Metabolomic systems biology of trypanosomes

MICHAEL P. BARRETT1*, BARBARA M. BAKKER2 and RAINER BREITLING3

1 Faculty of Biomedical and Life Sciences and Wellcome Trust Centre of Molecular Parasitology, University of Glasgow, Glasgow Biomedical Research Centre, Glasgow G12 8TA, United Kingdom
2 Department of Pediatrics, Center for Liver, Digestive and Metabolic Diseases, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands
3 Groningen Bioinformatics Centre, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands

(Received 26 October 2009; revised 11 January 2010; accepted 11 January 2010; first published online 17 February 2010)

SUMMARY

Metabolomics analysis, which aims at the systematic identification and quantification of all metabolites in biological systems, is emerging as a powerful new tool to identify biomarkers of disease, report on cellular responses to environmental perturbation, and to identify the targets of drugs. Here we discuss recent developments in metabolomic analysis, from the perspective of trypanosome research, highlighting remaining challenges and the most promising areas for future research.

Key words: Metabolomics, mass spectrometry, Orbitrap, ab initio networks, Trypanosoma brucei

INTRODUCTION

The African trypanosome, Trypanosoma brucei, the causative agent of human African trypanosomiasis (Barrett et al. 2003), has in recent years emerged as a front runner in systems biology analysis. Reproducible cultivation methods exist and quantitative analysis is relatively well developed. For example, a quantitative mathematical model of energy metabolism in the long slender form of the trypanosome (i.e. the form that replicates in the mammalian bloodstream) has been developed (Bakker et al. 1997) and iteratively updated after experimental testing (Bakker et al. 1999a,b; Albert et al. 2005; Haanstra et al. 2008a). Moreover, a model of the gene expression cascade, based on quantitative knowledge of transcription, RNA precursor degradation, trans-splicing and mRNA degradation for phosphoglycerate kinase (PGK) has recently been generated (Haanstra et al. 2008b). It is therefore possible to envisage the derivation of a fully comprehensive molecular model of trypanosome physiology. Another paper in this volume details the birth of the “silicon trypanosome” project and applications of dynamic modelling to understand function in the cells. Here we focus on the experimental approaches to metabolome analysis that are required to sustain this ambitious endeavour.

ACQUISITION OF METABOLOMICS DATA

In recent years, methods have become available for the rapid acquisition of whole genome data, from which it is possible to infer the entire repertoire of metabolically active enzymes present within a cell. Transcriptome analysis can reveal which genes are expressed at a given time, under a given set of environmental conditions. Proteomic techniques, although less able to achieve comprehensive coverage of expressed proteins, can reveal if specific proteins are indeed present under given conditions. Metabolomics, which aspires to the simultaneous measurement of all low molecular weight chemicals within a cell, has been less easy to achieve, mainly because of the wide range of physicochemical properties characteristic of cellular metabolites and the huge dynamic range over which these are present (Van der Werf et al. 2005; Breitling et al. 2006a). However, a variety of techniques are available to measure cellular metabolites. These include nuclear magnetic resonance (NMR) and a variety of mass spectrometry-based approaches, such as mass spectrometry coupled to liquid chromatography (LC-MS), gas chromatography (GC-MS), or capillary electrophoresis (CE-MS). Different types of mass spectrometer can be used to identify the masses of metabolites separated through the different chromatography platforms. Alternatively, nuclear magnetic resonance spectrometry (NMR) is used in highly quantitative analysis and also for identifying metabolites that are difficult to ionise or which are difficult to analyse by MS, due to their idiosyncratic behaviour on different chromatographic platforms. The most promising recent development is the combination of MS and

* Corresponding author: Faculty of Biomedical and Life Sciences and Wellcome Trust Centre of Molecular Parasitology, University of Glasgow, Glasgow Biomedical Research Centre, Glasgow G12 8TA, United Kingdom. Tel: 0141-330-6904. Fax: 0141-330-4600. Email: m.barrett@bio.gla.ac.uk
NMR technology in hybrid instruments that unite the sensitivity of MS and the superior quantitative capacity and specificity of NMR (Yang, 2006).

Relatively little previous work on identification and quantification of metabolites in *T. brucei* has been performed. The glycolytic intermediates have been identified as part of the programme investigating energy metabolism using HPLC (Visser and Oppordoes, 1980) or enzymatic analysis (Visser et al. 1980; Albert et al. 2005; Haanstra et al. 2008a). Polyamines and the thiol intermediates of the trypanothione pathway have also been characterized using HPLC-based methodology (Fairlamb et al. 1987; Shim and Fairlamb, 1988; Xiao et al. 2009). NMR has been used to trace steady-state concentrations of abundant phosphorylated compounds within *T. brucei* (Moreno et al. 2000) and also to assess the abundant end products of metabolism in procyclic trypanosomes from which key enzymes have been deleted in order to assess likely pathways of metabolism in this form of the parasite (Rivièvre et al. 2004; Coustou et al. 2005, 2008). Similar studies were also used to assess end products of glucose metabolism in bloodstream forms (Mackenzie et al. 1983).

