Comparison of extraction methods in *in vitro Plasmodium falciparum:* a ^{1}H-NMR and LC-MS joined approach

Lúcia Mamede[1], Fanta Fall[2], Matthieu Schoumacher[3], Allison Ledoux[1], Céline Bugli[4], Pascal De Tullio[3], Joëlle Quetin-Leclercq[2], Bernadette Govaerts[4], Michel Frédérich[1*]

[1] Laboratory of Pharmacognosy, Center of Interdisciplinary Research on Medicines (CIRM), University of Liège, Belgium

[2] Pharmacognosy research group, Louvain Drug Research Institute (LDRI), UCLouvain, , Brussels

[3] Laboratory of Pharmaceutical Chemistry, Center of Interdisciplinary Research on Medicines (CIRM), University of Liège, Belgium

[4] Statistical Methodology and Computing Service (SMCS/LIDAM), UCLouvain, Louvain-la-Neuve, Belgium

***Corresponding author**

Lúcia Mamede, Université de Liège, CIRM Laboratoire de Pharmacognosie, CHU Av Hopital 1, B36 4000 Liège, Belgium.

Phone : + 3243664330 Fax: + 3243664332 E-mail: lcccmamede@uliege.be

Keywords: Metabolomics, Malaria, Plasmodium sp., Mass Spectrometry, Nuclear Magnetic Resonance, Methodology

Abstract

Malaria is a parasitic disease that remains a global concern and the subject of many studies. Metabolomics has emerged as an approach to better comprehend complex pathogens and discover possible drug targets, thus giving new insights that can aid in the development of antimalarial therapies. However, there is no standardized method to extract metabolites from *in vitro Plasmodium falciparum* intraerythrocytic parasites, the stage that causes malaria. Additionally, most methods are developed with either LC-MS or NMR analysis in mind, and have rarely been evaluated with both tools. In this work, three extraction methods frequently found in the literature were reproduced and samples were analyzed through both LC-MS and ^{1}H-NMR, and evaluated in order to reveal which is the most repeatable and consistent through an array of different tools, including chemometrics, peak detection and annotation. The most reliable method in this study proved to be a double extraction with methanol and methanol/water (80:20, v/v). Metabolomic studies in the field should move towards standardization of methodologies and the use of both LC-MS and ^{1}H-NMR in order to make data more comparable between studies and facilitate the achievement of biologically interpretable information.

Introduction

Malaria is a vector-borne parasitic disease that remains a global health issue[1]. The latest data indicate that there were 249 million estimated malaria cases in 2022, with a death toll of 608 000 of which 76% are children[1]. These rates of incidence and mortality are still relevant and make eradication all the more important. The *Plasmodium sp.* is a protozoon that affects hundreds of species and when it infects the human host, its cyclic invasion of red blood cells leads to fevers in a specific diagnosis characteristic of malaria. This implies that the parasitic metabolism has uniquely adapted to this environment, making antimalarial drug discovery a challenging field[2].

The study of the metabolism was made possible with the emerging discipline that comprehensively studies a biological system through various lenses, including *in silico* models and analytical technologies, also broadly named omics[3]. These omics sciences began with genome sequencing, which for the *Plasmodia* started with the publishing of *P. falciparum*'s genome in 2002, and saw their exponential growth with the optimization of analytical tools and statistic models[2,3]. Specifically in the case of metabolomics, the metabolome is analyzed, which comprises metabolites, small molecules (<1500 Da) that reflect accurately and rapidly the activity of enzymes, proteins and pathways, leading to a faithful snapshot of the parasite's status[2,3]. Some examples include amino acids, vitamins, cofactors, nucleotides, fatty acids, among others; all compounds that provide energy, signaling or building blocks essential for parasitic survival. Powerful and robust techniques such as Liquid Chromatography (LC) coupled with mass spectrometry (MS), or Nuclear Magnetic Resonance (NMR) make metabolomics reliable because of their sensitivity, selectivity and reproducibility[4,5].

Extensive research published on the *Plasmodium* metabolome makes use of metabolomics analysis unrivaled advantages[2]. It has the capability of reflecting the adaptation of the parasite to exposure to a drug, revealing both the senescence cascade or the resistance mechanisms, crucial to find new targets and for rational drug development[3,6–8]. Notwithstanding the many studies on the *Plasmodium* metabolome, there is no consensus on a standardized method for metabolite extraction, despite it being the key departure point for metabolomic studies[9]. In the literature, dozens of methods exist that variate in the use of saponin for red blood cell (RBC) lysis, quenching, extraction solvents, among other factors. The disparity in methodology might mean a different range of metabolites is extracted with each study, consequently changing the biological deductions that can be inferred. Moreover, most literature methods were developed for LC-MS analysis, raising the question on whether they can be used for ^{1}H-NMR as well. Ideally, a metabolite extraction method should be repeatable, reproducible and extract a metabolome as representative as possible[9]. Additionally, most methods have been developed with either LC-MS or NMR in mind, and have rarely been evaluated with both tools. In this work, three methods frequently used in the studies of the *P. falciparum* metabolome *in vitro* were reproduced, analyzed through ^{1}H-NMR and verified once through LC-MS, and evaluated in order to reveal which parameters award the most reproducibility in order to achieve methodology optimization. These methods were chosen based on their relevance in the literature, with the addition of one method developed with ^{1}H-NMR analysis in mind[10–12].

Materials & Methods

Reagents

All pipettes, bottles and sterile materials to handle the culture were acquired from Greiner Bio-One. Sorbitol was obtained from Sigma. Saponin (Alfa Aesar), and PBS were obtained from Thermo Fisher Scientific. Methanol of HPLC grade was obtained from Merck, chloroform (Merck), acetonitrile LC-MS-grade was obtained from VWR. Milli-Q water was obtained with a milli-Q reference A+ system® from Merck. Trimethylsilyl-3-propionide acid-d4 (TMSP) and deuterium oxide (D_2O, 99.96% D) were purchased from CortecNet (France).

