

Metabolic Clustering Analysis as a Strategy for Compound Selection in the Drug Discovery Pipeline for Leishmaniasis

Emily G. Armitage,^{†,‡,§,||} Joanna Godzien,[†] Imanol Peña,[‡] Ángeles López-González,[†] Santiago Angulo,[†] Ana Gradillas,[†] Vanesa Alonso-Herranz,[†] Julio Martín,[‡] Jose M. Fiandor,[‡] Michael P. Barrett,^{§,||} Raquel Gabarro,[‡] and Coral Barbas^{*,†,||}

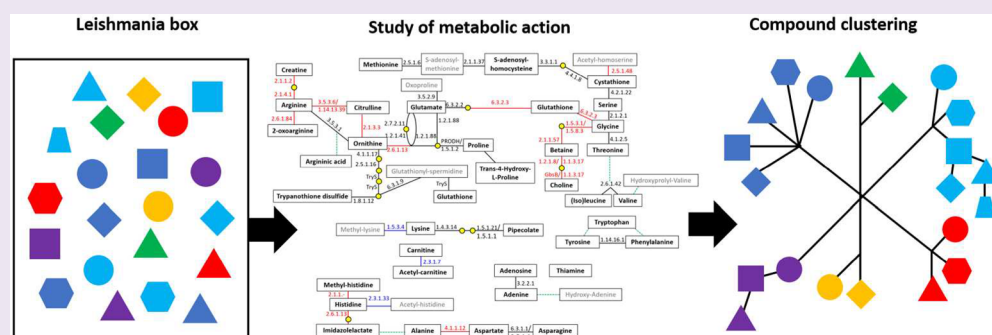
[†]Centre for Metabolomics and Bioanalysis (CEMBIO), Facultad de Farmacia, Universidad CEU San Pablo, Campus Montepríncipe, Boadilla del Monte, 28668 Madrid, Spain

[‡]GSK I+D Diseases of the Developing World (DDW), Parque Tecnológico de Madrid, Calle de Severo Ochoa 2, 28760 Tres Cantos, Madrid, Spain

[§]Wellcome Centre for Molecular Parasitology, Institute of Infection, Immunity and Inflammation, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8TA, United Kingdom

^{||}Glasgow Polyomics, Wolfson Wohl Cancer Research Centre, College of Medical Veterinary & Life Sciences, University of Glasgow, Glasgow G61 1QH, United Kingdom

Supporting Information



ABSTRACT: A lack of viable hits, increasing resistance, and limited knowledge on mode of action is hindering drug discovery for many diseases. To optimize prioritization and accelerate the discovery process, a strategy to cluster compounds based on more than chemical structure is required. We show the power of metabolomics in comparing effects on metabolism of 28 different candidate treatments for Leishmaniasis (25 from the GSK Leishmania box, two analogues of Leishmania box series, and amphotericin B as a gold standard treatment), tested in the axenic amastigote form of *Leishmania donovani*. Capillary electrophoresis–mass spectrometry was applied to identify the metabolic profile of *Leishmania donovani*, and principal components analysis was used to cluster compounds on potential mode of action, offering a medium throughput screening approach in drug selection/prioritization. The comprehensive and sensitive nature of the data has also made detailed effects of each compound obtainable, providing a resource to assist in further mechanistic studies and prioritization of these compounds for the development of new antileishmanial drugs.

Due to a lack of viable hits or increasing resistance to currently available treatments, the bottleneck in research toward new therapies for many different diseases is a growing concern. Limited knowledge on the mode of action (MoA) or polypharmacological effects of existing treatments could be hindering the discovery of new compounds. Studying compound MoA is valuable to understand how they could be improved, to propose combination therapies looking for synergistic actions and also to determine possible toxic effects. For diseases where drug repurposing is a popular approach, e.g., for neglected tropical diseases, MoA studies are specifically important since compounds were not originally designed to target the new disease type. Untargeted approaches to study

MoA are useful when compounds are suspected to have polypharmacy effects beyond known targets and to cluster compounds with the same MoA for improving the selection for further *in vivo* studies. Metabolomics offers a valuable approach to clustering compounds on MoA.¹

Neglected tropical diseases are a prime example where resistance to current treatments is problematic, funding is limited, and drug repurposing is popular. There is a requirement for novel strategies in the drug discovery pipeline

Received: March 2, 2018

Accepted: April 19, 2018

Published: April 19, 2018

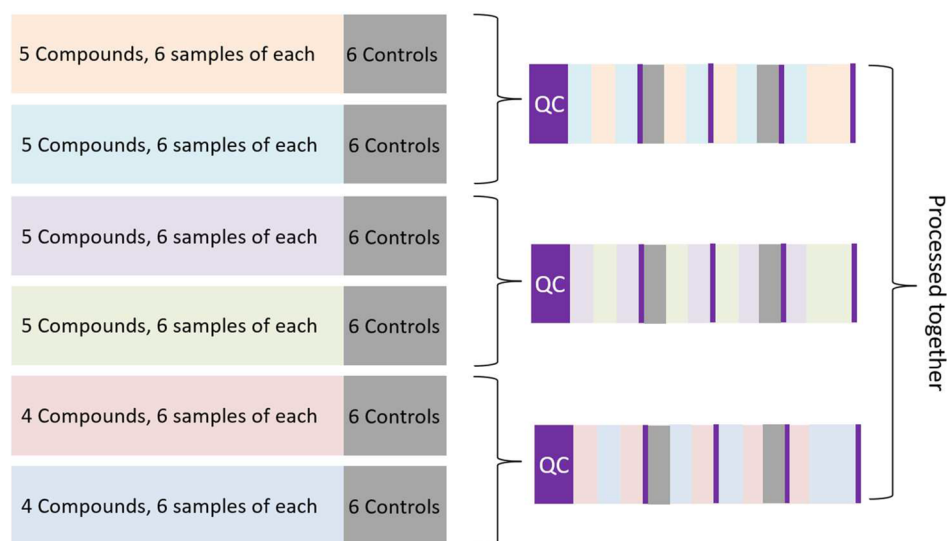


Figure 1. Overview of metabolomics experimental design. Samples were collected over six separate batches. Each batch consisted of six biological replicates of untreated axenic amastigotes and amastigotes treated with one of four or five different compounds in replicates of six. QC samples were prepared from a pool of extra control samples collected, and the same pool was used throughout the metabolomics experiment. Metabolomics was performed in three analytical batches of randomized samples from batch 1 and 2, then batch 3 and 4, and finally batch 5 and 6. All data were processed together as indicated.

