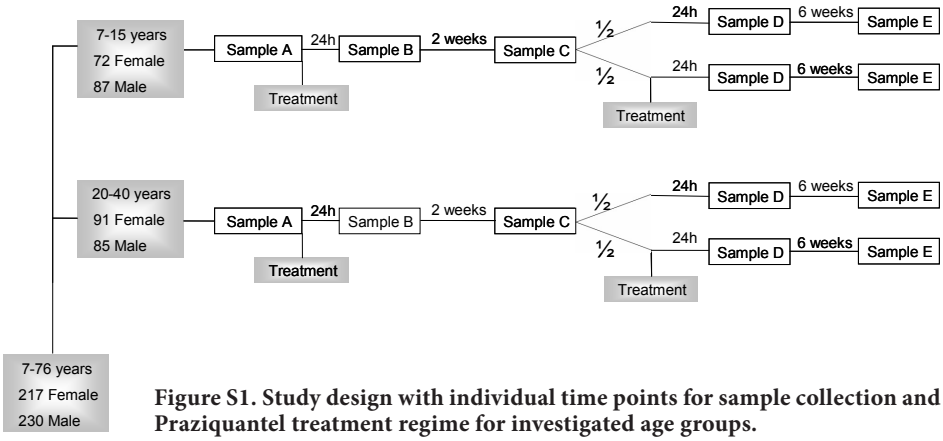


Figure S1. Study design with individual time points for sample collection and Praziquantel treatment regime for investigated age groups.

Structure elucidation of unknown metabolite M01

^{1}H detected 1D and 2D NMR spectra of urine sample 65803240905 TubeID: WELL0032#NMR#D03 were obtained using a Bruker 600 MHz AVANCE II spectrometer equipped with a 5mm TCI cryo probe and a z-gradient system. 1D proton spectra with water suppression (S1) were recorded using the first increment of a NOESY pulse sequence (S2) with presaturation (γB_1 = 50Hz) during 4 sec relaxation delay and a mixing time of 10 msec. 8 scans of 32768 points covering 6002.4 Hz were recorded at 300K and zero filled to 65536 complex points prior to Fourier transformation, an exponential window function was applied with a line-broadening factor of 1.0 Hz.

All 2D experiments were recorded at a temperature of 300K with presaturation (γB_1 = 50Hz) during a relaxation delay of 2 sec. For DQF-COSY (S3) spectra a data matrix of 256 x 2048 points covering 6009.6 x 6009.6 Hz was recorded with 8 scans for each increment. Data was zero filled to 2048 x 2048 points prior to States-TPPI type 2D Fourier transformation and a sine bell shaped window function shifted by $\pi/2$ in the F1 and $\pi/4$ in the F2 dimension was applied. Coherence order selective gradient HSQC (S4) spectra were recorded for a data matrix of 256 x 2048 points covering 24146 x 6009.6 Hz with 4 scans for each increment. Data was zero filled to 512 x 2048 points prior to echo-anti echo type 2D Fourier transformation and a sine bell shaped window function shifted by $\pi/2$ in both dimensions was applied. For HMBC (S5) spectra a data matrix of 256 x 2048 points covering 33202 x 6009.6 Hz with 16 scans for each increment. Data was zero filled to 512 x 2048 points prior to echo-anti echo type 2D Fourier transformation and a sine bell shaped window function shifted by $\pi/2$ in the F1 dimension and $\pi/6$ in the F2 dimension was applied. Magnitude mode spectra were obtained by magnitude calculation in F2. All spectra were referenced according to the internal TSP = 0.0 ppm.