Parasite culture and maintenance

Blood-stage *P. falciparum* 3D7 parasites obtained from the Malaria Research and Reference Reagent Resource 453 Center (MR4) were cultured in human erythrocytes at 3% hematocrit in various volumes in complete media[13]. This media consisted of RPMI 1640 (Gibco, Fisher Scientific, Loughborough, U.K.) containing $NaHCO_3$ (32 mM), HEPES (25 mM), and L-glutamine, supplemented with 1.76 g/L of glucose (Sigma-Aldrich, Machelen, Belgium), 44 mg/mL of hypoxanthine (Sigma-Aldrich, Machelen, Belgium), 100 mg/L of gentamycin (Gibco, Fisher Scientific, Loughborough, U.K.), and 10% human

pooled serum (A+). Cultures were microscopically verified for stage and parasitemia and were kept gassed (90% N_2, 5% O_2, 5% CO_2) and incubated at 37°C. Parasites were kept synchronously by weekly 5% sorbitol (w/v) treatment and additionally treated 24h before each assay.

Metabolomic extraction

In each assay, asexual intraerythrocytic ring-stage cultures not older than 8h as observed microscopically (> 90%) were aliquoted equally to a minimum of 10^8 parasites/sample and extracted by either method according to the design shown in Figure 1. One extra assay was performed in the same conditions, this time in six replicates per method, and analyzed through LC-MS. Samples were placed in an ice bath before extraction. Three methods from the literature were performed with slight adaptations as shown in Figure 2.

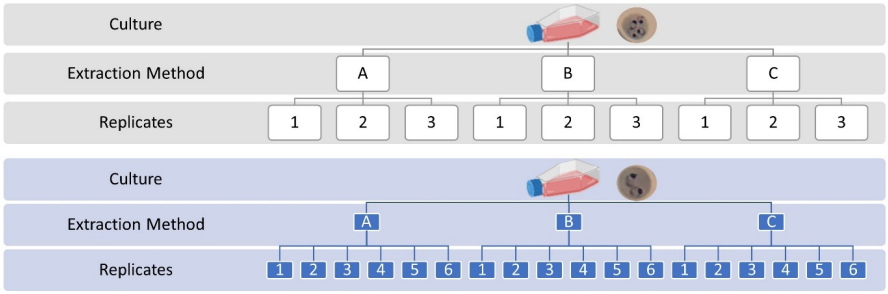

Figure 1 – Experimental design, in gray the N=3 analyzed by ^{1}H-NMR, and in blue the assay (N=1) analyzed by LC-MS. All cultures were synchronized with sorbitol and microscopically verified to start the extraction no later than 8h post-infection.

Method A was published by Vo Duy *et al* [10] involved RBC lysis with 0.01% saponin (w/v) for 3 min, followed by centrifugation (3000 rpm for 8 min) and a PBS wash. The parasite pellet is then extracted in two cycles: first, incubated 15 mins at -20°C with cold methanol (1:4, v/v), then centrifuged at 11,000 rpm for 10 min to keep the supernatant; second, resuspended with 2 pellet volumes of a mixture of cold methanol/water (80:20, v/v) followed by an incubation of 15 mins at -20°C, followed

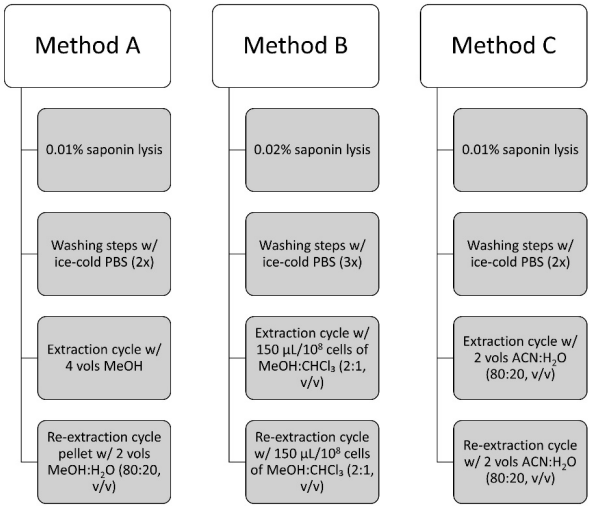

Figure 2 – Schematic of each extraction method under study.

by centrifugation. Both supernatants were pooled, evaporated, lyophilized and kept at -20°C until resuspension for analysis.

Method B was published by Teng *et al* [11] and it comprised saponin lysis (0.02%, w/v) for 90s, followed by centrifugation at 3000 rpm for 8 min and two subsequent PBS washes before the pellet was centrifuged 7700 rpm for 2 min before being stored at -20°C. The frozen pellet was extracted at first with a mixture of methanol/chloroform (2:1, v/v) at a proportion of 150 µL/10^8 cells and vortexed until a suspension is formed. Ice-cold water was then added at a proportion of 20 µL/10^8 cells and incubated for 15 min in an ice bath, followed by a thawing at 4°C for 3 minutes. A second volume of ice-cold water (150 µL/10^8 cells) was added, vortexed and centrifuged at 11,000 rpm for 10 min. After supernatant collection, the pellet was re-extracted with methanol/water (2:1, v/v) at a proportion of 150 µL/10^8 cells and incubated 5 minutes on ice. After centrifugation at 11,000 rpm for 10 min, the supernatant was collected, pooled, evaporated, lyophilized and kept at -20°C until resuspension for analysis.