for medium throughput screening that combines a balance on breadth and depth of knowledge on drug MoA. The leishmaniasis are a spectrum of neglected tropical diseases caused by protozoa of the genus *Leishmania*. *Leishmania donovani* provokes one of the most severe forms, that is, visceral leishmaniasis,² and existing therapeutic options for this are limited.³ From a recent screening of 1.8 million compounds against the three kinetoplastid parasites most relevant to human disease (*Leishmania donovani*, *Trypanosoma brucei*, and *Trypanosoma cruzi*), 192 noncytotoxic active hits against *Leishmania donovani* were selected to be included in the so-called *Leishmania* box.² While some general hypotheses were generated relating to compound structure, suggesting that many of them could target kinases, proteases, cytochromes, and host–pathogen interactions, the MoA of each is still unknown. Classification into MoA is an important element in analyzing activity data. To optimize prioritization and accelerate discovery, a strategy to cluster compounds based on more than chemical structure is required. Metabolomics and other hit-to-screen assays can be powerful tools in the analysis of MoA.⁴

The metabolomics approach to study the MoA of compounds for drug discovery purposes has been successfully applied in many fields. For recent reviews, see Vincent and Barrett⁵ for parasitology, Armitage and Southam⁶ for oncology, Rankin et al.⁷ for cardiology, Adamski⁸ for diabetes, Atzori et al.⁹ for perinatology, Gennari et al.¹⁰ for osteoporosis drug discovery, dos Santos et al.¹¹ for antibacterial MoA of plant derived products, Mikami et al.¹² for updates specific to MS-based metabolomics, and Hoerr et al.¹³ for updates specific to NMR-based metabolomics.

Studying compound MoA can be challenging, especially since it is difficult to distinguish drug effects from generic stress responses. A way to overcome this is to study many compounds in the same organism in parallel so that generic stress responses can be identified in all drug treated samples (therefore, not drug specific). Different approaches can be taken to study MoA, and the “omic” approaches can be particularly attractive due to

their medium-high-throughput screening capabilities combined with high sensitivity and coverage. Moreover, integration of omic data with *in silico* network analysis is a systems pharmacology approach that can be used to identify compound MOA on a multiscale.¹⁴

The metabolomics approach has recently been applied to study compounds of the Malaria box, another tropical disease with similar unmet medical needs.^{15,16} Metabolomics screening was applied to reveal the metabolic perturbations induced by 90 of the almost 30,000 compounds that were previously shown to selectively inhibit growth of cultured *P. falciparum* asexual red blood cell stages, in addition to samples treated with known antimalarials. The key features of the medium-high throughput screen were the use of the 96-well format and the use of high-sensitivity LC-MS to reproducibly detect 460 putatively annotated metabolites from a range of metabolic pathways. Though the number of compounds and the number of metabolites detected was high, authors of this study reported significant batch effects using this experimental design that were partially overcome by normalization of treated samples to untreated controls on each plate, but systematic variation was still observed in a subset of the drug treatments. Moreover, single doses of 1 μ M were studied for 5 h of exposure, irrespective of growth inhibition rates, meaning that some treatments did not elicit metabolic response under the conditions tested.

The *Leishmania* box contains a total of 192 compounds.² In the present research, lead compounds of the *Leishmania* box have been screened using metabolomics. Twenty-eight compounds (27 compounds or analogues from the box in addition to amphotericin B) have been studied in *Leishmania donovani* axenic amastigotes, chosen as the most relevant *in vitro* model of human leishmaniasis. Samples of parasites exposed to each compound were prepared in parallel with untreated control samples and analyzed using an untargeted metabolomics approach to reveal the similarities and differences in the metabolome following treatment. The dose of compound

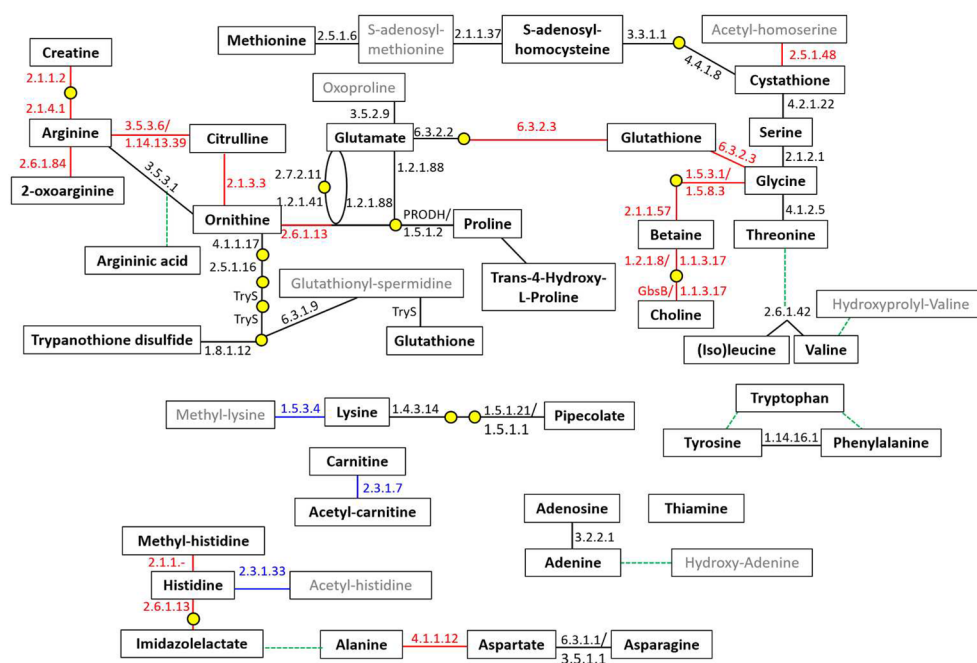


Figure 2. Metabolic network of identified metabolites in untreated *Leishmania donovani* axenic amastigotes. Metabolites in bold have been definitively identified to metabolomics standards initiative (MSI) level 1 through the analysis of authentic standards and standards spiked into samples. All other metabolites are annotated to MSI level 2. KEGG enzyme EC numbers are given for direct reactions. In black are those possible according to KEGG maps for *Leishmania donovani*. In blue are those possible specifically in this species according to KEGG, although the reactions are not provided in KEGG maps, and in red are reactions not known for *Leishmania donovani*. Yellow dots indicate proximal metabolites in the biochemical pathway to detected metabolites but that were not detected themselves.

and exposure time was chosen based on individual kill kinetics.^{1,14}

We aimed to reveal clusters of compounds with similar action against the parasite metabolome, as shown previously for antimalarials.¹⁵ The identification of clusters allows selection of compounds for further consideration in the drug discovery pipeline. Capillary electrophoresis–mass spectrometry (CE-MS) was chosen as the analytical tool to study the metabolomes of treated parasites, combined with definitive identification of the majority of the metabolic profile screened (metabolomics standards initiative (MSI) level 1¹⁸). Moreover, due to the scale of the study, analyses were performed in batches, and data were integrated for processing. As one of the largest scale metabolomics studies employing CE-MS, important strategies were identified in data treatment that build on previously observed limitations in metabolomics and could be useful in the field beyond the scope of this research.

