TrypanoCyc: a community-led biochemical pathways database for *Trypanosoma brucei*

Sanu Shameer[1], Flora J. Logan-Klumpler[2], Florence Vinson[1], Ludovic Cottret[3],
Benjamin Merlet[1], Fiona Achcar[4], Michael Boshart[5], Matthew Berriman[2], Rainer Breitling[6],
Frédéric Bringaud[7], Peter Bütikofer[8], Amy M. Cattanach[4], Bridget Bannerman-Chukualim[2],
Darren J. Creek[9], Kathryn Crouch[4], Harry P. de Koning[4], Hubert Denise[10],
Charles Ebikeme[11], Alan H. Fairlamb[12], Michael A. J. Ferguson[12], Michael L. Ginger[13],
Christiane Hertz-Fowler[14], Eduard J. Kerkhoven[15], Pascal Mäser[16], Paul A. M. Michels[17],
Archana Nayak[4], David W. Nes[18], Derek P. Nolan[19], Christian Olsen[20],
Fatima Silva-Franco[14], Terry K. Smith[21], Martin C. Taylor[22], Aloysius G. M. Tielens[23,24],
Michael D. Urbaniak[13], Jaap J. van Hellemond[24], Isabel M. Vincent[4], Shane R. Wilkinson[25],
Susan Wyllie[12], Fred R. Opperdoes[26], Michael P. Barrett[4,*] and Fabien Jourdan[1,*]

[1]Institut National de la Recherche Agronomique (INRA), UMR1331, TOXALIM (Research Centre in Food Toxicology), Université de Toulouse, Toulouse, France, [2]The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK, [3]Institut National de la Recherche Agronomique (INRA), UMR441, Laboratoire des Interactions Plantes-Microorganismes (LIPM), Auzeville, France, [4]University of Glasgow, Glasgow, Scotland, G12 8QQ, UK, [5]Ludwig-Maximilians-Universität München, Biocenter, 82152-Martinsried, Germany, [6]Manchester Institute of Biotechnology, Faculty of Life Sciences, University of Manchester, Manchester, UK, [7]CNRS, Bordeaux, 33076, France, [8]University of Bern, Bern, CH-3012, Switzerland, [9]Monash Institute of Pharmaceutical Sciences, Monash University, Parkville 3052, Australia, [10]European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, CB10 1SD, UK, [11]ISSC, UNESCO, F-75732 CEDEX 15, Paris, France, [12]University of Dundee, Dundee, Scotland, DD1 4HN, UK, [13]Divisionof Biomedical and Life Sciences, Lancaster University, Bailrigg, Lancaster, LA1 4YG, UK, [14]University of Liverpool, Liverpool, Merseyside L69 3BX, UK, [15]Chalmers University of Technology, Kemivägen 10, 412 96, Göteborg, Sweden, [16]Swiss Tropical and Public Health Institute, Socinstr. 57, Basel 4051, Switzerland, [17]University of Edinburgh, Mayfield Road, Edinburgh EH9 3JU, UK, [18]Texas Tech University, Lubbock, TX, USA, [19]Trinity College Dublin, Dublin 2, Ireland, [20]Biomatters Inc. 60 Park Place, Suite 2100, Newark, NJ, USA, [21]University of St Andrews, St Andrews, Scotland, KY16 9ST, UK, [22]LSHTM, London, WC1E 7HT, UK, [23]Utrecht University, Utrecht, 3508 TD, The Netherlands, [24]Erasmus University Medical Center, Rotterdam, 3015 CE, The Netherlands, [25]Queen Mary University of London, London E1 4NS, UK and [26]University of Louvain, Brussels, B-1200, Belgium

ABSTRACT

The metabolic network of a cell represents the catabolic and anabolic reactions that interconvert small molecules (metabolites) through the activity of enzymes, transporters and non-catalyzed chemical reactions. Our understanding of individual metabolic networks is increasing as we learn more about the enzymes that are active in particular cells under particular conditions and as technologies advance to allow detailed measurements of the cellular metabolome. Metabolic network databases are of increasing importance in allowing us to contextualise data sets emerging from transcriptomic, proteomic and metabolomic experiments. Here we present a dynamic database, TrypanoCyc (http://www.metexplore.fr/trypanocyc/), which describes the generic and condition-specific metabolic network of *Trypanosoma brucei,* a parasitic protozoan responsible for human and animal African trypanosomiasis. In addition to enabling navigation through the

*To whom correspondence should be addressed. Tel: +33 561 28 57 15; Fax: +33 561 28 52 44; Email: Fabien.Jourdan@toulouse.inra.fr
Correspondence may also be addressed to Michael P. Barrett. Tel: +44 141 330 6904; Fax: +44 141 330 4077; Email: michael.barrett@glasgow.ac.uk

BioCyc-based TrypanoCyc interface, we have also implemented a network-based representation of the information through MetExplore, yielding a novel environment in which to visualise the metabolism of this important parasite.

INTRODUCTION

Trypanosoma brucei is the causative agent of African trypanosomiasis (commonly known as sleeping sickness in humans and Nagana in animals). The disease is fatal if untreated in humans [1] and the economic impact of trypanosomes on agriculture in Africa is immense. The drugs available for the trypanosomiases are inadequate for a number of reasons and better therapeutic options are required [2]. Many drugs work through interfering with enzymes involved in cellular metabolism. The only anti-trypanosomal drug whose target is known is eflornithine, an inhibitor of ornithine decarboxylase [3], a key enzyme in the polyamine biosynthetic pathway. A comprehensive understanding of parasite metabolism therefore contributes to current efforts in drug discovery and understanding drug resistance [4].