Method C was published by Dickerman *et al* [12] and it encompassed lysis with 0.01% saponin (w/v), vortexing, pelleting by centrifugation at 3000 rpm for 8 min followed by a PBS wash. The pellet was extracted with acetonitrile/water (80:20, v/v), followed by centrifugation at 11,000 rpm for 10 min before the supernatant was collected and this process repeated. Supernatants were pooled, evaporated, lyophilized and kept at -20°C until resuspension for analysis.

Samples to be analyzed by NMR were dissolved in 400 µL of buffered D_20 at pH 7.4 with TMSP as internal reference, and transferred into 3 mm NMR tubes (Bruker) for analysis. For LC-MS analysis, mixture of 100 µL of formate/acetonitrile (20:80 v/v) was used for resuspension and transferred in an LC-HRMS vial.

Instrumentation

NMR spectra were acquired on a Bruker NEO Ultrashield Plus 700 MHz equipped with a helium cold probe (cryoprobe). ^{1}H-NMR experiments were performed with a CPMG sequence with 128 scans collected over a spectral width of 20 ppm. All spectra were phased and baseline-corrected manually using TopSpin v4. Spectra were stacked, aligned and integrated between δ0.5-9.5 ppm using MestReNova v14. Spectra were divided into buckets of 0.04 ppm, integrated to the sum of intensities and normalized to the number of parasites per sample.

Other samples were analyzed using a LC-HRMS system consisting in a Thermo Accela pump, autosampler, photodiode array detector and Thermo Scientific LTQ orbitrap XL mass spectrometer at the MASSMET platform of UCLouvain. Samples were injected (10 µL) into a hydrophilic interaction liquid chromatography (HILIC) column with a Phenomenex Luna 3 mm x 150 mm, 200 A HILIC (Louvain, Belgium). The mobile phase consisted of A: 10 mM pH 3.8 ammonium formate, and B: acetonitrile and the gradient elution started with 5% solvent A until 3 min, then increased to reach 95% at 25 min and maintained for 5 more min, then back to 5% and equilibrated for 10 min. Flow rate was 0.3 mL/min; oven temperature was 40°C and total run time was 40 min.

Metabolomic Data analysis

The bin tables generated by the NMR spectra were analyzed using MetaboAnalyst v5.0 and *R* (packages MBXUCL, PepsNMR and limpca)[14–16]. The NMR spectra were annotated using Chenomx NMR Suite 9.0 database and the Human Metabolome Database (HMDB), as according to literature[10,11].

Data acquisition was done in positive mode and raw LC-HRMS data profiles were converted into mzXML format with msConvert (ProteoWizard) using the Filter "Peak Picking". MzXML format files

were then processed using XCMS package to Worklow4Metabolomics 3.3 (W4M). The CentWave algorithm was used for automatic peak detection. Statistical analysis was performed with MetaboAnalyst 5.0.

Results

Chemometric Visualization and Description

Metabolomics is a discipline that generates complex datasets with multidimensional data. Generally, the first step in data analysis is data visualization through a chemometric tool, such as Principal Component Analysis (PCA), which can then be followed by more telling models, such as ASCA+ (ANOVA-Simultaneous Component Analysis) and APCA+ (ANOVA-Principal Component Analysis)[15,17]. These enhanced versions of the original statistical methods use general linear models instead of ANOVA to correct the bias of unbalanced experimental designs to generate a tool that

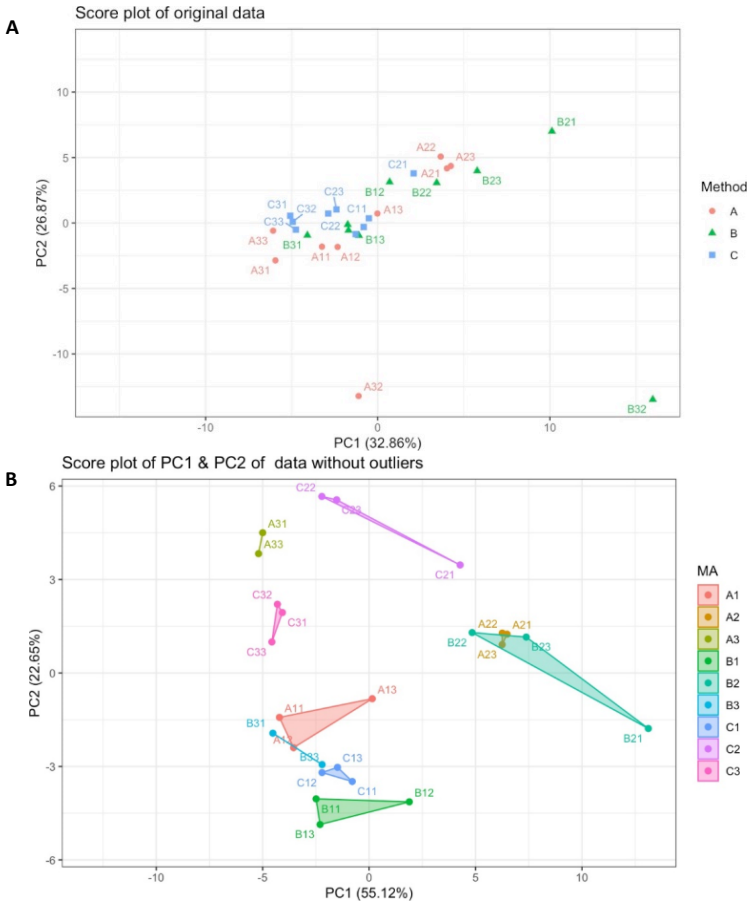

*Figure 3 – (*A) PCA score plot of the original ^{1}H-NMR bin data organized by method where the two outliers are clear. (B) PCA score plot of the original ^{1}H-NMR bin data without outliers organized by method and assay in which groups cluster mostly by assay. Points represent replicates. (MA – Method-Assay)

incorporates multivariate analysis of variance with PCA for eased data visualization in a reduced space[15]. This tool was used in this context to analyze the assay effect and remove its interference from the outcome.