2. RESULTS AND DISCUSSION

2.1. Assessment of Data Quality and Overview of Entire Analysis. Following filtration to remove features directly associated with specific compounds and therefore likely metabolites of the compounds themselves, in addition to filtration by QC RSD (keeping those features with RSD < 30%), 174 features remained, and data were assessed for quality and batch effect. [Supporting Information Figure 1S](#) shows an overview of the analyses from three analytical batches considering the internal standard signal, total useful signal, and total number of features (see also [Figure 1](#)). As shown, certain samples had particularly low numbers of features and total useful signal. These samples corresponded to five specific compound groups in addition to one anomalous sample from another group. Parasite numbers calculated for each sample

before and after washing were consulted to confirm that these lower profiles did not occur because of a lower parasite number in those samples. Trends in signal were observed in the internal standard and total useful signal. Three methods of normalization were performed to observe how data quality could be improved to remove this batch effect. [Supporting Information Figure 2S](#) shows the scores plots generated for the first two PCs before and after normalization by total useful signal, internal standard, and a commonly used method in metabolomics—locally estimated scatterplot smoothing (LOESS). Normalization by internal standard was deemed most appropriate since it did not skew the remaining samples based on the number of features present as total useful signal normalization did.

Batch effects are common and often unavoidable in large-scale studies.^{19,20} The challenge has been addressed previously, mainly for gas chromatography–mass spectrometry (GC-MS) and liquid chromatography–mass spectrometry (LC-MS) data,^{21,22} although to our knowledge it has not been addressed for CE-MS based metabolomics. Moreover, CE-MS is often discounted based on its reputation for irreproducibility, particularly in migration time, although advancements in technology and methodology are making CE-MS increasingly popular.²³ In our experience, careful choice of analytical method, experimental design, and studying of raw data to find the best parameters for alignment of analytical batches, makes CE-MS a robust and viable choice in multiple-batch studies, especially for ionic and polar metabolites where the alternative mechanism would be to use HILIC based LC-MS, that has a deeper complexity of issues surrounding robustness and reproducibility.²⁴

2.2. Identification of *Leishmania donovani* Axenic Amastigote Metabolic Profile. Before further multivariate analysis, identification of the entire CE-MS profile was

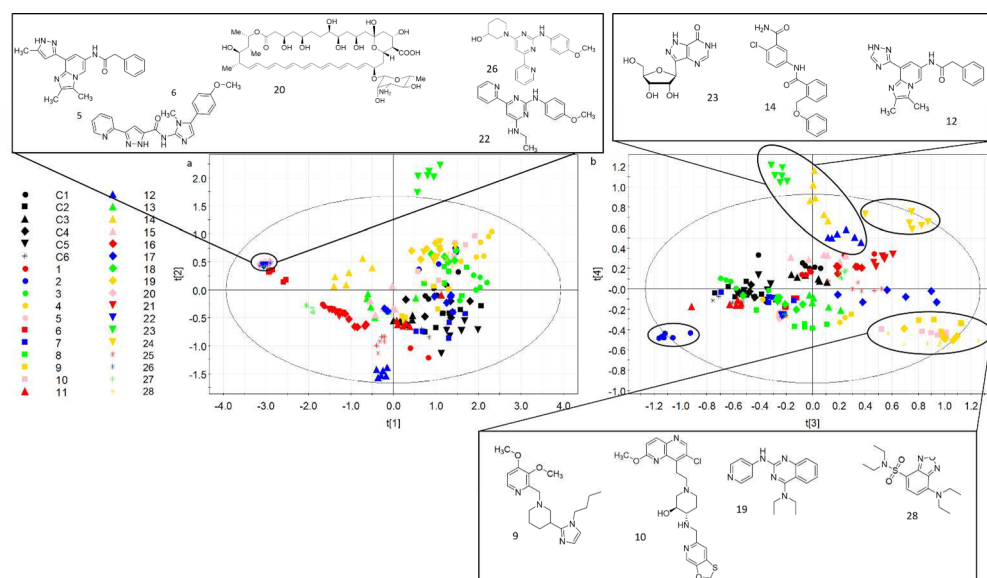


Figure 3. Principal components analysis (PCA) of compound clusters considering all groups. Explained variances for each component are PC1 ($R^2 = 0.49$), PC2 ($R^2 = 0.22$), PC3 ($R^2 = 0.13$), PC4 ($R^2 = 0.04$). Scores plots for PC1 vs PC2 are shown in a and for PC3 vs PC4 in b.

performed for untreated parasite samples. The 174 peaks following filtration and normalization were first annotated, and from this 105 were found to be unique features. The remaining features were identified as fragments, dimers, or artifacts of other metabolites present in the data and were therefore removed in all further multivariate analysis. All features passed filters for presence in untreated parasites and as such this identified profile serves as the first complete metabolic profile of *Leishmania donovani* axenic amastigotes in CE-MS. A total of 36 metabolites could be definitively identified to MSI level 1, determined through analysis of authentic standards, and a further 10 were identified to level 2. A network showing metabolic interactions is shown in Figure 2, where MSI level 1 identified metabolites are highlighted in bold. KEGG enzyme numbers are shown. Yellow dots indicate metabolites closely relating detected metabolites but that were not detected themselves. Supporting Information Table 2 details the experimental m/z , migration time, and MSI level of identification for all 105 uniquely distinguished metabolites.

2.3. Metabolic Clustering Analysis by Principal Components. Multivariate analysis of data using PCA can be challenging, especially when the experimental and biological complexity increases. For example, in the related study of metabolomics on Malaria box compounds, the first two PCs revealed only stochastic biological/experimental variation, while usable information was embedded in PC3 onward,¹⁵ representing a low proportion of total variability in the model. To explore this in our data, two approaches of metabolic clustering were employed to study the similarities and differences in parasites treated with one compound compared with untreated cells. These approaches together revealed complementary information on the likely MoA of different compound clusters that could be used to select a subset of compounds for further analysis in the drug discovery pipeline for leishmaniasis. In both cases, all features (identified, annotated, or unidentified, as detailed in Supporting Information Table 1) were included in the multivariate analyses. In the first approach, a further filter to keep only those features with RSD < 40% in untreated, control parasite samples was applied to reduce intragroup variation. This resulted in 94 metabolic features in total.

The first approach was through analysis of sequential principal components (PCs) to study the different degrees of variation among the compounds. The scores for the first four PCs are shown in Figure 3. The first PC in this model showed the variation due to total useful signal and number of features. This, as previously mentioned, was not associated with parasite number, nor to the analysis itself. Therefore, although it cannot be discounted since it accounts for almost 50% of the total variation in the model, it does not show the most informative separation for interpretation and is representative of stochastic biological/experimental variation as seen previously.¹⁵ The variation of biological interest is shown in PCs two, three, and four (collectively accounting for around 40% of the total variation). The idea is that this type of model can be used to assess the distance between different compounds on the scores plots to observe which compounds share similar effects on the metabolome of parasites and which are more unique. In terms of the drug discovery process for selecting a subset of candidates for further analysis, it may not be necessary to perform a deeper analysis of which metabolites contribute to the separation. The initial conclusions that can be drawn at this stage are that compounds 5, 6, 20 (amphotericin B), 22, and 26 form a very tight cluster of compounds exhibiting an extreme effect on CE-MS detected metabolites (decreasing most internal metabolites, mainly amino acids and derivatives). Relative to other experimental groups, the profiles were generally lowered in abundance, although on assessment of parasite number, this was not due to a lower number of parasites in the sample, which were counted at the time of harvesting and washing cells prior to extraction. It is known that the MoA of amphotericin B is to create pores in the membrane, and therefore it is likely that these other compounds cause a similar effect and the low profile observed in internal metabolites is due to leakage through these pores. It remains to be seen in a lipidomic analysis if there are any biochemical changes in the membrane associated with the MoA or if this is entirely structural.