Global untargeted molecular profiling data sets (e.g. trancriptomics, proteomics and metabolomics data) are now being generated for trypanosomes and the effects of life cycle, environmental perturbation, specific genetic manipulation and drug action are being dissected in a systematic manner [5]. Interpreting and integrating these data to allow biological inference and hypothesis generation is a major challenge. Metabolic network-based methods offer a means to contextualise and integrate data to help inference of biological function [6]. Reliable, comprehensive databases collating information on metabolic networks and pathways are therefore crucial to optimize understanding derived from postgenomic data sets. Metabolic databases such as LeishCyc [7] (for *Leishmania*), and the Library of Apicomplexan Metabolic Pathways [8] (for apicomplexan parasites) are among the few examples where this information is available in parasitology.

Creation of a metabolic network database is achieved by gathering information on all of the metabolic transformations an organism can perform [9]. A first outline of this information is generally retrieved from genomic orthology. Genes coding for enzymes are identified through sequence similarity searches and then, using enzyme activity information, metabolic reactions catalyzed by these enzymes are added to the network database. Several automatic and semiautomatic tools are available to perform these genome-based metabolic network reconstructions [10–13].

In spite of their undoubted utility, genome-based reconstruction has limitations since it is based primarily on sequence homology comparisons between the organism of interest and databases encompassing information from a multitude of organisms [14]. Incorrect annotations readily propagate across databases [15]. Moreover, evolution works through modification of function following alteration of genes encoding proteins. For instance, trypanosomes use N^1,N^8 bis-glutathionyl spermidine (trypanothione) [16] as a key cellular redox-associated metabolite. Trypanothione is retained in its reduced form by the enzyme trypanothione reductase (EC 1.8.1.12). This enzyme is evolutionarily derived from glutathione reductase (EC 1.8.1.7), with which it shows great homology. In the absence of accompanying biochemical evidence, genome annotations would simply predict trypanosomes as possessing a glutathione reductase, and metabolic reconstructions would assume trypanosomes employ canonical glutathione-based redox balancing. Cases like this highlight the necessity of refining genome-based metabolic reconstructions by incorporating advanced biochemical knowledge [15].

Moreover, simple genome reconstructions do not take into account the sub-cellular localization of the enzymes (although various methods are now being developed to tackle this issue as canonical signals determining cellular localization come to light [17]). Finally, the genome provides a view of the total metabolic capability of an organism, regardless of environmental and genetic conditions. In trypanosomes, however, different metabolic strategies are used at different points in the life cycle. In the tsetse fly, the trypanosome's main carbon source is proline [18] while in the human-host it is glucose [19]. Some reactions are active in one condition but not in another. This information is particularly important when looking for potential drug targets.

Web servers such as KEGG [20] and BioCyc [14] represent metabolism as a set of pathways, reflecting classical textbook views of biochemistry. However, the pathway approach fragments metabolism in ways which constrain our ability to decipher the broader impact on the metabolic network; hence, methods that also enable connected network views of metabolism are desirable. We have therefore combined building a pathway-based TrypanoCyc database with its integration into the MetExplore web server [21], to offer both pathway and network-based inference and visualization.

A HIGHLY CURATED DATABASE OF *T. BRUCEI* METABOLISM

The *T. brucei* TREU 927 genome is 26 Mb in size, with a karyotype of 11 megabase chromosomes [22] and containing a predicted 9068 protein-coding genes. In a collaborative project between the International Trypanotolerance Centre in The Gambia and the Sanger Institute, the genome sequence was processed using the Pathologic metabolic network reconstruction tool of Pathway Tools [23], creating a Pathway/Genome Database (PGDB) where gaps (called 'pathway holes') in the predicted metabolic pathways were filled by hypothetical reactions, even without an obvious gene association. The result of this first automatic reconstruction was the starting point of the current TrypanoCyc database.