Figure 3A shows the PCA score plot of the original ^{1}H-NMR data for each extraction method (of which one representative spectrum can be found in Figure S1 in the Supplementary Data. It reveals two clear outliers, one for method A (sample A32) and one for method B (B32), which were removed for subsequent analysis. A new PCA scores plot was generated and is shown in Figure 3B in which it is clear that each group of samples separated by assay and extraction method group together. However, it also becomes clear that samples tend to gather in either quadrant of the PCA in regards to the assay, which shows the influence of this factor that was then removed.

Because assays were conducted at three independent times, it was important to decompose the outcomes matrix in the effect matrices: Method + Assay + Residuals. This was done using ASCA+ with the package limpca and the results are shown in Figure 4A and 4B. Figure 4A shows the percentage of variance explained by the effects and it is perceivable how the effect 'assay' represents the biggest variance. Because this effect is not interesting for the purpose of this study, it was removed by centering the data for the assay variability and a new PCA score plot was generated – Figure 4B. Method A has less intragroup variability, as seen by the distribution across the PC1

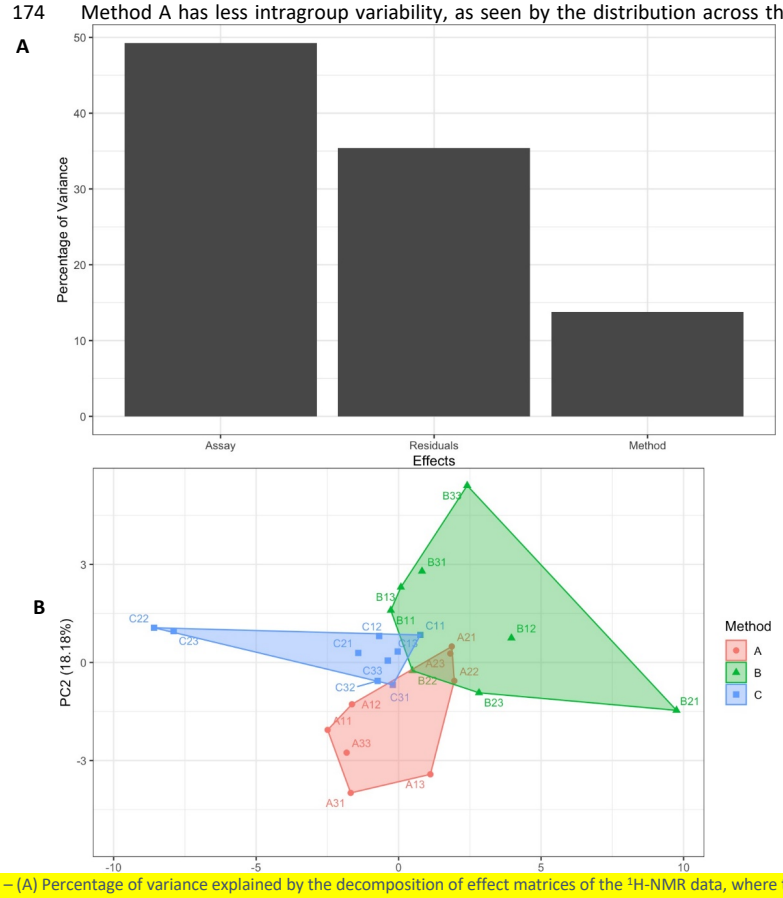

Figure 4 – (A) Percentage of variance explained by the decomposition of effect matrices of the ^{1}H-NMR data, where the matrix effect of the assays represents the biggest variance. (B) PCA score plot obtained by centering the data to the 'assay' effect after rebuilding the data matrix with the 'limpca' package. This allows to observe directly the variability of the extraction methods across two PCs and conclude that method B is the most variable.

(56.5%), which reveals this method as the most homogenic. Method C has some variability associated with two samples from the second assay and Method B is the most variable across both PC1 and PC2. PC2 (18.29%) has the least variability for method C, which would indicate that this PC is related to a source of variability that is more impactful on the other two methods.

Figure 5 shows the PCA scores plot of the LC-MS-analyzed assay, where groups correspond to each individual method A, B or C. All methods are clearly separated from each other with their respective samples clustering closely. The variation explained by components 1 and 2 accounts for a big part of the variation of the data (70.5 %). Method B would appear the most variable, as perceivable by its variance across the PC1, followed by method A and C.

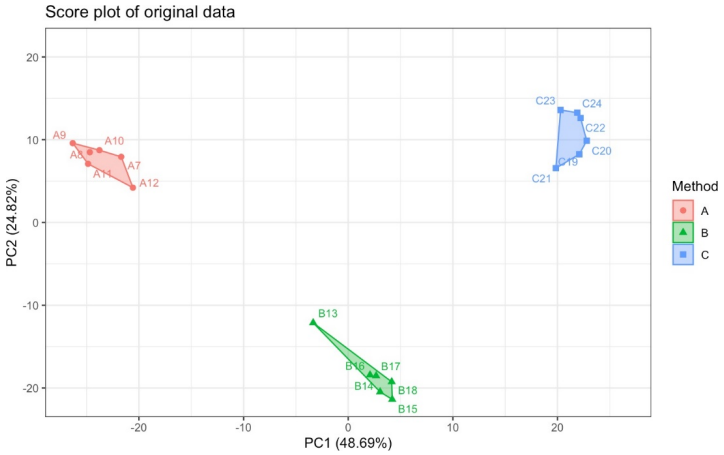

Figure 5 – PCA scores plot of the LC-MS data organized by method, which confirms the variability of extraction method B with a different analytical technique.

