

Visual workflows for ^{13}C -metabolic flux analysis

Katharina Nöh*, Peter Droste and Wolfgang Wiechert

Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: The precise quantification of intracellular metabolic flow rates is of fundamental importance in bio(techno)logy and medical research. The gold standard in the field is metabolic flux analysis (MFA) with ^{13}C -labeling experiments. ^{13}C -MFA workflows orchestrate several, mainly human-in-the-loop, software applications, integrating them with plenty of heterogeneous information. In practice, this had posed a major practical barrier for evaluating, interpreting and understanding isotopic data from carbon labeling experiments.

Results: Graphical modeling, interactive model exploration and visual data analysis are the key to overcome this limitation. We have developed a first-of-its-kind graphical tool suite providing scientists with an integrated software framework for all aspects of ^{13}C -MFA. Almost 30 modules (plug-ins) have been implemented for the Omix visualization software. Several advanced graphical workflows and ergonomic user interfaces support major domain-specific modeling and proofreading tasks. With that, the graphical suite is a productivity enhancing tool and an original educational training instrument supporting the adoption of ^{13}C -MFA applications in all life science fields.

Availability: The Omix Light Edition is freely available at <http://www.omix-visualization.com>

Contact: k.noeh@fz-juelich.de, p.droste@omix-visualization.com

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on April 30, 2014; revised and accepted on August 25, 2014

1 INTRODUCTION

1.1 ^{13}C -metabolic flux analysis

Fluxomics, which is also known as metabolic flux analysis (MFA; Winter and Krömer, 2013), deals with the detailed quantification of metabolic reaction rates inside living cells. Fluxes are material flows through biochemical pathways, rather than metabolite concentrations. Fluxomics is unique among the ‘omics’ methods because it does not—like the other methods—characterize the mere potential of an organism to perform a certain function, but rather it reflects its physiological *status quo* (Sauer, 2006; Wiechert, 2001).

A distinguishing feature of MFA is that metabolic fluxes cannot be directly quantified but must be computed from measured data based on a mathematical model. This particularly applies to MFA with stable isotopic carbon atoms (^{13}C -MFA), which is considered the gold standard in the field, calculating the *in vivo* fluxes from stable ^{13}C enrichments of intracellular metabolites described by a mathematical model (Sauer, 2006;

Stephanopoulos, 1999). The unique information quality of isotope-based fluxomics is increasingly recognized as enabling technology in metabolic engineering, systems biology and biomedical research.

The methodology of model-based ^{13}C -MFA is continuously developed since the early 1990s. The basics are described in tutorial papers (Tang *et al.*, 2009; Wiechert *et al.*, 2001; Zamboni *et al.*, 2009) and the state of the art is covered by several reviews (Kruger and Ratcliffe, 2009; Wiechert and Nöh, 2013; Zamboni, 2011). In addition to the experimental procedures of carbon labeling experiments (CLEs) and the protocols for biochemical analytics, the computational aspects play a central role in ^{13}C -MFA. The latter involve a sequence of modeling, simulation, experimental design, flux estimation and statistical evaluation steps. In all steps, the biological expert must have the ability to intervene and stir the process. Up to now, establishing tailored computational evaluation pipelines for ^{13}C -MFA is a time-consuming exercise. Easing these tasks is the subject of this article.

1.2 Modeling CLEs

Label distribution over a metabolic network is governed by well-established physical laws. This means that all frameworks for modeling and simulating CLEs involve the same basic principles and thus require the same input, i.e. mass balances, reaction stoichiometry, carbon mappings between reactants and measurements (Wiechert *et al.*, 2001). On the other hand, various computational formalisms have also emerged, which process the physical laws in different ways (Antoniewicz *et al.*, 2007; Schmidt *et al.*, 1997; van Winden *et al.*, 2002; Wiechert *et al.*, 1999). More importantly, these formalisms impact on the performance of the computational procedures but not on the calculated flux results (Weitzel *et al.*, 2007).

Unfortunately, structural and measured input data are both extensive and heterogeneous. A lot of information must be compiled on the metabolic network structure, carbon atom mappings (CAMs) for reactions, measurement configurations, flux constraints, input substrate compositions, etc. This implies that ^{13}C -MFA involves a painstaking configuration procedure requiring considerable experience. To this end, a general evaluation workflow for modeling and simulation of CLEs was proposed in (Wiechert *et al.*, 2001).