An international consortium of investigators, expert in various aspects of trypanosome metabolism, was assembled to produce a highly annotated TrypanoCyc database. As recommended by Thiele & Palsson [24] we started the TrypanoCyc initiative in 2012 with a two-day 'jamboree'. Each expert was offered a specific set of pathway(s) in his/her area of expertise to curate. A dedicated web interface, called TrypAnnot (a password protected part of the website available to annotators, not described here) stores submitted annotations in a curation database, making it possible to track all annotations, which are automatically taken from
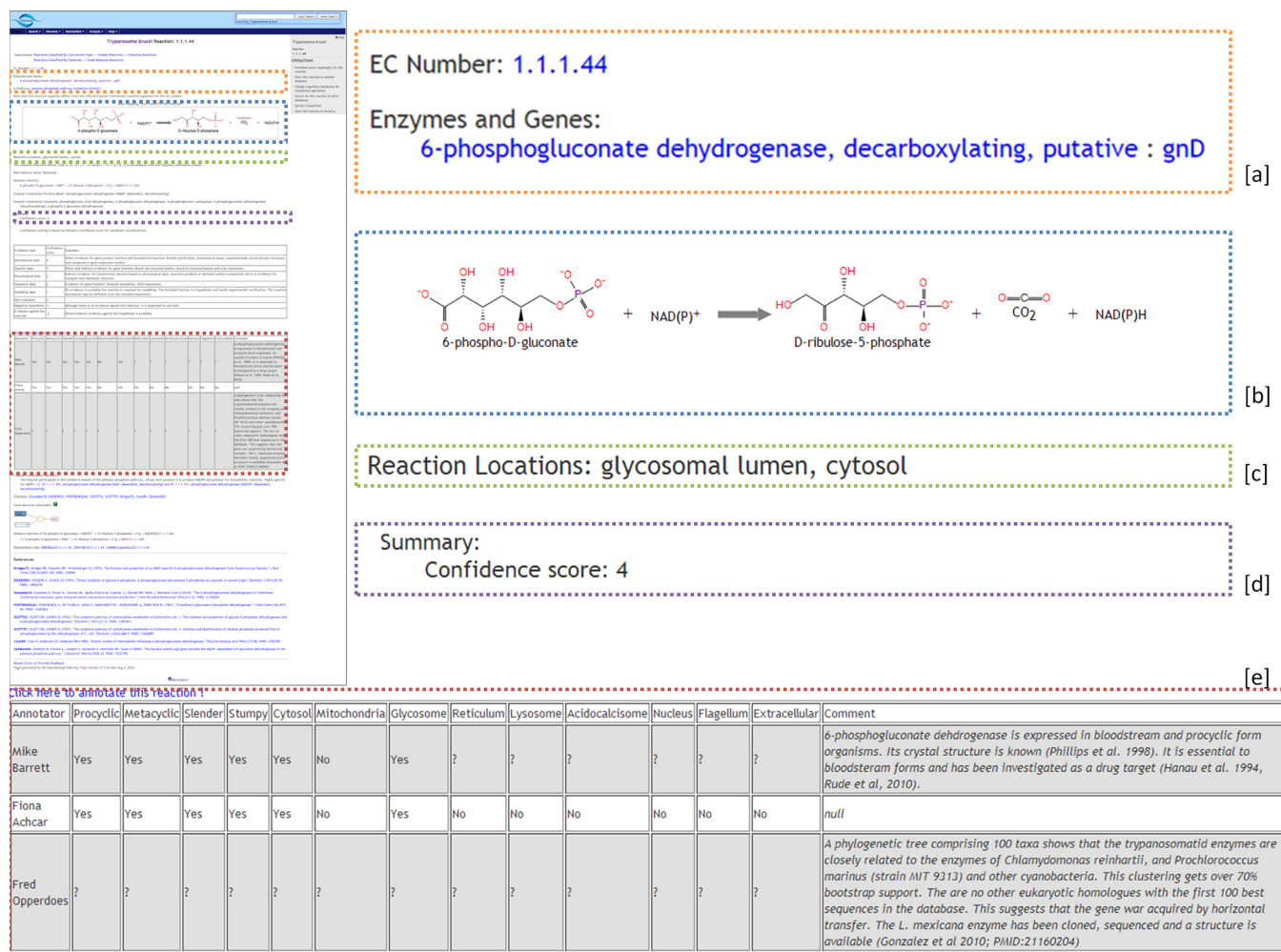
Figure 1. TrypanoCyc page for the 6-phosphogluconate dehydrogenase (1.1.1.44) reaction. (**a**) Reaction name and GeneDB link (specific to TrypanoCyc), (**b**) Detailed description of the reaction, (**c**) Localizations of the reactions as suggested by annotators, (**d**) Confidence score for the reaction (specific to TrypanoCyc), (**e**) Annotation tables displaying content of the TrypAnnot database (specific to TrypanoCyc).

the database and added to the web page of the corresponding reaction. The TrypanoCyc project has so far had 1368 editing events, among which are 653 annotations made on 464 reactions. Furthermore, since the first automated reconstruction in 2008, 17 pathways, 35 enzymatic-reactions, 10 transport reactions, 41 enzymes, 2 protein complexes and 104 metabolites have been added to TrypanoCyc. Extended summaries for some pathways have also been made available in the database.

T. brucei cells contain multiple membrane-bounded organelles, including the mitochondrion and an unusual peroxisome-related organelle, the glycosome (25,26), in which the first seven steps of glycolysis occur, as well as a series of other pathways (19). Annotators, therefore, specify the sub-cellular localization of reactions, if known, in the annotation interface. Life cycle stage specificity for each reaction is also important, since trypanosomes use different metabolic pathways in different environments; hence annotators can specify one or more developmental stages in which reactions occur. Note that this information is not available in the reconstruction provided by KEGG (see Ta-