Aside from those compounds which appear to damage the membrane, seemingly causing leakage of internal metabolites, this CE-MS based metabolomic approach allowed the

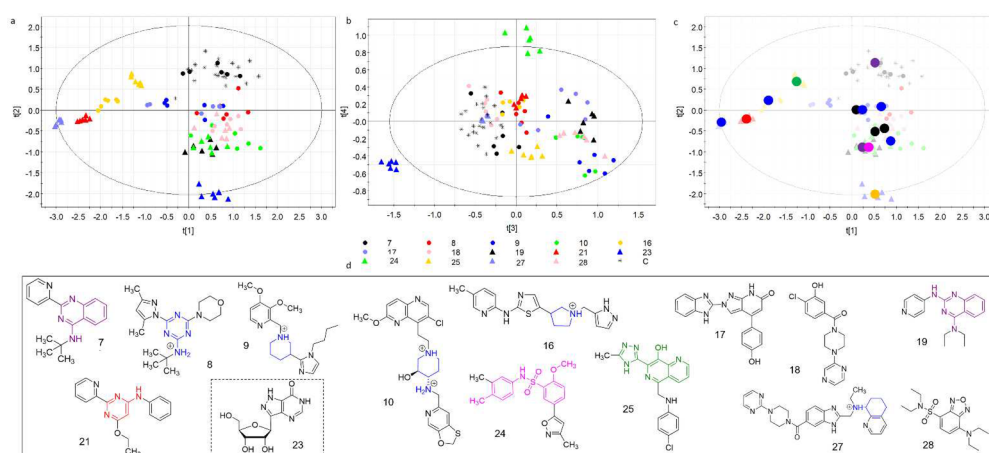


Figure 4. PCA of compound clusters after removal of sample groups from batch 1, batch 3, and compound groups yielding a reduced metabolic profile in terms of signal and number of features. Explained variances for each component are PC1 ($R^2 = 0.49$); PC2 ($R^2 = 0.22$); PC3 ($R^2 = 0.13$); PC4 ($R^2 = 0.04$). (a) Scores plot for PC1 vs PC2. (b) Scores plot for PC3 vs PC4. (c) Scores plot for PC1 vs PC2 with compound groups overlaid. (d) Compound structures with corresponding structural groups marked.

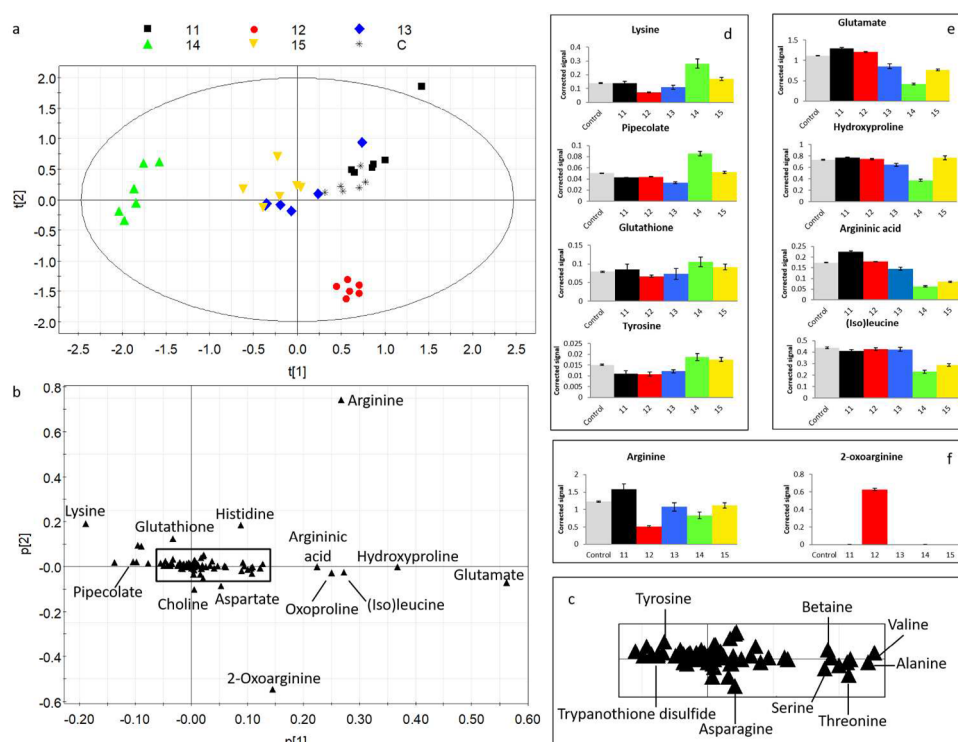


Figure 5. Batch 3 clustering. (a) Scores plot of PC1 ($R^2 = 0.44$) vs PC2 ($R^2 = 0.29$) from PCA analysis. (b) Loadings plot for PC1 vs PC2 from the same model. (c) Zoom of central region of loadings plot in b. (d) Column plots for discriminative metabolites in the negative loadings of PC1. (e) Column plots for discriminative metabolites in the positive loadings of PC1. (f) Column plots for discriminative metabolites in the loadings of PC2.

observations that compounds 9, 10, 19 and 28 form another cluster, compounds 2 and 24 seem to be unique, as does 23, although this has a closer neighbor in compound 14 which in turn shares similarities with compound 12. As shown in Figure 3, the metabolic clusters show little relation to the chemical structures of the compounds. These clusters were related neither to the LogP nor LogD (predicted values calculated at both pH 5.5 and 7.4 using Marvin Sketch version 15.5.25, ChemAxon) physicochemical properties of the compounds, as detailed in Supporting Information Table 2.