It is imperative that metabolism is quenched as rapidly as possible prior to extracting metabolites. Directly applying trypanosomes in suspension to boiling ethanol achieves this (Kamleh et al. 2008a), although it is necessary also to measure medium without cells to distinguish intracellular and extracellular metabolic content. Rapidly chilling cells in culture in a dry ice-ethanol bath, followed by centrifugation, washing and extraction in chloroform-water-methanol, as used for Leishmania (Robinson et al. 2007; Ruben T’Kind, personal communication) offers advantages in terms of allowing removal of extracellular medium and gentler extraction conditions than boiling ethanol. Continued development of novel extraction methods will enhance metabolite coverage.

ADVANTAGES OF ULTRA-HIGH RESOLUTION, ULTRA-HIGH MASS ACCURACY MASS SPECTROMETRY

We have recently adopted a platform of ultra-high resolution, ultra-high mass accuracy mass spectrometry to systematically acquire metabolite data in *T. brucei* (Breitling et al. 2006a, b, 2008). Modern Fourier transform instruments can measure the mass of analytes to within 1 ppm, so a metabolite like glucose (or any of its structural isomers; Mw = 180.0634) can be distinguished from a metabolite of similar mass such as theophylline (Mw = 180.0647) in a manner that less accurate machines could not achieve. The problem of structural isomers, however, requires further analysis to resolve: e.g. fructose, mannose, galactose, inositol have the same mass as glucose. Orthogonal approaches such as chromatographic behaviour or analysis of metabolite fragments in GC-MS or tandem mass spectrometry are required to resolve this issue.

In a first study, trypanosome extracts were generated from parasites isolated from rats and separated as a buffy coat. The extracts were directly injected into a Fourier transform ion cyclotron resonance mass spectrometer. On the order of 1,000 peaks were detected (Breitling et al. 2006a). Many of these were associated with the *T. brucei* lipidome, and it is possible that the highly ionisable nature of these metabolites suppressed the ionisation of many non-lipid metabolites (in spite of using a range of different extraction solvents to recover a balanced set of metabolites). Notwithstanding these problems, the initial datasets were rich enough to enable the inference of large networks of metabolites that could be linked through potential metabolic or chemical links (Breitling et al. 2006a). Such *ab initio* networks reveal the advantage of using high accuracy–high resolution approaches in which each metabolite can be assigned an exact mass: since every possible biochemical transformation (phosphate addition, hydroxylation, methylation, prenylation etc.) will alter the size of a metabolite by a unique exact mass, each metabolite measured in a dataset can be related to others through the simple addition or subtraction of those characteristic metabolic masses (Breitling et al. 2006a, 2008; Rogers et al. 2009; Fig. 1). This *ab initio* network building technique has a major role in the inference of novel metabolic pathways (Jourdan et al. 2008).

Magnetic sector ion cyclotron resonance instruments, as used in that first study, are relatively expensive and complex to operate. The development of the Orbitrap mass spectrometer, which can achieve similar levels of accuracy and resolution, represented an important advance in metabolomics research (Breitling et al. 2006; Lim et al. 2007; Ding et al. 2007; Dunn et al. 2008; Kamleh et al. 2008a, b; Kiefer et al. 2008). Furthermore, coupling Orbitrap mass spectrometry to a range of chromatographic systems, most importantly HILIC columns that resolve hydrophilic compounds, has greatly enhanced our ability to identify many polar cellular metabolites (Kamleh et al. 2008a; Cubbon et al. 2009).

HILIC chromatography linked to an Orbitrap has now become our standard approach to trypanosome metabolomics, and early runs have shown that hundreds of metabolites can be successfully identified through this route, including signature trypanosome metabolites such as trypanothione (Fig. 2). Using these data it has been possible to develop algorithms to enhance the accuracy of mass spectra by calibration based on background ions constantly detected in the chromatograms (Scheltema et al. 2008). Furthermore, software that can remove mass spectrometry artefacts including
transformations. In (B) the central panel, the network is extended by identifying further metabolites that can be linked through a list of the mass differences associated with known biochemical, metabolic, and regulatory events. In (C) the central panel, the network is extended by identifying further metabolites that can be linked to the connected masses by multiscale metabolic transformations (second level network), and by (C) the network has been extended by six of these iterations to identify potential connectivity between the observed masses.

Fig. 2. Extracted mass chromatogram for trypanothione.

common adducts (Na⁺, K⁺, NH₄⁺, Cl⁻, etc.), identify isotopomers (based on ¹³C abundance) and compare spectra along time series, have been applied to identify those peaks that represent true metabolites (Scheltema et al. 2009). This approach allows the identification of several hundred metabolites. However, this coverage is significantly lower than the thousand or so expected for most cells. Given the chemical diversity of the small chemical species that comprise the metabolome, comprehensive coverage
can only be achieved if the same sample is exposed to a range of different analytical platforms, including different types of liquid chromatography and solvents, as well as gas chromatography-MS approaches. The latter platform has underpinned much of the pioneering work in metabolomics of Leishmania (Doyle et al. 2009), and also offers a valuable tool for trypanosome research.