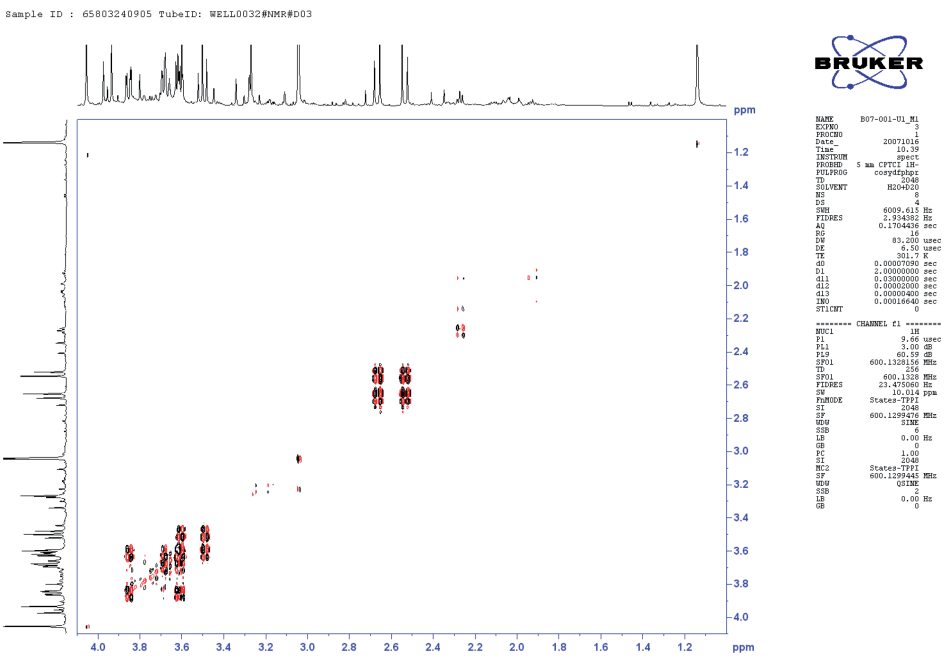


Figure S2. Excerpt of DQF-COSY spectrum of sample 65803240905.

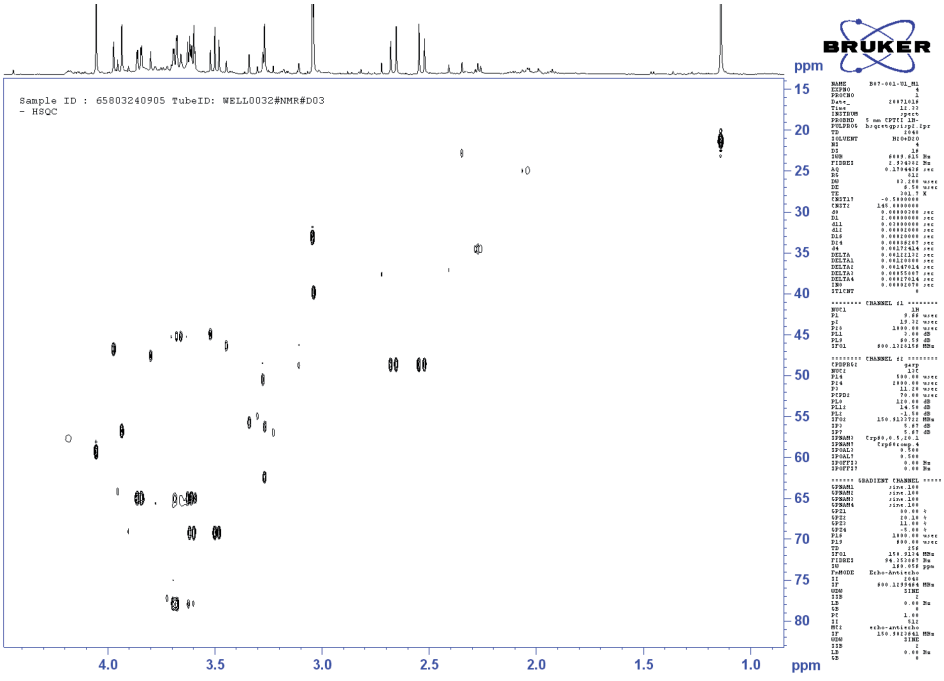


Figure S3. Excerpt of HSQC spectrum of sample 65803240905.