The second approach was to perform clustering analysis by deduction of sample groups. For example, those experimental

groups from parasites treated with compounds that resulted in low levels of all metabolites, presumed due to leakage if the MoA involves making pores in the membrane (compounds 5, 6, 20 (amphotericin B), 22, and 26), were removed from the analysis. Also, batches one and three that separated from the rest were removed to be analyzed separately. This left a total of 14 compound groups and related untreated controls that could be directly compared, and the 40% filter of RSD did not need to be applied. Figure 4 shows the results from PCA of these groups whereby clusters can be observed. To see whether compound clusters were related to structure, functional moieties were identified and colored in the structures to

classify them before comparing to the metabolic PCA. As can be observed, separation based on metabolic profile was not correlated with structure, further highlighting the potential of metabolomics as a genuine contender in the drug selection process. While the first method of clustering shown in Figure 3 offers a certain advantage in that all compounds can be simultaneously compared in one multivariate analysis, the deduction method allows ease of interpretation and eliminates the skew on the data created by outlying groups and removes the problem of stochastic biological/experimental variation that heavily contributes to variation in the model. Thus, although not all compounds can be directly compared using this deduction approach, the interpretation of this subset is much clearer. Of course, the fewer groups in a multivariate analysis, the clearer the separations become; however it is still valid to use this approach for three or more experimental groups. In the first two PCs representing the largest amount of variance between compounds, all show different profiles to the untreated controls except compound 7 (which was not so clear in Figure 3). As before, compound 23 is still observed as a singlet, and compound 24 separates further in PCs 3 and 4 although it is clustered with 8, 18, 10, 19, and 28 in the first two PCs. This indicates that these compounds share a similar effect on most metabolites, although in certain features (responsible for less of the total variation) compound 24 separates further from the rest of the cluster.

To observe the metabolite-by-metabolite differences for compounds in Figure 5, the abundance of each metabolite was plotted separately and used to aid interpretation of the figure. Supporting Information Figure 3S shows these data. Compound 23 was found to exhibit a metabolic profile quite different from all other groups. Relative to untreated controls and all other treatment groups, compound 23 causes a marked increase in the aromatic amino acids tyrosine, tryptophan, and phenylalanine (see Supporting Information Figure 4S). Likewise, it causes a marked increase in lysine and methyl-lysine, along with decreases in carnitine, acetyl-carnitine, and pipercolate, highlighting the activity of this compound in lysine and carnitine metabolism. On the other hand, compounds 16, 21, and 27 were observed as a cluster, having similar effects on the metabolism of parasites. Supporting Information Figure 5S shows the similar effect of these compounds on the metabolic pathways detected using CE-MS. As shown, key increases were S-adenosyl-methionine and S-adenosyl-homocysteine, while methionine itself was decreased. A decrease in arginine, citrulline, and ornithine was coordinated with an increase in 2-oxoarginine. A decrease in lysine was observed with a simultaneous increase in methyl-lysine, and a decrease in histidine was coordinated with an increase in acetyl-histidine.

To observe whether batches one and three that were previously removed could be directly comparable, PCA was performed on these batches, and results are shown in Supporting Information Figure 6S. There is a clear separation between untreated parasite controls for each, and one compound was completely irreproducible and therefore could not be used for clustering. This was used to confirm that batch three should be considered in isolation (results shown in Figure 5). Compounds 12 and 14 have unique effects. Compounds 13 and 15 are somewhat similar and do not differ much from the untreated parasites, nor does 11, which is largely similar to the untreated metabolome. The concentration of 2-oxoarginine distinguishes compound 12 from all other groups, while

compound 14 causes the relative increases in metabolites of Figure 5d while decreasing metabolites of Figure 5e.

2.4. Summary of Most Notable Features of Each Compound's MoA. Following multivariate analysis of clusters, each compound group was considered alone against its respective untreated control group to observe the MoA in more detail for each individually. Pairwise comparisons between treated and untreated samples were performed using Student's two tailed *t* test assuming unequal variance in addition to fold change and log 2 fold change calculations. Results from this are detailed in Supporting Information Table 3 and can be used as a resource to assist in further mechanistic studies and prioritization of these compounds for the development of potential antileishmanial drugs. To observe trends, log 2 fold changes are highlighted in pale red for increases and pale blue for decreases, and those fold changes of greatest magnitude (± 2 -fold) are highlighted by darker red and blue for increases and decreases, respectively.

As shown through the clustering, compound 7 caused very little variation from the untreated controls. Compounds 3 and 4 cause an increase in hydroxy-adenine and asparagine, though they do not cause significant decreases in any metabolite. These increases could be indicative of an oxidative stress response. Compound 1 causes the same 2-fold increase in hydroxy-adenine and, along with compound 2, causes 2–3-fold decreases in proline and aspartic acid. Compound 1 decreases several other metabolites, most notably related to arginine metabolism such as arginine, citrulline, and glutathionyl-spermidine. Conversely, proline is increased by more than 2-fold in compounds 8, 9, 10, 17, 18, 19, 23, 24, and 28, of which 10, 19, 23, and 24 simultaneously cause a greater than 2-fold decrease in glutamate, while 17 and 23 simultaneously decrease ornithine levels by more than 2-fold. Glutamate, proline, and ornithine are closely related. As shown in Figure 2, glutamate-5-semialdehyde (not detected in this study) can be produced from ornithine and 1-pyrroline-5-carboxylate (also not detected) from proline, both of which can be used to make glutamate *via* enzyme 1.2.1.88 (known as L-glutamate gamma-semialdehyde dehydrogenase or 1-pyrroline-5-carboxylate dehydrogenase). An increase in proline and coordinated decrease in glutamate and ornithine may suggest this enzyme as a target. Compounds 11 and 12 are centered on reducing ornithine and hydroxyprolyl-valine levels, while compound 11 also decreases proline and compound 12 arginine. Compounds 16 and 21 most notably increase acetyl-histidine and S-adenosylmethionine, while decreasing histidine and methyl-histidine, metabolites associated with polyamine-related metabolites (ornithine, glutathionyl-spermidine and arginine), aspartic acid, and hydroxyprolyl-valine. Compounds 14 and 15 both cause a 2–3-fold increase in acetyl-histidine, and compound 14 additionally increases lysine by more than 2-fold along with around 2-fold reductions in threonine, glycine, and cystathionine in addition to L-proline and glutamate related metabolites.

2.5. Conclusion. Using metabolomics, we have shown that compounds with different chemical structure and physicochemical properties can disturb the same metabolic pathways, while others with more similar structures can have different downstream effects. For that reason, novel approaches considering the effects of drugs in real biological or clinical settings could be highly valuable in the drug discovery pipeline, rather than selecting compounds based only on chemical structure. Here, we show the power of metabolomics in