**ANALYSIS OF METABOLIC FLUX AND DYNAMICS**

The ability to generate static views of the metabolome is now available. However, much of what we wish to understand about the system must come from dynamic views of how metabolites flow through the metabolic network. In the case of the glycolytic pathway, a combination of modelling and experimentation has shown that the control of flux through the pathway is distributed between several steps including the glucose transporter (THT1), which exerts some 40% of the flux control (Bakker et al. 1999b), and most likely glyceraldehyde 3-phosphate dehydrogenase, glycerol phosphate dehydrogenase, phosphoglycerate mutase and the mitochondrial glyceral-3-phosphate oxidase (Albert et al. 2005). In contrast, in mammals, hexokinase and phosphofructokinase assume most control over the pathway.

A great deal of further work is required to refine even these current relatively simple models. For example, kinetic data are usually obtained from purified enzymes in relatively simple buffer systems, that by definition lack much of the complexity that will govern kinetic behaviour in vivo. Teusink et al. (2000) have shown in yeast that such differences between in vivo and in vitro parameters currently limit our capability to explain pathway behaviour ab initio from enzyme kinetic properties. Time-dependent metabolic patterns have a much richer information content than steady-state data and will be vital to bridge the gap between in vitro biochemistry and the in vivo behaviour of metabolic pathways (Hynne et al. 2001). Metabolomics approaches, particularly those using mass spectrometry, now offer the possibility to measure flux through multiple pathways simultaneously (Weichert et al. 2007; Tang et al. 2009; Niittylae et al. 2009). In addition to providing a means of testing mathematical models of metabolism, these approaches can provide data that might even allow the prediction of kinetic parameters (or at least a range of parameters that could account for observed behaviour) (Kümmel et al. 2006; Hynne et al. 2001). Such model improvement will be important to make better predictions of the adaptive capacity of metabolism towards environmental changes, such as the addition of drugs and/or the signals that trigger parasite differentiation.

Metabolomics approaches will also reveal the relative importance of parallel routes in the network. Pulse-chase experiments can follow the distribution of a precursor metabolite through the metabolome in time. For example, by adding a pulse of heavy atom-labelled glucose (with one or more of the carbon atoms in glucose labelled with $^13$C) the accumulation of the heavy isotope in different metabolites of the glycolytic and pentose phosphate pathways can be measured. In our rapid-stop protocol (Kamleh et al. 2008a), where trypanosome metabolism is quenched instantaneously by squirting trypanosomes in suspension into boiling ethanol, we can, in principle, make repeated measurements at very short intervals (down to seconds), to measure the dynamic distribution of label over time. The same principle can be applied to other pathways; for example, the flux of labelled cysteine, glutamate and glycine into glutathione and trypanothione pathway intermediates. This ability to actually measure fluxes will play an invaluable role in learning about those parameters that control dynamic flux through the metabolome.

**GENOME-WIDE MODELS**

Existing kinetic models of trypanosome metabolism describe relatively small parts of the complete metabolic network, while the metabolomics techniques discussed in this paper aim at covering a large part of the metabolic network. They are usually interpreted by mapping the results onto genome-based reconstructions of the metabolic potential of the organism of interest. This is, for instance, possible using the Pathway Tools software (Karp et al. 2002) that allows interactive interrogation of the results and has already proven useful for many organisms. A genome-wide reconstruction of the Leishmania metabolic network (LeishCyc) has recently been published (Doyle et al. 2009). A T. brucei reconstruction (TrypanoCyc) is also being made (Chukualim et al. 2008; Jourdan et al. unpublished). The KEGG environment (Kanehisa et al. 2008) also represents a good background in which to represent measurements. Software that allows direct projection of high accuracy metabolomics datasets onto the KEGG contextual overview is available (Suhre and Schmitt-Kopplin, 2008).

In addition to the simple, static views of the metabolome provided by the MetaCyc and KEGG environments, other computational techniques, that have been developed to study the metabolic potential at a genome scale in a more dynamic fashion, are available (Breitling et al. 2008; Feist et al. 2009). Constraints-based models based on stoichiometric matrices have become widely used (Edwards et al. 2001) and have recently been generated for Leishmania (Chavali et al. 2008) and T. cruzi (Roberts et al. 2009). These can assist in drug target discovery by identifying essential enzymes whose absence is catastrophic to cellular functioning (Jamshidi and Palsson, 2007).
In the long run, the ambitious aim of this work is to help identify those parts of metabolism most amenable to targeting by novel drugs.

ACKNOWLEDGMENTS

MPB gratefully acknowledges funding from the BBSRC in the “SysStryp” consortium funded jointly by the BBSRC and the ANR. RB is supported by a NWO-Vidi fellowship. BMB is supported by a Rosalind Franklin Fellowship at the University of Groningen.

REFERENCES