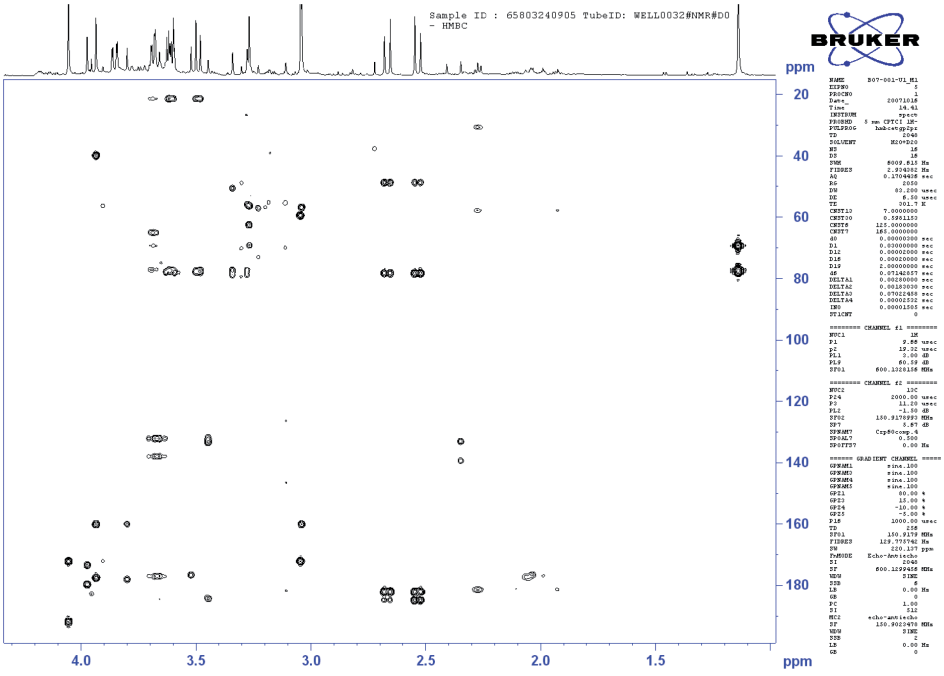


Figure S4. Excerpt of HMBC spectrum of sample 65803240905.

Age related effects

The strategy of focusing on individual age groups for investigation of *S. mansoni* infection was chosen based on the fact that in the initial PCA analysis an age related trend could be observed in the scores plot already for the first two components as seen in Figure S5. Such age group related clustering has previously been described in the literature for large population studies (S6). Therefore, selection of individual age groups will reduce age related artifacts and background variability within the data.

The age related trend was further investigated by building an OPLS model with the age group as the response variable as seen in Figure S6. In this model a clear separation between children and adults is visible with samples from adolescent individuals (age group 15-19) located in between these two clusters. No trend or clustering for the children age groups 7-9 and 10-14 years can be observed. For the individual adult age groups there is a slight trend visible, however, it is much less pronounced as for the separation between children and adults.