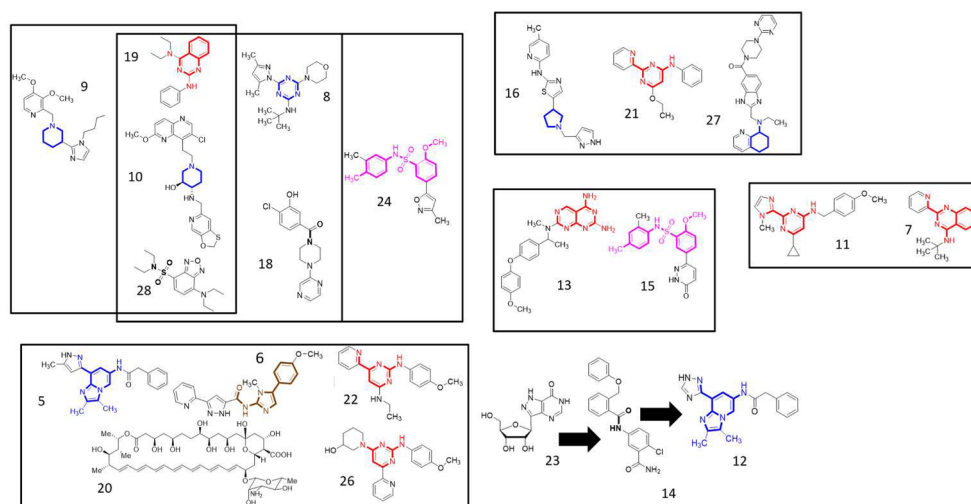


Figure 6. Compound clusters identified using the PCA clustering techniques described. Compounds 23, 14, and 12 were identified largely as singlets, although some metabolic similarity was found relating 23 to 14 and 14 to 12.

comparing different candidate treatments for leishmaniasis on endogenous metabolism. Two approaches in clustering using PCA have been assessed for 28 different compounds: 25 from GSKs *Leishmania* box² plus two box analogues and amphotericin B as a gold standard treatment. To be useful as a high-throughput screening mechanism, this metabolomics approach does not require biological interpretation of each compound's MoA that can be timely to perform. Identification of the metabolic profile to ensure separations of compound clusters are based on biological features is enough to enable the decision to take representative candidates from each cluster forward for further analysis in the drug discovery pipeline. That said, it is an advantage of this approach that this level of data generated can be used when required to study the MoA of any candidate further without requiring a different assay or a repeated assay to be performed. This highlights the potential of metabolomics over other types of assay which are neither so sensitive nor so comprehensive. In this research, a significant number of compounds have been simultaneously analyzed together using metabolomics. In doing so, it has been possible to assess different methods of data treatment in the combination of data from several analytical batches and to present the challenges and advantages of different approaches for the use of PCA to analyze the separation of groups, which can be useful in the omics wide community. Combining the information generated from metabolomics and the clustering techniques, the clusters depicted in Figure 6 were identified as sharing a similar metabolic response and can be useful in the next steps of prioritization of anti-leishmanial candidates.

3. MATERIALS AND METHODS

3.1. Compound Selection. Twenty-five compounds from the *Leishmania* box² in addition to two further analogues of this box and amphotericin B (chosen as a “gold standard treatment” for leishmaniasis) were selected for comparison by metabolomic analysis. Compounds were selected to cover a range of kill-kinetics (as determined for these compounds in Tegazzini et al.¹⁷), aiming for a wider scope of potential action within the *Leishmania* box. Based on these data,¹⁷ compounds showing the maximum activity by 6 h and 24 h as observed at 72 h were chosen to study the initial effects of them on the parasite metabolome. Compound concentrations were 1 and 2 times the EC₅₀ for compounds showing maximum response at 6 h and 24 h, respectively. For some slower acting compounds that only

reached maximum activity at 72 h, samples were taken at 24 h at two times the final observed EC₅₀. Even if the apparent final EC₅₀ observed at 72 h is not observed at the times the samples were collected, we decided to use the 72 h value as we assumed that the compound was already acting at those times even if the phenotypic response was not reached at the time samples were collected and that the apparent EC₅₀ observed at shorter times was due to slower MoA rather than slower distribution or binding kinetics of the compound. A summary of compounds selected for analysis is given in Supporting Information Table 1.

3.2. Sample Collection and Analysis by CE-MS. Samples were collected and analyzed using CE-MS as described. The overall design of the metabolomics experiment is depicted in Figure 1. Stratified randomization was used to decrease the batch effect on replicates of the same compound treatment. In addition to the samples shown, extra samples of untreated parasites were collected alongside each batch and pooled to make a quality control (QC) sample that was injected in each analytical batch as described for the analysis.

3.2.1. Chemicals and Reagents. The axenic culture medium used in all experiments was prepared as a single batch “in-house,” following the protocol described in Peña et al.² All methanol used was HPLC-grade, and formic acid was analytical grade. These chemicals in addition to formaldehyde solution and PBS were purchased from Sigma-Aldrich, as were authentic standards used in identification. Ultrapure water was obtained using a Milli-Qplus 185 system (Millipore, Billerica, MA, USA).

3.3. Sample Collection and Analysis by CE-MS. *Leishmania donovani* strain 1S2D (WHO designation: MHOM/SD/62/1S-CL2D)²⁵ was cultured by cycling between promastigotes and axenic amastigotes using protocols from prior work.² Amastigote forms were grown at 37 °C with 5% CO₂ in media adapted from De Rycker et al.²⁶ To prepare metabolomics samples for each group, *Leishmania donovani* axenic amastigotes were cultured in three T75 flasks each containing 30 mL of culture with an initial density of 6.67×10^6 parasites mL⁻¹, and either compound in DMSO (comparable volumes) or DMSO alone (untreated samples) was added at the respective concentrations stated for each (Supporting Information Table 1). After an incubation period of 5 h, two samples were obtained from each flask by equally dividing the culture into two 15 mL falcon tubes, resulting in six replicates for each group. Before the division, 50 μL of each culture was collected into Eppendorf tubes, to which 50 μL of formaldehyde was mixed with each and samples were stored at 4 °C to be counted later to record the exact number of parasites from each flask at the time of harvesting. At the time of harvesting and throughout the subsequent processes, samples were maintained at 4 °C.

After collection of culture into falcon tubes, samples were centrifuged at 1500g at 4 °C for 15 min, after which culture medium was decanted and parasites were resuspended in 2 mL of PBS. Parasites were washed in PBS (maintained at 4 °C) by gentle mixing, after which samples were transferred to 2 mL Eppendorf tubes. To record the exact number of parasites in each sample immediately before quenching, 10 μ L aliquots were collected at this stage, were fixed with 10 μ L of formaldehyde, and were stored at 4 °C to be counted later. Samples for metabolomics were subsequently centrifuged at 1500g at 4 °C for 15 min. PBS was decanted, and 200 μ L of ice cold methanol was added to each sample, which were immediately stored at -80 °C until extraction and metabolomics analysis. After storage of these samples for metabolomics analysis, samples collected for counting were analyzed using the CASY cell counter, and the total number of parasites in each sample was recorded.