Inertias

One technique developed specifically for the evaluation of the Metabolomic Informative Content (MIC) in NMR metabolomics studies was also used to analyze this dataset[18]. Contrary to using just repeatability in spectrometry, the MIC uses a clustering approach that evaluates the amount of captured information (*i.e.* signal) compared to noise that would unintentionally come from other factors. The method considers that a signal is responsible for group clustering, thus dependent on the group's characteristics and non-identical between groups, whereas noise is independent and identically distributed, as it is likely to come from other factors such as the experimental design, the operator, analytical apparatus, etc. To differentiate them, the model decomposes the total variance into two parts: between groups, *i.e.* intergroup, and within the groups of observations, *i.e.* intragroup[19]. Ideally, the variance, or inertia, within groups should be small, as it would mean that observations are the most similar. Likewise, if the inertia between groups is big, then it is indicative of the quantity of captured signal that has informative value.

Table 1 resumes the inertia of the three methods as analyzed by ^{1}H-NMR. Method A has the least intra group variability demonstrated by the lowest measure of inertia within assays of the three methods (13.2%). Method B and C have similar inertias, 25.41% and 25.65%, respectively, which indicates similar levels of variability. A high inertia between assays and consequently small inertia within groups translates into a better signal over noise ratio with more metabolomic informative content, which reveals repeatability and robustness across assays.

	Inertia Between Assays	Inertia Within Assays
Method A	86.8%	13.2%
Method B	74.59%	25.41%
Method C	74.36%	25.65%

Table 1: Measurement of the Metabolomic Informative Content (MIC) through inertia per method of the original ^{1}H-NMR data (without outliers).

Because the MIC measures inertia between repetitions, it could not be applied to the LC-MS assay, which was conducted only once. However, a measure of inertia can be performed by calculating the volume of the PCA ellipses for all PCs for each method. This calculation can be found in Table 2 and it translates an abstract measure of variability that allows to compare the inertia per method between analytical tools. Similarly with the MIC, the smaller the value, the smaller the variability associated with a method analyzed with that tool. Interestingly, the methods that have the least inertia are not the same for both analytical tools – method A is the least variable when analyzed by NMR, followed closely by method C, whereas through LC-MS method C is clearly less variable. Unsurprisingly, method B is the most variable with both analytical tools.

	NMR	**LC-MS**
Method A	0. 06860552	1.308199
Method B	0.1092255	1.578083
Method C	0.07758823	0.9162912

Table 2: Inertias calculated by method through both the NMR (assay centered data) and LC-MS data.

Through exploration models, a general overview of the datasets is possible. PCA of the original data exposed two outliers for methods A and B, respectively, in ^{1}H-NMR analysis. ASCA+ and APCA+ analysis, in which the 'assay' effect was removed, confirmed the variability of method B regardless of this effect. Through MIC analysis, method A displayed the minimal within assays inertia and maximal between assays inertia, which correlates with the least intragroup variance and the highest metabolomic informative content. Lastly, LC-MS analysis generally demonstrated a good separation between samples extracted by method A, B or C. A second calculation of inertias shown in Table 2 to compare the variability across analytical tools showed that the methods perform differently between platforms – method A shows less variability for NMR and method C for LC-MS. Chemometrics analysis points to a close variability and repeatability between methods A and C while confirming that method B is not reliable.

Metabolite Detection

Metabolite detection can be broadly assumed as the number of signals detected in the analysis of a metabolic extract. As one single molecule may have multiple protons, which will elicit multiple ^{1}H-NMR peaks for one sole metabolite, the number of peaks cannot be translated to number of metabolites. Still, this number can be correlated with the success of the metabolic extraction, as if there are no metabolites there will be no signals.

For this reason, automatic peak picking was performed with MestReNova v14, which uses a Global Spectral Deconvolution algorithm to automatically pick only positive peaks, in this case. The average number of peaks variated between assays, though it remained similar across methods: 281 ± 34 peaks for method A, 289 ± 65 peaks for method B and 295 ± 37 peaks for method C. Through the standard deviation (SD) it is evident that the method B has the biggest variation, whereas the other two methods are fairly similar.

In LC-MS, features are detected, i.e., a two-dimension signal of retention time per *m/z*. Though the whole chromatographic run is analyzed, missing values are a common occurrence in which zero intensity is detected. These can have biological or technical reasons, and the way of dealing with this issue is a topic of discussion in the metabolomics field[20]. Regardless, missing values may also be used as a metric of the quality of a metabolic extraction method, as it can play a role in the detection by MS. Table 3 summarizes the results obtained. The mean peak intensities for each extraction method along with reproducibility were calculated, which in this context is evaluated as the same metabolite being detected in at least 5 out of 6 replicates. After the application of filters to remove peaks present in the blanks, a total of 187 peaks were detected. When evaluating reproducibility, 144 peaks are detected in at least 5/6 of the replicas with method A, followed by method C, with 139 peaks and method B, with 117 peaks. Method A displayed a higher number of detected features with higher average intensities and fewer missing values.

	Method A	Method B	Method C
Missing Values (MV)	192	345	254
Mean with MV	5.03	4.05	4.61
Mean without MV	6.07	5.85	5.96
Peaks in at least 5/6 samples	144	117	139
Within SD	0.35	0.42	0.35

Table 3: LC-MS analysis: missing values, mean of peak intensities with and without MV, reproducibility and intragroup SD for each extraction method.

Detection of peaks through both ^{1}H-NMR and LC-MS hence points method B as the least consistent, with method C showing the biggest average peak number through NMR and method A through LC-MS. Method A appeared more consistent through LC-MS with the least missing values and highest number of features detected in at least five out of six samples.

Annotation

Annotation of metabolites is a crucial step in processing metabolomic data. It is effectively the step that allows for biological inference and result interpretation. However, it is also the most complex and time consuming stage, as the degree of certainty of annotation can play a large role in the confidence of the results[17]. Previous metabolomic studies with *P. falciparum* successfully detected and annotated parasitic metabolites, thus annotation can be used as a measurement of quality for an extraction method in this context.