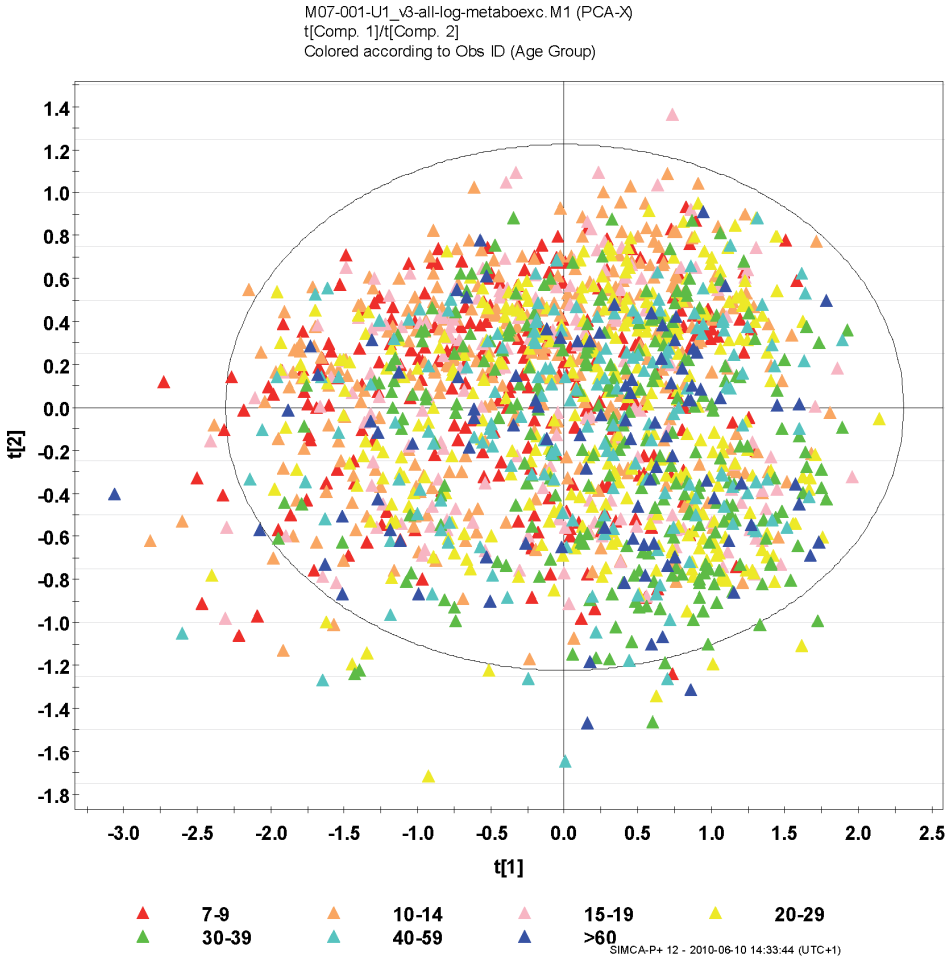


Figure S5. Scores plot of first two components from initial PCA model using all data colored according to age groups.

For better visualization of these age-group related changes the response variable for each individual sample were plotted against the predicted values as calculated using the OPLS model (Figure S7). In this representation an age related trajectory can be drawn with the above described characteristics. Major metabolic discriminators as identified from the loadings plot are Creatinine, Lactate and PAG that show an increase with advancing age, whereas Hippurate, *trans*-Aconitic acid, Citrate and Formiate are decreasing with age.

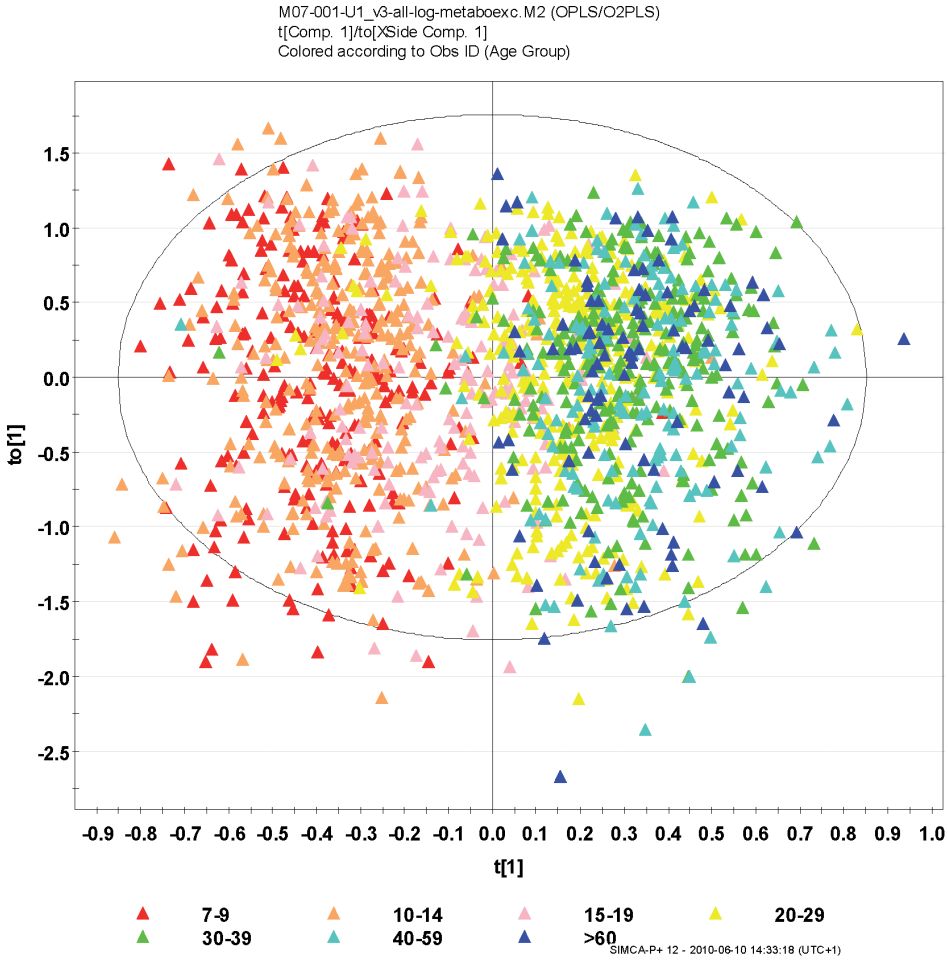


Figure S6. Scores plot of first two orthogonal components from OPLS model build using age groups as response variable.