3.3.1. Metabolite Extraction. On the day of each analysis (three days for each of the three analytical batches), metabolites were extracted, and resulting extracts were analyzed by CE-MS. Samples were evaporated to dryness using a speed vacuum concentrator (Eppendorf, Hamburg, Germany), after which 200 mg of 425–600 μ m acid-washed glass beads was added. Then, 575 μ L of 100% methanol was added. Samples were vortex mixed for 10 min before being placed in a tissue lyzer for 30 min at 50 Hz and were then centrifuged at 16,000g at 4 °C for 10 min. After subtraction of 80 μ L from each sample to be stored for later analyses, 165 μ L of water was added; samples were vortex mixed for 30 min and finally centrifuged at 16,000g at 4 °C for 10 min. Following centrifugation, supernatants were collected into Eppendorf tubes for CE-MS analysis. These were evaporated to dryness using the speed vacuum concentrator, and once dry, 100 μ L of water containing 0.2 mM methionine sulfone used as an internal standard and 0.1 M formic acid was added to each. On the first extraction day, all samples collected to make the QC pool were extracted individually as described for all other samples, before being pooled and finally aliquoted for use in each of the three analytical batches. Extraction blanks were prepared following all steps of the extraction.

3.3.2. Analysis of Extracts by CE-MS. Analyses were conducted over three analytical batches as shown in Figure 1. Each batch started with the injection of extraction blanks and 10 QCs, then samples injected randomly with QC injections after every sixth sample injection. At the end of each analytical batch, a representative of each biological group was reinjected at a higher voltage to induce a higher degree of in-source fragmentation to be used later in metabolite identification, as described in Godzien et al.²⁷

The instrument consisted of capillary electrophoresis (7100 Agilent) coupled to a TOF Mass Spectrometer (6224 Agilent) equipped with an ESI source, whereby the CE mode was controlled by ChemStation software (B.04.03, Agilent) and MS mode by Mass Hunter Workstation Data Analysis (B.02.01, Agilent). The separation occurred in a fused-silica capillary (Agilent; total length, 100 cm; i.d., 50 μ m). All separations were carried out in normal polarity with a background electrolyte containing 0.8 M of formic acid solution in 10% methanol (v/v) at 20 °C. In our laboratory, new capillaries are preconditioned with a flush of 1.0 M NaOH for 30 min followed by Milli-Q water for 30 min and background electrolyte for 30 min (although only one capillary was used in the analysis of all samples for this research). Before each analysis, the capillary was conditioned with a flush of background electrolyte for 5 min. The sheath liquid (6 μ L min⁻¹) was MeOH/H₂O (1:1) containing 1.0 mM formic acid with two reference masses: m/z 121.0509 ([C₅H₄N₄ + H]⁺) and m/z 922.0098 ([C₁₈H₁₈O₆N₃P₃F₂₄ + H]⁺), which allowed correction and higher mass accuracy in the MS. Samples were hydrodynamically injected at 50 mBar for 50 s. The stacking was carried out by applying background electrolyte at 100 mBar for 10 s. The separation voltage was 30 kV. The internal pressure was 25 mBar, and the analyses were carried out in 35 min. The MS parameters were fragmentor, 100 V; skimmer, 65 V; octopole, 750 V; drying gas temperature, 200 °C; flow rate, 10 L min⁻¹; and capillary voltage, 3500 V. Data were acquired in positive mode with a full scan from m/z 85 to 1000 at a rate 1.41 scan

s⁻¹. For the samples run at higher voltage, all parameters described here were maintained the same, except for the fragmentor voltage that was set at 175 V.

3.4. Data Treatment and Feature Identification. Data from all batches were processed together using recursive analysis in Mass Hunter Profinder (B.06.00, Agilent) software. Data were reprocessed considering ions [M + H]⁺, [M + Na]⁺, and [M + K]⁺ and neutral water loss. The maximum permitted charge state was double. Alignment was performed based on m/z and RT similarities within the samples. Parameters applied were 1% for the RT window and 20 ppm for mass tolerance. These were selected based on assessment of raw data of all three analytical batches to align all data together into one set.

Data were filtered to remove those features consistently present in certain compound samples (one to three groups at most) and consistently absent in all others, which were presumed to be masses representing individual compounds and their metabolites rather than molecules from the parasite metabolome. All remaining features were present in all untreated control samples in addition to QCs. These data were refiltered based on relative standard deviation (RSD) in QCs to retain only those with less than 30% RSD. A total of 174 features remained.

Identification was performed across the entire profile of 174 features by searching m/z against Metlin (<http://metlin.scripps.edu>) and considering the same adducts as those described for data reprocessing. Putative identities were assigned to m/z values for metabolite features considering (i) mass accuracy (maximum mass error 10 ppm), (ii) isotopic pattern distribution, (iii) possibility of ion formation, and (iv) adducts formation. For as many metabolites as possible, authentic standards were analyzed, both separately and spiked into quality control samples to definitively identify them. All features identified as fragments, dimers, or ringing artifacts as described in Godzien et al.²⁷ were removed from the data set. This resulted in a peak table of 105 features, including those definitively identified with standards (MSI level 1), those remaining putatively identified (MSI level 2), and those unidentified (MSI level 4), according to the metabolomics standards initiative.¹⁸

Data normalization was performed separately for comparison, by total useful signal (dividing each signal by the sum of signals for each of the features for the corresponding sample) and by dividing each signal by the signal analyzed for the internal standard in each sample. Metabolic clustering analysis was performed using principal components analysis (PCA) in SIMCA-P 12.0 software (Umetrics, Umea, Sweden).

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acchembio.8b00204.

Supporting Figures (PDF)

Supporting Tables (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +34 913724711. E-mail: cbarbas@ceu.es.

ORCID

Emily G. Armitage: 0000-0001-6948-3854

Coral Barbas: 0000-0003-4722-491X

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge funding provided by the Tres Cantos Open Lab Foundation (TCOLF: program code TC132). In addition, the authors at Universidad San Pablo CEU acknowl-

edge Spanish Ministerio de Economía y Competitividad (CTQ2014-55279-R), and M.P.B. is funded by a core grant to the Wellcome Centre for Molecular Parasitology (104111/Z/14/Z).

ABBREVIATIONS

MoA, mode of action; CE-MS, capillary electrophoresis–mass spectrometry; MSI, metabolomics standards initiative; QC, quality control; PCA, principal components analysis; GC-MS, gas chromatography–mass spectrometry; LC-MS, liquid chromatography–mass spectrometry