As such, the table of detected metabolites published by Teng *et al* [11], where method B was published, was used as a frame of reference for the annotation of ^{1}H-NMR data. Chenomx and HMDB databases were used to identify specific chemical shifts attributed to each metabolite. The full list of metabolites can be found in the Supplementary Data, and it consists broadly of amino acids, membrane precursors, nucleotides, carboxylates and contaminants (ethanol, methanol, and for this study, acetonitrile which was added as the extraction solvent of method C). In total, 53 metabolites and contaminants were searched in the spectra, 43 through Chenomx and 10 manually with HMDB's spectra references.

For all methods the same range of metabolites was annotated, with emphasis on the lowest annotated spectra in the case of method B (13 metabolites). Method B's variability didn't make possible to annotate metabolites reported to be extracted through this method, such as γ-aminobutyric acid, putrescine or spermidine[11]. Between method A and C, the range was similar, between 18-32 and 17-34 metabolites, respectively. It was not possible to annotate every metabolite as reported in the

reference table, however, this might be a consequence of the lower amount of parasites used in these assays (~ 10^8), in regards to the amount used by Teng *et al.* (~1-4 x 10^8)[11].

Visibly, all methods were in the same range of metabolites annotated, but there were differences in the metabolites, as shown in Table 4. A total of 13 metabolites from diverse classes were identified differently across extraction methods. Method C accounts for more consistency as all the 13 metabolites were identified in most samples. Method B displays the highest variation regarding detection of these metabolites, with neither being found in all samples. For method A, only $NADP^+$ was not found in any sample, possibly because of the quicker experimental time or another technical factor, as Vo Duy *et al.* reported annotation of this cofactor through this extraction method[10].

Class	Metabolite	Method A	Method B	Method C
Amino acids	Asparagine	Yes[1]	Yes[1]	Yes[1]
	Glutamate	Yes	Yes[1]	Yes[1]
	Glutamine	Yes[1]	Yes[1]	Yes[1]
	Phenylalanine	Yes[1]	Yes[1]	Yes[1]
	Serine	Yes[1]	Yes[1]	Yes[1]
	Tyrosine	Yes[1]	Yes[1]	Yes[1]
Nucleotides and related compounds	AMP	Yes	Yes[1]	Yes
	Hypoxanthine	Yes[1]	Yes[1]	Yes[1]
	IMP	Yes	Yes[1]	Yes
	$NADP^+$	No	Yes[1]	Yes[1]
Glutathione	Reduced	Yes[1]	Yes[1]	Yes[1]
Carboxylates	Fumarate	Yes	Yes[1]	Yes[1]
Soluble membrane precursors	*myo-Inositol*	Yes	Yes[1]	Yes[1]

Table 4: Differences in the annotated metabolites per method, according to the class. Annotation performed with Chenomx. [1] – not found in all samples.

Discussion

In this study, three extraction methods from the literature were compared by ^{1}H-NMR and verified once by LC-MS analysis in order to choose the most robust and reliable method towards studying the *Plasmodium falciparum* metabolome. Several methods in this context have been published in the literature, but the evolution of metabolomics technology warrants for further changes in order to develop one or more methods that are reproducible and robust across all analytical platforms. Method A, developed by Vo Duy *et al* [10], was validated for quantitative LC-MS and it was a starting point to more recent extraction methods in the field, including method C, adapted from by Dickerman *et al* [12][9]. Both methods have not been, to the authors knowledge, studied and evaluated through ^{1}H-NMR, hence the interest in this study. Method B was published by Teng *et al* [11] and was the first research paper in the literature to compare *in vitro* metabolomic extraction methods and analyze them by ^{1}H-NMR, hence assuring that the signal was repeatable enough for accurate characterization. Other methods have been published and adapted in the last decade in the field, but for the sake of time and complexity, only these three were studied in detail. Additionally, as most have already been studied or adapted in LC-MS studies, the focus stayed mainly on ^{1}H-NMR analysis with one LC-MS assay confirmation.

The *Plasmodium* spp. complex lifecycle that involves human infection presents multiple opportunities for effective blocking of malaria, be it as a prevention, treatment or transmission blocking. However, as the mortality and mobility associated with this disease remain high and resistance is a growing concern, studies that aid in antimalarial drug research can still focus on the life stages that elicit

Commenté [CCMLC1]: The asexual blood stages, often the most interesting target for antimalarials, have a relatively fast and complex cycle, so the choice of life stage to study is relevant. As referred previously, the ring stage is metabolically less active and thus difficult to study, whereas late trophozoites are more active and easier to study due to non-invasive purification techniques such as magnetic cell sorting. There are other synchronization and purification techniques such as sorbitol and Percoll, but these may introduce contaminants if performed immediately before an assay. Additionally, it is not viable to wait for merozoite invasion post magnetic purification because it relies on available non-infected RBCs that will then also decrease the purity of sample (bellow 100% concentration) and re-introduce RBC-related contaminants.

symptoms and complications, namely the intraerythrocytic stages. Within this cycle, the parasite evolves morphologically and can be distinguished both metabolically and visually from an early trophozoite, also named ring-stage, to late trophozoite and into a schizont that can release multiple merozoites that will reinfect new red blood cells. The ring-stage was chosen in the context of this study to pave the way for assays that could routinely be implemented with resistant strains. Whether to evaluate the effects of promising antiplasmodial compounds or facilitate the discovery of compounds with activity in the early stages, characteristic for being the stage to support artemisinin resistance, this stage seemed the most challenging.[21] Despite reports of the ring-stage being the least metabolically active, our study supports the exploration of this stage with robustness as long as a few parameters are controlled, like synchrony and high enough number of parasites per sample.[22,23] Additionally, none of the extraction methods studied are limited to the ring-stage; purification with other methods like magnetic cell sorting or Percoll are described in the literature and could be implemented prior to these methods.[8,9] This could be an interesting future venue to explore.