ble 1 for comparison). The level of knowledge on each reaction varies from experimentally verified to indirect evidence of activity regardless of manual curation. To reflect the level of confidence of the annotation we have used the scoring system proposed by Thiele & Palsson (9) (see Table 2). For instance, of the 464 annotated reactions, 84 were annotated based on direct evidence from protein purification, biochemical assays or comparative gene expression studies and hence can be considered with the highest confidence. During curation we found numerous falsely predicted reactions and pathways; 60 pathways, 14 enzymatic reactions, 20 enzymes and 56 metabolites have been removed from the original reconstruction. Nevertheless we retained some reactions if they are known to occur in related trypanosomatids, or else when they have been proposed to exist, erroneously, in the literature. Although such reactions are kept, they are not linked to any pathway and they are assigned a negative confidence score to highlight the fact that according to our present knowledge they are not actually present. For example, a methionine cycle that regenerates methionine from methylthioadenosine resulting from polyamine biosynthe-
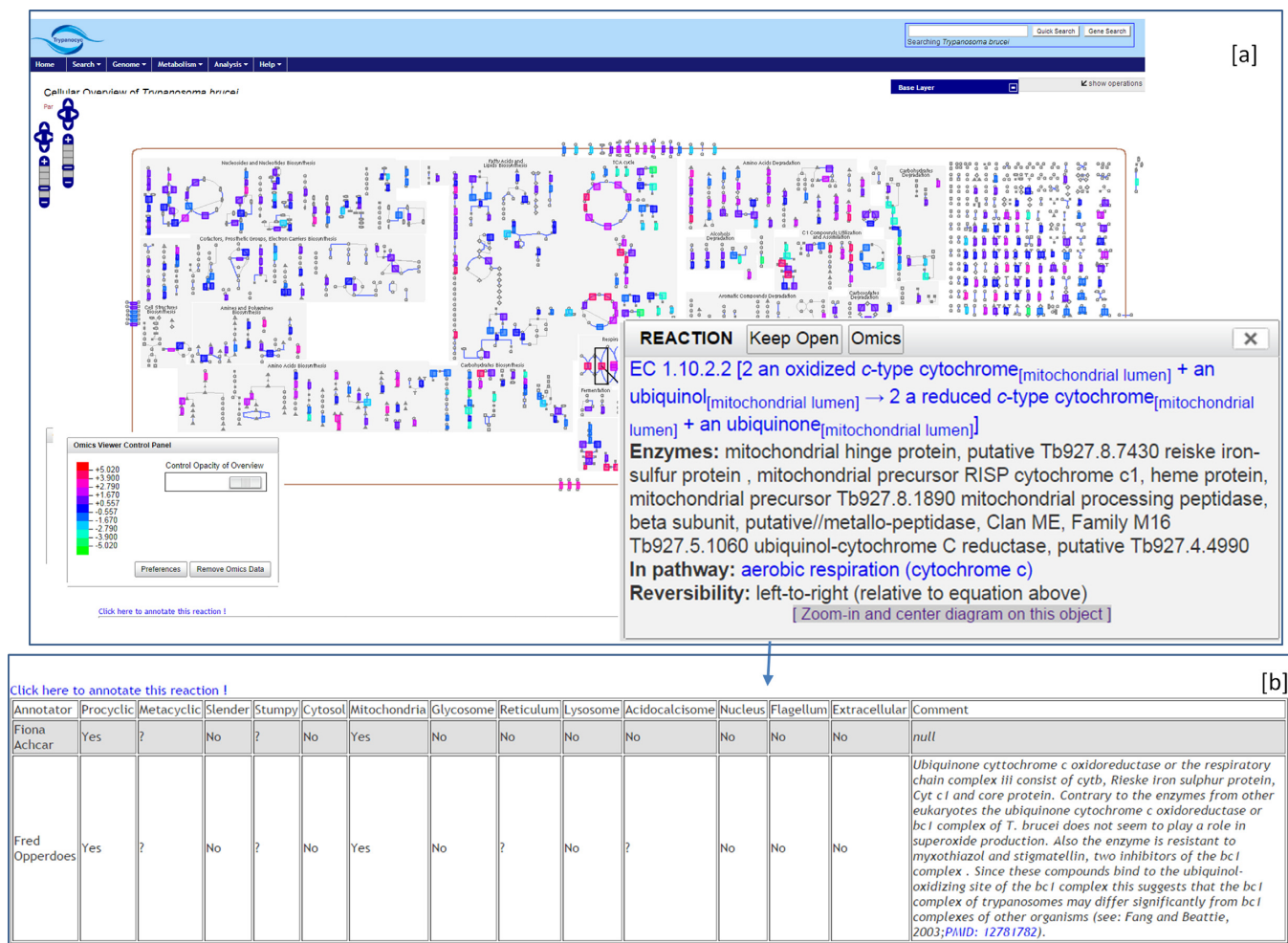
Figure 2. Proteomics data loaded in TrypanoCyc using the cellular overview tool. (**a**) The diagram shows all the metabolic pathways in gray boxes. Colored squares correspond to reactions with associated proteomics values. The color scale is displayed in the 'Omics Viewer Control Panel'; it can be tuned using dedicated parameters. The 'REACTION' dialog appears when clicking on a reaction. (**b**) It is then possible to get back to the corresponding reaction page and read the annotators' comments.

sis has been proposed (27). However, metabolic labelling experiments have subsequently indicated that the pathway is not active in trypanosomes, at least in the conditions used (28). The reactions EC 4.2.1.109 (methylthioribulose 1-phosphate dehydratase) and EC 3.1.3.77 (5-(methylthio)-2,3-dioxopentyl-phosphate phosphohydrolase), required to complete the pathway, are included in the database, but assigned negative scores to highlight that they are undetectable in spite of previous predictions in the literature (27). We consider it useful to keep such entries such that users of the database can find explicit reference to these reactions they might seek upon reading literature pertaining to these reactions.

Since metabolic databases focus mainly on pathways and seldom consider sub-cellular compartments, they usually lack information on intracellular transport reactions. Currently, TrypanoCyc contains only 35 such reactions. This is because we did not incorporate transport reactions into our annotation platform and because experimental knowledge on intracellular transport processes is still sparse. However, the dynamic nature of TrypanoCyc means additional anno-

tation and incorporation of measured and probable transport reactions (e.g. taken from existing manually curated metabolic models of the closely related organism *Leishmania major* (29)) will form part of the iterative process of database refinement. We also perform gap filling in each compartment using graph approaches and testing metabolic scenarios as suggested (9,10) and successfully implemented for other organisms (30).