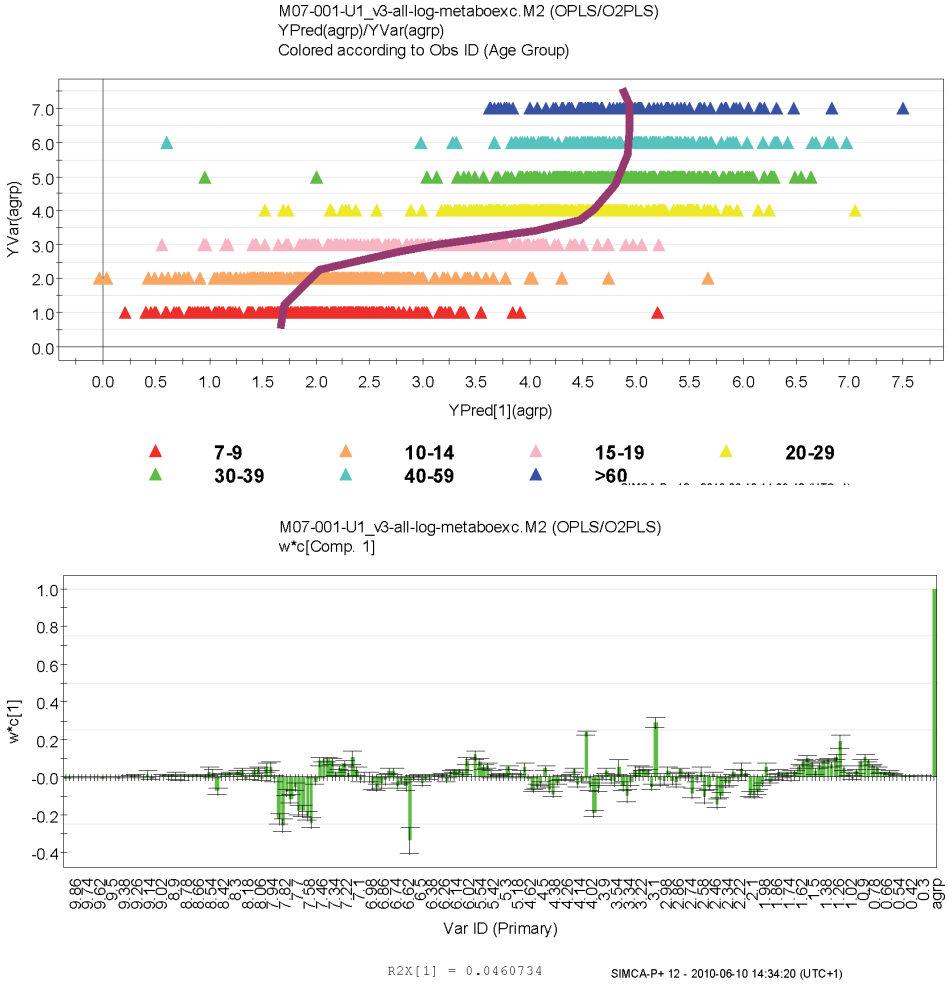


Figure S7. Observed vs. predicted plot and loadings plot for first component of OPLS model build using age groups as response variable.