Samples were processed as indicated previously and prepared for either analytical platform. As conservation was not a parameter of this study, all samples were evaporated and freeze-dried to remove solvents and assure stability until analysis. Despite this, traces of water, ethanol and methanol could still be found through ^{1}H-NMR, noting how difficult it is to eliminate solvent traces. These residues were however significantly minimized through this processing and the implementation of a water suppressing sequence during acquisition, which made their impact negligible.

Additionally, it is worth mentioning that extraction methods were performed as indicated in the literature, but new approaches and factors have since been acknowledged as important in the context of these types of metabolomics assays.[24] In culture, even high levels of parasitemia (*e.g.* 10%) remain relatively low in what might represent significant RBC contamination. Other techniques, like culturing the *Plasmodium* sp. without the RBC or through other enrichment techniques, like magnetic cell sorting or the SLOPE method have since become more frequent than methods with saponin, which lyse RBC but has been shown to be associated with the presence of "ghost contamination" from the RBC membranes.[24–26] In this work, all samples were treated with saponin to induce RBC lysis and its effect should be similar across all samples, regardless of the following extractions steps. Still, for the development of future extraction methodologies, this parameter is essential and should be the focus of careful research.

Chemometrics analysis was done through PCA for both analytical platforms as it is the most often used model to graphically give an overview of the variability of the data. Additionally for the NMR data, two other approaches were used: an ASCA+ analysis, to eliminate the assay effect from the analysis, and the MIC algorithm, which consists of a more recently developed model in NMR metabolomics[17,18]. It is important to consider factors external to the method of extraction such as the different assays, which can change the outlook on the results and the choice of method of extraction for further experiments. The assay effect was chosen as a known fixed categorical factor that could be easily evaluated and was relevant in this study. The consensus of the models used for ^{1}H-NMR data analysis is that method A has less intragroup variability and is the most repeatable. Through LC-MS, it is clear that the three methods are distinct from one another, with method A and C clustering in different sides of the PC1 (48.69%), which would indicate relevant differences between the two. The inertia per method for both analytical tools was different and it could be linked to the fact that method C's extraction solvent is also present in the LC-MS mobile phase. Despite lyophilization prior to reconstitution and analysis, the use of acetonitrile at both stages could facilitate solubilization and hence remove a factor that could introduce variability. Of the three methods, method B had the most variability across both ^{1}H-NMR and LC-MS data.

Other chemometrics and statistical models could have been used, but all would have suffered from the same issue: the number of replicates. This study had three triplicates in three independent assays for ^{1}H-NMR, and one set of six replicates for LC-MS. These numbers are considered small in what statistical analysis is concerned, as the number of replicates translates directly into the statistical model's robustness and predictability. This remains a bottleneck in statistical analysis of metabolomics data, as the number of variables is much bigger than the replicates. Indeed, it is still very challenging to obtain high quantities of purified parasite *in vitro* which explains the limitations in the number of replicates. Still, PCA, ASCA+, APCA+ and MIC are not predictable tools and are less affected by this issue than other models, hence their application in this study.

Additionally, metabolite detection was evaluated through the spectral signals. It is important to note that not all signals are metabolites, as background noise or matrix effects can also produce misleading signals. Still, the number of peaks or features is indicative of the number of metabolites in a sample, which can give information on the capability of an extraction method. In this context, method C was found the most repeatable, maintaining a high number of peaks without variating as much as methods A and B between assays for ^{1}H-NMR. It is worth noting that particularly in NMR, superposition plays a role in metabolite detection and peak picking, as metabolites with the same functional groups will have clusters around the same chemical shift window. Fortunately, even in such cases, most metabolites can be distinguished by other characteristic chemical shifts (if there is more than one), allowing annotation even in the presence of hundreds of compounds. Techniques like 2D NMR can be used to tackle this issue, but have further sensitivity issues and are thus less used in metabolomics, especially in malaria *in vitro* studies, where biological material is a limitation.

Through LC-MS, method A was found to be the most repeatable, with the highest number of features detected in at least five out of six samples, and the method with the least missing values. This result could be somewhat predictable, as the method was originally developed for LC-MS, but there might be other factors to consider. As LC-MS is a more sensitive technique, it could be that it can detect more metabolites more consistently across the different methods, hence reporting faithfully a greater consistency for method A. Simultaneously, method A did show the best results through chemometrics and the MIC approaches to ^{1}H-NMR data, so there is a consistency in the results.

Annotation is the analysis step that transforms all the metabolomic data into biologically interpretable results and its success is paramount. As such, a table published by Teng *et al.* was used as a reference to search for *P. falciparum* metabolites and evaluate if the extractions were successful[11]. All three methods succeeded in extracting parasite related metabolites, though to different degrees. All methods allowed the identification of between 13 and 34 metabolites, with method B being the most variable. Differences in annotation were revealed across diverse metabolic classes, such as amino acids, nucleotides and related compounds, and others. Method C was the most consistent method in regards to annotation, followed by method A. The use of a list of metabolites as reference was important in this context to reliably search for metabolites previously reported in *P. falciparum* and detected through ^{1}H-NMR. However, it should be noted that the use of a list as a framework fails to account for other metabolites that a method might have successfully extracted that were not part of the list, but could prove significant in establishing a method as more thorough than the others.