Many additional databases provide information that can complement metabolic network databases. Linking to these other data sources enhances our ability to learn about an organism's metabolism. TrypanoCyc, therefore, links to multiple databases including BRENDA (31,32), expasy.org (33), ExplorEnz (34), Pubmed and UniProt (35). The TritrypDB database (36) is the central resource for trypanosomatid genomes and associated functional genomics data, while GeneDB houses the sequence information gathered and annotated through the Wellcome Trust Sanger Institute (37). For each gene, TrypanoCyc offers a direct link to the corresponding TritrypDB and GeneDB pages.
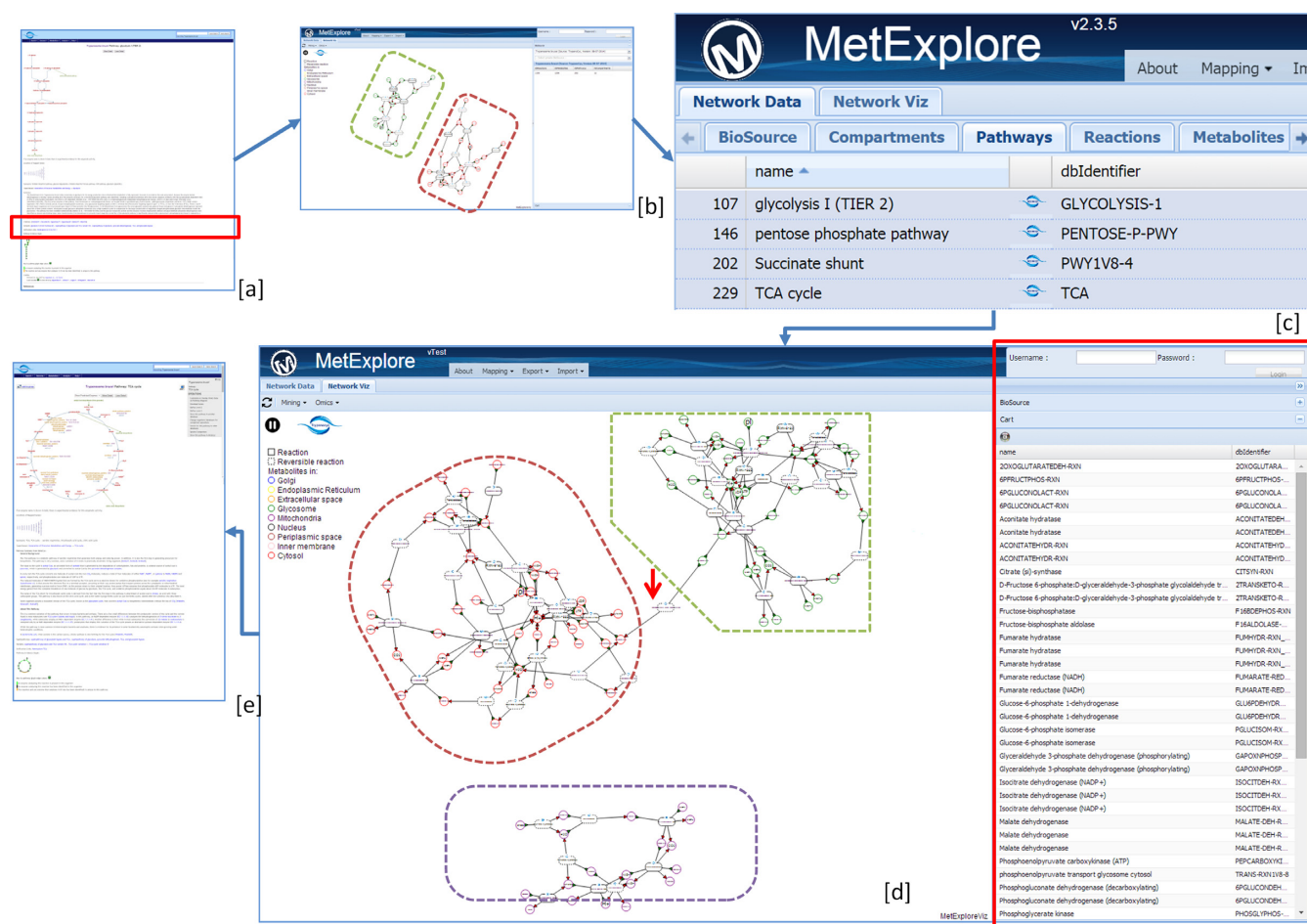
Figure 3. Navigation between pathway and network representation using MetExplore and TrypanoCyc. (**a**) Each pathway page has an hyperlink allowing to load and visualize the pathway in MetExplore (circled in red on the pathway page screenshot). (**b**) When clicking on this link in the Glycolysis page, it is loaded in MetExplore; the red box corresponds to the cytosolic part and the green one to the glycosomal part. (**c**) Using MetExplore, it is then possible to generate a combination of various pathways. TCA cycle, succinate shunt, glycolysis and the pentose phosphate pathway were selected. (**d**) All reactions of these pathways are added to the cart (red box on the right). A third compartment, mitochondrion, appears (purple box). A reaction allowing transport between cytosol and glycosome appears in the network (red arrow). (**e**) In the tabular view of MetExplore, a TrypanoCyc button (visible in the third column of [c] table) allows to link back to TrypanoCyc.

Table 1. Description of the confidence score system used in TrypanoCyc to evaluate the level of curation of each reaction

Reconstruction	Compartments	Life cycle stages	Pathways	Enzymatic reactions	Unique metabolites
Draft reconstruction 2008	1	0	238	1120	796
KEGG August 2014	1	0	61	656	646
TrypanoCyc August 2014	9	4	209	1025	842

The BioCyc library is a collection of 3563 PGDBs. Based on the quality of the PGDBs and the level of manual curation, this central repository classifies them into Tier 1 (highly curated), Tier 2 (moderately curated) and Tier 3 (non-curated) categories. Prior to the release of BioCyc v18.1, only 6 PGDBs (EcoCyc (38), MetaCyc (14), HumanCyc (39), AraCyc (40), YeastCyc and LeishCyc (7)) were published in the Tier 1 category. Due to the quality of information being made available on TrypanoCyc, it was included in BioCyc's Tier 1 category with the release of BioCyc v18.1 in June 2014.