Treatment effect

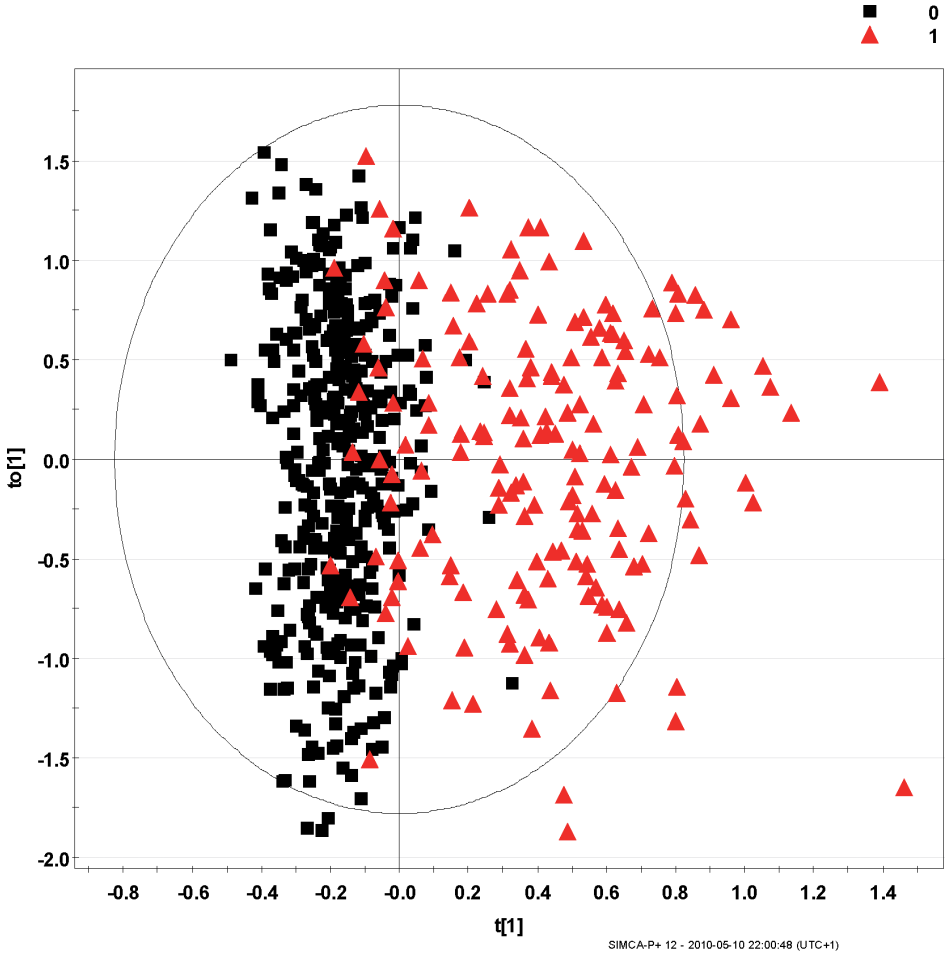


Figure S8. Scores plot of first and orthogonal component from OPLS/O2PLS-DA two class model built using 602 NMR spectra of urine using treatment status as response variable illustrating Praziquantel treatment effect. Data points are colored according to before treatment (■) and 24h post treatment (▲) time points including initial and 2 weeks follow-up visits.

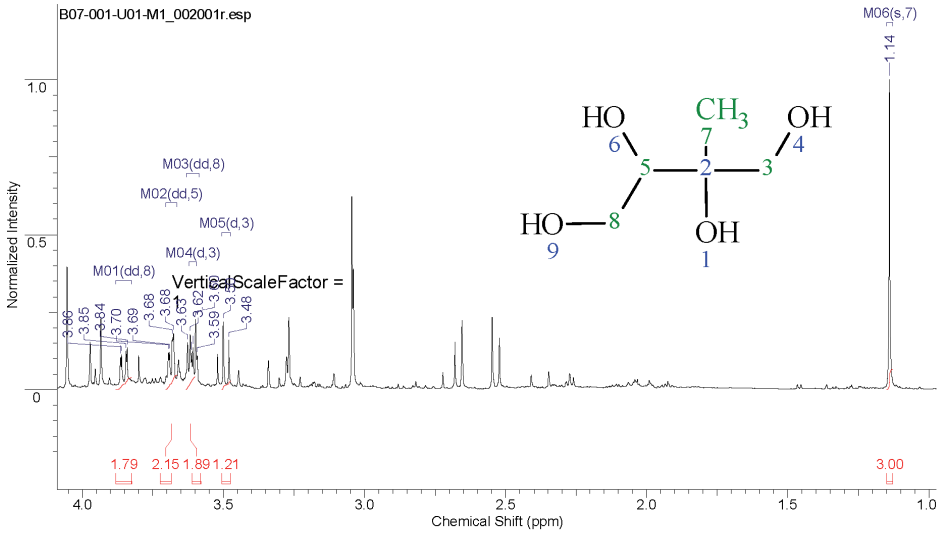
Unknown Metabolite M01 from Urine - 3-C-Methylerythritol