REFERENCES

- (1) Kurita, K. L., Glassey, E., and Lington, R. G. (2015) Integration of high-content screening and untargeted metabolomics for comprehensive functional annotation of natural product libraries. *Proc. Natl. Acad. Sci. U. S. A.* *112*, 11999–12004.
- (2) Peña, I., Pilar Manzano, M., Cantizani, J., Kessler, A., Alonso-Padilla, J., Bardera, A. I., Alvarez, E., Colmenarejo, G., Cotillo, I., Roquero, I., de Dios-Anton, F., Barroso, V., Rodriguez, A., Gray, D. W., Navarro, M., Kumar, V., Sherstnev, A., Drewry, D. H., Brown, J. R., Fiandor, J. M., and Julio Martin, J. (2015) New compound sets identified from high throughput phenotypic screening against three kinetoplastid parasites: an open resource. *Sci. Rep.* *5*, 8771.
- (3) Pedrique, B., Strub-wourgaft, N., Some, C., Olliaro, P., Trouiller, P., Ford, N., Pécoul, B., and Bradol, J. (2013) The drug and vaccine landscape for neglected diseases (2000 – 11): a systematic assessment. *Lancet Glob. Heal.* *1*, 371–379.
- (4) Powers, R. (2014) The Current State of Drug Discovery and a Potential Role for NMR Metabolomics The Current State of Drug Discovery and a Potential Role for NMR Metabolomics. *J. Med. Chem.* *57*, 5860–5870.
- (5) Vincent, I. M., and Barrett, M. P. (2015) Metabolomic-Based Strategies for Anti-Parasite Drug Discovery. *J. Biomol. Screening* *20*, 44–55.
- (6) Armitage, E. G., and Southam, A. D. (2016) Monitoring cancer prognosis, diagnosis and treatment efficacy using metabolomics and lipidomics. *Metabolomics* *12*, 146.
- (7) Rankin, N. J., Preiss, D., Welsh, P., and Sattar, N. (2016) Applying metabolomics to cardiometabolic intervention studies and trials: past experiences and a roadmap for the future. *Int. J. Epidemiol.* *45*, 1351–1371.
- (8) Adamski, J. (2016) Key elements of metabolomics in the study of biomarkers of diabetes. *Diabetologia* *59*, 2497–2502.
- (9) Atzori, L., Griffin, J., Noto, A., and Fanos, V. (2012) Review Metabolomics: A New Approach to Drug Delivery in Perinatology. *Curr. Med. Chem.* *19*, 4654–4661.
- (10) Gennari, L., Rotatori, S., Bianciardi, S., Gonnelli, S., Nuti, R., and Merlotti, D. (2015) Appropriate models for novel osteoporosis drug discovery and future perspectives. *Expert Opin. Drug Discovery* *10*, 1201–1216.
- (11) dos Santos, B. S., da Silva, L. C. N., da Silva, T. D., Rodrigues, J. F. S., Grisotto, M. A. G., dos Santos Correia, M. T., Napoleão, T. H., da Silva, M. V., and Paiva, P. M. G. (2016) Application of omics technologies for evaluation of antibacterial mechanisms of action of plant-derived products. *Front. Microbiol.* *7*, 1–13.
- (12) Mikami, T., Aoki, M., and Kimura, T. (2012) The Application of Mass Spectrometry to Proteomics and Metabolomics in Biomarker Discovery and Drug Development. *Curr. Mol. Pharmacol.* *5* (2), 301–316.
- (13) Hoerr, V., Duggan, G. E., Zbytnuik, L., Poon, K. K. H., Große, C., Neugebauer, U., Methling, K., Löffler, B., and Vogel, H. J. (2016) Characterization and prediction of the mechanism of action of antibiotics through NMR metabolomics. *BMC Microbiol.* *16*, 1–14.
- (14) Zhao, S., and Iyengar, R. (2012) NIH Public Access. *Annu. Rev. Pharmacol. Toxicol.* *52*, 505–521.
- (15) Creek, D. J., Chua, H. H., Cobbold, S. A., Nijagal, B., Macrae, J. I., Dickerman, B. K., et al. (2016) Metabolomics-Based Screening of the Malaria Box Reveals both Novel and Established Mechanisms of Action. *Antimicrob. Agents Chemother.* *60*, 6650–6663.
- (16) Allman, E. L., Painter, H. J., Samra, J., Carrasquilla, M., and Llinás, M. (2016) Metabolomic profiling of the Malaria Box reveals antimalarial target pathways. *Antimicrob. Agents Chemother.* *60*, 6635–6643.
- (17) Coteron, J. M., Tegazzini, D., Cantizani, J., Pena, I., and Martin, J. (2017) Unravelling the rate of action of hits in the Leishmania donovani box using standard drugs amphotericin B and miltefosine. *PLoS Neglected Trop. Dis.* *11*, e0005629.
- (18) Fiehn, O., Robertson, D., Griffin, J., van der Werf, M., Nikolau, B., Morrison, N., Sumner, L. W., Goodacre, R., Hardy, N. W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., van Ommen, B., Lindon, J. C., and Sansone, S. A. (2007) The metabolomics standards initiative (MSI). *Metabolomics* *3*, 175–178.
- (19) Wehrens, R., Hageman, J. A., van Eeuwijk, F., Kooke, R., Flood, P. J., Wijnker, E., Keurentjes, J. J. B., Lommen, A., van Eekelen, H. D. L. M., Hall, R. D., Mumm, R., and de Vos, R. C. H. (2016) Improved batch correction in untargeted MS-based metabolomics. *Metabolomics* *12*, 88 DOI: 10.1007/s11306-016-1015-8.
- (20) Hendriks, M. M. W. B., Eeuwijk, F. A. va., Jellema, R. H., Westerhuis, J. A., Reijmers, T. H., Hoefsloot, H. C. J., and Smilde, A. K. (2011) Data-processing strategies for metabolomics studies. *TrAC, Trends Anal. Chem.* *30*, 1685–1698.
- (21) Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B., and Goodacre, R. (2011) Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* *6*, 1060–1083.
- (22) Fernández-Albert, F., Llorach, R., Garcia-Aloy, M., Ziyatdinov, A., Andres-Lacueva, C., and Perera, A. (2014) Intensity drift removal in LC/MS metabolomics by common variance compensation. *Bioinformatics* *30*, 2899–2905.
- (23) Ramautar, R., and de Jong, G. J. (2014) Recent developments in liquid-phase separation techniques for metabolomics. *Bioanalysis* *6*, 1011–26.
- (24) García, A., Godzien, J., López-González, Á., and Barbas, C. (2017) Capillary electrophoresis mass spectrometry as a tool for untargeted metabolomics. *Bioanalysis* *9*, 99–130.
- (25) Dwyer, D. M. (1976) Antibody-Induced Modulation of Leishmania Donovanii Surface Membrane Antigens Information about subscribing to The Journal of Immunology is online at: Antibody-Induced Modulation of Leishmania donovani Surface Membrane Antigens. *J. Immunol.* *117*, 2081–2091.
- (26) De Rycker, M., Hallyburton, I., Thomas, J., Campbell, L., Wyllie, S., Joshi, D., Cameron, S., Gilbert, I. H., Wyatt, P. G., Frearson, J. A., Fairlamb, A. H., and Gray, D. W. (2013) Comparison of a high-throughput high-content intracellular Leishmania donovani assay with an axenic amastigote assay. *Antimicrob. Agents Chemother.* *57*, 2913–2922.
- (27) Godzien, J., Armitage, E. G., Angulo, S., Martinez-Alcazar, M. P., Alonso-Herranz, V., Otero, A., Lopez-Gonzalez, A., and Barbas, C. (2015) In-source fragmentation and correlation analysis as tools for metabolite identification exemplified with CE-TOF untargeted metabolomics. *Electrophoresis* *36*, 2188–2195.