REACTIONS, PATHWAYS AND NETWORK MINING

Browsing TrypanoCyc content and expert annotations

As a Pathway Tools-based website, TrypanoCyc provides a dedicated web page for each metabolic network entity (pathways, reactions, metabolites, enzymes, proteins and genes). The reaction page architecture was, however, modified in order to allow additional annotation information. These include the annotation confidence score (Figure 1d), stage specificity and compartmentation with links to key literature (Figure 1e). A comment box is also included, containing detailed free-text information on the reaction. Fig-

Table 2. Overview of TrypanoCyc content before and after curation and comparison with the KEGG database

Evidence type	Confidence score	Description
Biochemical data	4	Direct evidence for gene product function and biochemical reaction: protein purification, biochemical assays, experimentally solved protein structures and comparative gene-expression studies.
Genetic data	3	Direct and indirect evidence for gene function: knock-out characterization, knock-in characterization and over expression.
Physiological data	2	Indirect evidence for biochemical reaction based on physiological data: secretion products or defined medium components serve as evidence for transport and metabolic reactions.
Sequence data	2	Evidence for gene function: genome annotation, SEED annotation.
Modelling data	1	No evidence is available but reaction is required for modelling. The included function is a hypothesis and needs experimental verification. The reaction mechanism may be different from the included reaction(s).
Not evaluated	0	
Negative hypothesis	–1	Although there is no evidence against this reaction, it is expected to not exist
Evidence against the reaction	–2	Direct/indirect evidence against the hypothesis is available

ure 1 shows the webpage for the pentose phosphate pathway enzyme, 6-phosphogluconate dehydrogenase (EC. 1.1.1.44).

Search requests on database content can be made through a quick search box found at the top right-hand corner of the interface page, as well as through the advanced search options available from the menu bar. Each pathway representation is available with different levels of detail, the simplest view displaying only the reactions and metabolites while the detailed view displays all available information including the molecular structure of all metabolites involved. Additionally, for every pathway in TrypanoCyc, we provide a link to visualize the pathway in MetExplore.

Mining stage specific metabolism using cellular overview

To exemplify the integration of molecular profiling data in the TrypanoCyc database we used published results from a Stable Isotope Labelling of Amino acids in Cell culture experiment, comparing protein levels in bloodstream form (BSF) and procyclic form (PCF) trypanosomes (41). The data set contains 3552 gene IDs along with their relative protein levels in the two tested stages of *T. brucei* (expressed as log PCF/BSF values). A TrypanoCyc cellular overview shows enzymes that differ in abundance between the two life cycle forms (Figure 2; for step by step instructions see Supplementary Data S1).

Mapping other molecular profiling data in TrypanoCyc can be achieved using the Pathway Tools Omics Viewer (42), which displays all pathways in a single representation. Data sets can be loaded using the options listed on the right-hand side of the page (Supplementary Material S2 is a version of this data set in an Omics Viewer-compliant format). Figure 2a shows an image of the overview after loading the proteomics data of (41). Individual reactions can be viewed by moving the mouse over them and clicking the link in the pop-up dialog box. This opens the related reaction page containing the annotation table, giving access to specific

TrypanoCyc annotators' comments about the enzyme and its activity as well as the generic information pertaining to that reaction in the MetaCyc database. For example, the overlaid data clearly show that the respiratory chain is up-regulated in procyclic stages. Browsing the reaction page of any of those up-regulated proteins shows additional information from the annotators. For example, for ubiquinone-cytochrome C reductase (EC 1.10.2.2), two TrypanoCyc annotators report that this reaction is active in the PCF but not in the long slender BSF of *T. brucei* (see Figure 2b), thus agreeing with the observations from the proteomics experiment.

Using MetExplore to create user-defined sub-networks from TrypanoCyc

To complement the classical pathway-oriented BioCyc representation of data, we also offer a novel way to visualize the content of TrypanoCyc *via* our MetExplore web server (21) (for step by step instructions see Supplementary Data S3). Each pathway page contains a hyperlink (Figure 3a), that opens MetExplore with the selected pathway (Figure 3b). Importantly, the MetExplore viewer takes into account the localization of reactions. For example, Figure 3b shows how the glycolytic pathway is divided into two compartments (glycosome and cytosol represented by green and red boxes, respectively).

Another advantage of MetExplore is that it provides a tabular representation of compartments, pathways, reactions, enzymes, genes and metabolites in the database. It is also possible to filter these tables by compartments, pathways or reactions. For instance, by filtering simultaneously on the pentose phosphate pathway, TCA cycle, succinate shunt and glycolysis, only reactions and metabolites related to these pathways are displayed in their respective tables (Figure 3c). The user can also add reactions of interest to a 'cart' (red box on Figure 3d). It is then possible to visualize

the content of this cart in the network representation. From Figure 3d, it is evident that the network perspective is much more effective in representing compartments and transport reactions. Furthermore, the glycosome (green box) and cytosol (red box) are demonstrably connected by a reaction involved in the succinate shunt (marked by a red arrow on Figure 3d). For a more flexible representation MetExplore also offers a downloadable version of the Cytoscape visualization software (43), pre-loaded with the cart content.

Finally, each MetExplore reaction/pathway with a description in TrypanoCyc has been hyperlinked to the corresponding reaction/pathway pages, allowing the user to go back to the expert annotations anytime (Figure 3e).

CONCLUSION

Since 2012, TrypanoCyc has been under extensive curation with the help of the scientific community and is now counted among the seven Tier 1 databases within the BioCyc repository. Collaborative annotations help in improving the quality of the database by reducing errors, reducing the workload for individual annotators and also providing inferences from multiple perspectives given the various types of experts in the community.