Conclusion

It should be noted that in the case of a metabolomic extraction method, repeatability and reproducibility should be valued as the most important criteria, as metabolomics suffers from metabolites with very fast turnover and technical problems, like matrix effects, missing values, among others. These factors influence significantly the amount of reliable information that can be deduced

from a study's dataset, which is already limited by the number of samples in comparison to the hundreds to thousands of variables to be analyzed. As such, a method that is the most repeatable and robust will diminish, though not eliminate, experimental bias and should be selected as a standard. Accordingly, the most reliable method evaluated in this study proved to be method A, *i.e.,* two cycles of extraction with methanol and methanol/water (80:20, v/v), followed closely by method C, a double extraction with acetonitrile/water (80:20, v/v). Metabolomic studies, whose workflows should be optimized towards reproducibility, are pivotal in the learning of plasmodial pathology and in rational drug design in the path towards malaria eradication.

Author contributions

AL, JQL and MF had the idea for the essay. LM and FF performed the essays and ran the NMR and LC-MS analysis. LM, MS and PT generated and analyzed the NMR data. FF generated and analyzed the LC-MS data. CB and BG contributed with the statistical analysis. LM wrote the main manuscript text. All authors read the manuscript and participated in the preparation of the final version of the manuscript. All authors reviewed and approved the final manuscript.

Acknowledgments

This work was supported by the Belgian National Fund for Scientific Research (FNRS, grant PDR T.0092.20). Authors would like to thank Arianna Cirillo (University of Liège) for her insights on the NMR analysis.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1 World Health Organization (2023) *World Malaria Report 2023*,

2 Yu, X. *et al.* (2021) From Metabolite to Metabolome: Metabolomics Applications in Plasmodium Research. *Front. Microbiol.* 11, 1–13

3 Sexton, A.E. *et al.* (2019) Post-Genomic Approaches to Understanding Malaria Parasite Biology: Linking Genes to Biological Functions. *ACS Infect. Dis.* 5, 1269–1278

4 Wishart, D.S. (2010) Computational Approaches to Metabolomics. In *Bioinformatics Methods in Clinical Research* 593 (Matthiesen, R., ed), pp. 283–313, Humana Press

5 Emwas, A.M. (2015) The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research. In *Methods in molecular biology* 1277 (Bjerrum, J. T., ed), pp. 161–193, Springer New York

6 Krishnan, A. *et al.* (2020) Vitamin and cofactor acquisition in apicomplexans: Synthesis versus salvage. *J. Biol. Chem.* 295, 701–714

7 Plata, G. *et al.* (2010) Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. *Mol. Syst. Biol.* 6, 408

8 Cobbold, S.A. and McConville, M.J. (2019) Determining the Mode of Action of Antimalarial Drugs Using Time-Resolved LC-MS-Based Metabolite Profiling. In *Methods in Molecular Biology* 1859pp. 225–239

9 Mamede, L. *et al.* (2022) Recent metabolomic developments for antimalarial drug discovery. *Parasitol. Res.* DOI: 10.1007/s00436-022-07673-7

10 Vo Duy, S. *et al.* (2012) A quantitative liquid chromatography tandem mass spectrometry method for

metabolomic analysis of Plasmodium falciparum lipid related metabolites. *Anal. Chim. Acta* 739, 47–55

11 Teng, R. *et al.* (2009) Metabolite profiling of the intraerythrocytic malaria parasite Plasmodium falciparum by 1H NMR spectroscopy. *NMR Biomed.* 22, 292–302

12 Dickerman, B.K. *et al.* (2016) Identification of inhibitors that dually target the new permeability pathway and dihydroorotate dehydrogenase in the blood stage of Plasmodium falciparum. *Sci. Rep.* 6, 1–15

13 Trager, W. and Jensen, J.B. (1976) Human malaria parasites in continuous culture. *Science (80-.).* 193, 673–675

14 Martin, M. *et al.* (2018) PepsNMR for 1H NMR metabolomic data pre-processing. *Anal. Chim. Acta* 1019, 1–13

15 Thiel, M. *et al.* (2023) limpca: An R package for the linear modeling of high-dimensional designed data based on ASCA/APCA family of methods. *J. Chemom.* DOI: 10.1002/cem.3482

16 Pang, Z. *et al.* (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* 49, W388–W396

17 Phelan, V. V. (2020) *Computational Methods and Data Analysis for Metabolomics*, 2104Springer US.

18 Féraud, B. *et al.* (2015) Statistical treatment of 2D NMR COSY spectra in metabolomics: data preparation, clustering-based evaluation of the Metabolomic Informative Content and comparison with 1H-NMR. *Metabolomics* 11, 1756–1768

19 Féraud, B. *et al.* (2019) Two data pre-processing workflows to facilitate the discovery of biomarkers by 2D NMR metabolomics. *Metabolomics* 15,

20 Wei, R. *et al.* (2018) Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Sci. Rep.* 8, 1–10

21 Mok, S. *et al.* (2021) Artemisinin-resistant K13 mutations rewire Plasmodium falciparum's intra-erythrocytic metabolic program to enhance survival. *Nat. Commun.* 12, 1–15

22 Beri, D. *et al.* (2019) Insights into physiological roles of unique metabolites released from Plasmodium-infected RBCs and their potential as clinical biomarkers for malaria. *Sci. Rep.* 9, 1–11

23 Murithi, J.M. *et al.* (2020) Combining Stage Specificity and Metabolomic Profiling to Advance Antimalarial Drug Discovery. *Cell Chem. Biol.* 27, 158-171.e3

24 Carey, M.A. *et al.* (2018) Influential Parameters for the Analysis of Intracellular Parasite Metabolomics. *mSphere* 3, e00097-18

25 Brown, A.C. *et al.* (2020) Cholesterol-dependent enrichment of understudied erythrocytic stages of human Plasmodium parasites. *Sci. Rep.* 10, 1–15

26 Yu, X. *et al.* (2023) Ring-stage growth arrest: Metabolic basis of artemisinin tolerance in Plasmodium falciparum. *iScience* 26, 105725