17/10/2007 14:04:18

Formula $C_5H_{12}O_4$	FW	136.1464

Acquisition Time (sec)	2.7296	**Comment**	Sample ID : 65803240905 TubeID: WELL0032#NMR#D03		
Date	16 Oct 2007 08:17:04			**Date Stamp**	16 Oct 2007 08:17:04
File Name	C:\Documents and Settings\All Users\Shared Documents\data\Biomarker\nmr\B07-001-U01-M1_002001r				
Frequency (MHz)	600.13	**Nucleus**	1H	**Number of Transients**	8
Origin	spect	**Original Points Count**	16384	**Owner**	nmrsu
Points Count	65536	**Pulse Sequence**	noesygppr1d	**Receiver Gain**	16.00
SW(cyclical) (Hz)	6002.40	**Solvent**	H2O+D2O	**Spectrum Offset (Hz)**	2868.6404
Sweep Width (Hz)	6002.31	**Temperature (degree C)**	28.700		

1H NMR (600 MHz, H2O+D2O) δ ppm 1.14 (s, 3 H) 3.49 (d, *J*=11.72 Hz, 1 H) 3.61 (dd, *J*=11.54, 8.52 Hz, 1 H) 3.61 (d, *J*=11.72 Hz, 1 H) 3.69 (dd, *J*=8.56, 2.34 Hz, 1 H) 3.85 (dd, *J*=11.63, 2.56 Hz, 1 H)



No.	Shift1 (ppm)	H's	Type	J (Hz)	Atom1	Multiplet1	(ppm)
1	1.14	3	s	-	7	M06	[1.13 .. 1.15]
2	3.49	0	d	11.72	3	M05	[3.48 .. 3.51]
3	3.61	0	d	11.72	3	M04	[3.60 .. 3.62]
4	3.61	1	dd	11.54, 8.52	8	M03	[3.59 .. 3.63]
5	3.69	1	dd	8.56, 2.34	5	M02	[3.66 .. 3.70]
6	3.85	0	dd	11.63, 2.56	8	M01	[3.83 .. 3.88]

No.	(ppm)	(Hz)	Height
1	1.14	683.8	1.0000
2	3.48	2089.2	0.1605
3	3.50	2100.9	0.2190
4	3.59	2156.8	0.1101
5	3.60	2159.6	0.2275
6	3.61	2165.2	0.1361
7	3.61	2168.2	0.1167
8	3.62	2171.3	0.1785
9	3.63	2176.8	0.1522
10	3.68	2206.8	0.1819
11	3.68	2209.0	0.1678
12	3.69	2215.2	0.1207
13	3.70	2217.7	0.1205
14	3.84	2305.4	0.1356
15	3.85	2308.0	0.1246
16	3.86	2317.0	0.1137
17	3.87	2319.6	0.1052

^{13}C shift (ppm)	type	Atom
77.0	Cq	2
69.2	CH_2	3
77.9	CH	5
21.2	CH_3	7
64.9	CH_2	8

REFERENCES

[S1] Price W.S. Water Signal Suppression in NMR Spectroscopy. *Annual Reports on NMR Spectroscopy*, **1999**, *38*, 289-354.

[S2] Kumar A., Ernst R.R., Wüthrich K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem Biophys Res Commun.*, **1980**, *95*, 1-6.

[S3] Derome A., Williamson M. Rapid pulsing Artifacts in Double-Quantum-Filtered COSY. *J. Magn. Reson.,* **1990**, *88*, 177-185.

[S4] Kay L.E., Keifer P., Saarinen T. Pure absorbtion gradient enhanced hetero-nuclear Single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.,* **1992**, *114*, 10663-10665.

[S5] Bax A., Summers M.F. 1H and 13C Assignments from Sensitivity-Enhanced Detection of Heteronuclear Multiple-Bond Connectivity by 2D Multiple Quantum NMR. *J. Am. Chem. Soc.,* **1986**, ***108***, 2093.

[S6] Singer G.H., Utzinger J., Ryff C.D., Wang Y., Holmes E. Exploiting the Potential of Metabonomics in Large Population Studies: Three Venues, in *The Handbook of Metabonomics and Metabolomics,* 1st ed.; Lindon, J. C., Nicholson, J. K., Holmes, E., Eds.; Elsevier: Oxford, 2007; pp 289-325.