T. brucei metabolic plasticity allows the parasite to adapt to divergent nutritional environments offered by different hosts. For drug target identification, for example, focusing on enzymes and metabolic pathways expressed in the parasite-stages that are replicative in the mammalian host is critical. TrypanoCyc is the first comprehensive metabolic network database for parasites including stage specificity as a key component of the collected data. LeishCyc (7), for the related parasite *L. major*, has also been established, and in the future these two databases should, ideally, be linked, given the significant degree of similarity in the metabolic networks of these evolutionarily related parasites.

TrypanoCyc and the related annotation database allow anyone with an interest to join the annotation team. The size of the consortium helps guarantee the sustainability of TrypanoCyc as does the involvement of permanent staff both at INRA, Toulouse, and the University of Glasgow. The Toulouse bioinformatics facility provides the TrypanoCyc server. TrypanoCyc is freely available and is not password protected.

TrypanoCyc database content can be mined in a pathway-oriented manner using the BioCyc-like web interface but also in a network perspective using the MetExplore web server, which allows tailored building and visualization of sub-networks. Two options are available to programmatically access TrypanoCyc: through pathway tools using JavaCyc or PerlCyc and through MetExplore using its web service.

TrypanoCyc is a unique knowledge source for people investigating *T. brucei* metabolism. The availability of SBML (44) files (provided as Supplementary Material S4) based on the curated network reconstruction in TrypanoCyc will underpin efforts to explore trypanosome metabolism using flux balance analysis (45) or other constraints-based techniques. It will also serve as a potential model organism for early eukaryotes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Brun,R., Blum,J., Chappuis,F. and Burri,C. (2010) Human African trypanosomiasis. *Lancet*, **375**, 148–59.
2. Barrett,M.P. (2010) Potential new drugs for human African trypanosomiasis: some progress at last. *Curr. Opin. Infect. Dis.*, **23**, 603–608.
3. Vincent,I.M., Creek,D.J., Burgess,K., Woods,D.J., Burchmore,R.J.S. and Barrett,M.P. (2012) Untargeted metabolomics reveals a lack of synergy between nifurtimox and eflornithine against Trypanosoma brucei. *PLoS Negl. Trop. Dis.* **6**, e1618.
4. Creek,D.J. and Barrett,M.P. (2014) Determination of antiprotozoal drug mechanisms by metabolomics approaches. *Parasitology*, **141**, 83–92.
5. Achcar,F., Kerkhoven,E.J. and Barrett,M.P. (2014) Trypanosoma brucei: meet the system. *Curr. Opin. Microbiol.*, **20C**, 162–169.
6. Cottret,L. and Jourdan,F. (2010) Graph methods for the investigation of metabolic networks in parasitology. *Parasitology*, **137**, 1393–1407.
7. Doyle,M.A., MacRae,J.I., De Souza,D.P., Saunders,E.C., McConville,M.J. and Likić,V.A. (2009) LeishCyc: a biochemical pathways database for Leishmania major. *BMC Syst. Biol.*, **3**, 57.
8. Shanmugasundram,A., Gonzalez-Galarza,F.F., Wastling,J.M., Vasieva,O. and Jones,A.R. (2013) Library of Apicomplexan Metabolic Pathways: a manually curated database for metabolic pathways of apicomplexan parasites. *Nucleic Acids Res.*, **41**, D706–D713.
9. Thiele,I. and Palsson,Ø.B. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.*, **5**, 93–121.
10. DeJongh,M., Formsma,K., Boillot,P., Gould,J., Rycenga,M. and Best,A. (2007) Toward the automated generation of genome-scale metabolic networks in the SEED. *BMC Bioinform.*, **8**, 139.
11. Whitaker,J.W., Letunic,I., McConkey,G.A. and Westhead,D.R. (2009) metaTIGER: a metabolic evolution resource. *Nucleic Acids Res.*, **37**, D531–D538.
12. Agren,R., Liu,L., Shoaie,S., Vongsangnak,W., Nookaew,I. and Nielsen,J. (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. *PLoS Comput. Biol.*, **9**, e1002980.
13. May,J.W., James,A.G. and Steinbeck,C. (2013) Metingear: a development environment for annotating genome-scale metabolic models. *Bioinformatics*, **29**, 2213–2215.
14. Caspi,R., Altman,T., Billington,R., Dreher,K., Foerster,H., Fulcher,C.A., Holland,T.A., Keseler,I.M., Kothari,A., Kubo,A. *et al.* (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.*, **42**, D459–D471.

15. Ginsburg,H. (2006) Progress in in silico functional genomics: the malaria Metabolic Pathways database. *Trends. Parasitol.*, **22**, 238–240.

16. Fairlamb,A.H., Blackburn,P., Ulrich,P., Chait,B.T. and Cerami,A. (1985) Trypanothione: a novel bis(glutathionyl)spermidine cofactor for glutathione reductase in trypanosomatids. *Science*, **227**, 1485–1487.

17. Tardif,M., Atteia,A., Specht,M., Cogne,G., Rolland,N., Brugière,S., Hippler,M., Ferro,M., Bruley,C., Peltier,G. *et al.* (2012) PredAlgo: a new subcellular localization prediction tool dedicated to green algae. *Mol. Biol. Evol.*, **29**, 3625–3639.

18. Lamour,N., Rivière,L., Coustou,V., Coombs,G.H., Barrett,M.P. and Bringaud,F. (2005) Proline metabolism in procyclic Trypanosoma brucei is down-regulated in the presence of glucose. *J. Biol. Chem.*, **280**, 11902–11910.

19. Michels,P.A., Hannaert,V. and Bringaud,F. (2000) Metabolic aspects of glycosomes in trypanosomatidae - new data and views. *Parasitol. Today*, **16**, 482–489.

20. Kanehisa,M., Goto,S., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–205.

21. Cottret,L., Wildridge,D., Vinson,F., Barrett,M.P., Charles,H., Sagot,M.-F. and Jourdan,F. (2010) MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res.*, **38**, W132–W137.

22. Berriman,M., Ghedin,E., Hertz-Fowler,C., Blandin,G., Renauld,H., Bartholomeu,D.C., Lennard,N.J., Caler,E., Hamlin,N.E., Haas,B. *et al.* (2005) The genome of the African trypanosome Trypanosoma brucei. *Science*, **309**, 416–422.

23. Karp,P.D., Latendresse,M. and Caspi,R. (2011) The pathway tools pathway prediction algorithm. *Stand. Genomic Sci.*, **5**, 424–429.

24. Thiele,I. and Palsson,B.Ø. (2010) Reconstruction annotation jamborees: a community approach to systems biology. *Mol. Syst. Biol.*, **6**, 361.

25. Opperdoes,F.R., Borst,P. and Spits,H. (1977) Particle-bound enzymes in the bloodstream form of Trypanosoma brucei. *Eur. J. Bioche.*, **76**, 21–28.

26. Opperdoes,F.R. and Borst,P. (1977) Localization of nine glycolytic enzymes in a microbody-like organelle in Trypanosoma brucei: the glycosome. *FEBS Lett.*, **80**, 360–364.

27. Berger,B.J., Dai,W.W., Wang,H., Stark,R.E. and Cerami,A. (1996) Aromatic amino acid transamination and methionine recycling in trypanosomatids. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 4126–4130.

28. Creek,D.J., Chokkathukalam,A., Jankevics,A., Burgess,K.E. V, Breitling,R. and Barrett,M.P. (2012) Stable isotope-assisted metabolomics for network-wide metabolic pathway elucidation. *Anal. Chem.*, **84**, 8442–8447.

29. Chavali,A.K., Whittemore,J.D., Eddy,J.a., Williams,K.T. and Papin,J.a. (2008) Systems analysis of metabolism in the pathogenic trypanosomatid Leishmania major. *Mol. Syst. Biol.*, **4**, 177.

30. Thiele,I., Swainston,N., Fleming,R.M.T., Hoppe,A., Sahoo,S., Aurich,M.K., Haraldsdottir,H., Mo,M.L., Rolfsson,O., Stobbe,M.D. *et al.* (2013) A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.*, **31**, 419–425.

31. Schomburg,I., Chang,A., Ebeling,C., Gremse,M., Heldt,C., Huhn,G. and Schomburg,D. (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.*, **32**, D431–D433.

32. Schomburg,I., Chang,A., Placzek,S., Söhngen,C., Rother,M., Lang,M., Munaretto,C., Ulas,S., Stelzer,M., Grote,A. *et al.* (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.*, **41**, D764–D772.

33. Artimo,P., Jonnalagedda,M., Arnold,K., Baratin,D., Csardi,G., De Castro,E., Duvaud,S., Flegel,V., Fortier,A., Gasteiger,E. *et al.* (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.*, **40**, W597–W603.

34. McDonald,A.G., Boyce,S. and Tipton,K.F. (2009) ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res.*, **37**, D593–D597.

35. The UniProt Consortium (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.*, **41**, D43–D47.

36. Aslett,M., Aurrecoechea,C., Berriman,M., Brestelli,J., Brunk,B.P., Carrington,M., Depledge,D.P., Fischer,S., Gajria,B., Gao,X. *et al.* (2010) TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.*, **38**, D457–D462.

37. Logan-Klumpler,F.J., De Silva,N., Boehme,U., Rogers,M.B., Velarde,G., McQuillan,J.A., Carver,T., Aslett,M., Olsen,C., Subramanian,S. *et al.* (2012) GeneDB-an annotation database for pathogens. *Nucleic Acids Res.*, **40**, D98–D108.

38. Keseler,I.M., Mackie,A., Peralta-Gil,M., Santos-Zavaleta,A., Gama-Castro,S., Bonavides-Martínez,C., Fulcher,C., Huerta,A.M., Kothari,A., Krummenacker,M. *et al.* (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.*, **41**, D605–D612.

39. Romero,P., Wagg,J., Green,M.L., Kaiser,D., Krummenacker,M. and Karp,P.D. (2005) Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.*, **6**, R2.

40. Zhang,P., Foerster,H., Tissier,C.P., Mueller,L., Paley,S., Karp,P.D. and Rhee,S.Y. (2005) MetaCyc and AraCyc. Metabolic pathway databases for plant research. *Plant Physiol.*, **138**, 27–37.

41. Urbaniak,M.D., Guther,M.L.S. and Ferguson,M.A.J. (2012) Comparative SILAC proteomic analysis of Trypanosoma brucei bloodstream and procyclic lifecycle stages. *PloS One*, **7**, e36619.

42. Paley,S.M. and Karp,P.D. (2006) The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Res.* **34**, 3771–3778.

43. Cline,M.S., Smoot,M., Cerami,E., Kuchinsky,A., Landys,N., Workman,C., Christmas,R., Avila-Campilo,I., Creech,M., Gross,B. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.

44. Hucka,M., Finney,A., Sauro,H.M., Bolouri,H., Doyle,J.C., Kitano,H., Arkin,A.P., Bornstein,B.J., Bray,D., Cornish-Bowden,A. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

45. Schellenberger,J., Que,R., Fleming,R.M.T., Thiele,I., Orth,J.D., Feist,A.M., Zielinski,D.C., Bordbar,A., Lewis,N.E., Rahmanian,S. *et al.* (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.*, **6**, 1290–1307.