Most software systems for ^{13}C -MFA follow a universal approach (e.g. *influx_s*, *NMR2Flux+*, *Metran*, *OpenFLUX*, *13CFLUX2*, cf. Supplementary Table S1.4). In this context, ‘universal’ means that a modeling environment is available for the specification of arbitrary metabolic networks, input substrate compositions, measurement configurations and

*To whom correspondence should be addressed.

additional constraints. Available software tools for ^{13}C -MFA rely heavily on text-, table-, XML- or matrix-based input documents to specify models. From these files, all equations needed to describe the CLE are automatically generated. It is neither easy nor straightforward to build up the complex configuration for ^{13}C -MFA, to check its correctness and consistency, to trigger simulation and data evaluation runs or to automatically visualize the final flux maps in the metabolic network context.

Notably, in all parts of a ^{13}C -MFA model, 100% correctness is needed. If, for example, one single CAM has been incorrectly specified, a measurement has been incorrectly related to a metabolite's fragment or the actual input substrate has not been accurately mirrored, the CLE analysis will fail to calculate the correct flux map. At present, it is difficult to detect and, in turn, locate the error in a model file or table covering several printed pages. Although more than 400 possible syntactic and semantic errors in the model specification can be detected, even the recently published 13CFLUX2 compiler, representing a decade of experience in ^{13}C -MFA, cannot prevent systematic mistakes or fundamental misunderstandings in the model formulation (Weitzel *et al.*, 2013).

1.3 Visual workflows

Experience has shown that building up the first model description of a complex CLE usually cannot be managed by a newcomer without experienced help. This simple fact is certainly one of the major reasons why ^{13}C -MFA is still not as commonly used as other omics techniques. This contribution goes one step further by addressing this very problem. An innovative collection of visual workflows is presented, which are intuitive to use and cover the ^{13}C -MFA model setup, all computational steps of simulation, experimental design, parameter fitting and statistical analysis, as well as the visual presentation of results.

The comprehensive graphical tool suite that we have developed builds on the functionality of the network-based visualization software Omix® (Droste *et al.*, 2013). The base functionalities of Omix allow even an inexperienced user to quickly create and edit metabolic networks (Droste *et al.*, 2011a, 2011b). Because all metabolic flux maps generated during the analysis are later visualized in the same network diagram, Omix users can design the network to suit their personal preferences and context-dependent priorities. This is essential for intuitive and effective visual recognition. The Omix Visualization Language (OVL) makes network drawings scriptable, which enables the easy and flexible user-defined mapping of data to the visual attributes of each single network node and edge (Droste *et al.*, 2011b).

To draw the maximum advantage from a graphical configuration of CLEs, it is important that graphical tools are not mere user interfaces to the underlying data structures but that they represent powerful, easy-to-learn, interactive and intuitive tools for immediate consistency, syntactical/semantical correctness and completeness checking, modeling and visual analysis. We achieved this goal by developing several advanced visual workflows supporting the model building processes (cf. Supplementary Table S1.3 for a complete list).

In the subsequent sections, we follow the general ^{13}C -MFA workflow outlined in (Wiechert *et al.*, 2001) and illustrate how the visual model configuration is implemented for each of the steps. Orchestration of the single workflow steps is made easier by a graphical workflow manager, facilitating plug-and-play assembly and coordinated execution of each step. The established comprehensive visual ^{13}C -MFA workflow, as illustrated in Supplementary Figure S1.1, consists of a unique collection of almost 30 Omix plug-ins (cf. Supplementary Table S1.2). Several figures illustrating the core elements of the graphical workflow along with a detailed explanation of non-trivial algorithms and additional applications are provided as Supplementary Material. Having briefly outlined the general aspects of metabolic network modeling, we will focus, in the following, on the ^{13}C -MFA-specific aspects of the visual workflows.

2 METABOLIC NETWORK ASSEMBLY

2.1 Metabolic network specification and annotation

The first step in any ^{13}C -MFA workflow is the collection of reactions, i.e. the stoichiometric relation between enzymes and metabolites. Clearly, building on expert-curated information sources minimizes the error in model configuration. There is a wide range of publicly available information sources, and a wealth of Omix plug-ins have been developed for importing metabolic networks into the workflow from databases (KEGG, BioCyc), files (SBML, BioPAX, SBGN-ML, GraphML) and reaction lists. In addition to the network structure, subsequent ^{13}C -MFA modeling steps rely heavily on accessory information such as the chemical structure formulas of metabolites, or common names and database IDs of reactions. Our graphical workflow helps to cut the costs of assembly and data integration by compiling tailored information sets automatically. In turn, these sets provide a valuable knowledge base for the subsequent modeling workflow (cf. Supplementary Fig. S2.1).

2.2 Flux constraints

From the network topology, one stoichiometric constraint for each internal metabolite node is derived by balancing the rates of incoming and outgoing reactions. In ^{13}C -MFA, reversible reaction steps have to be taken into account because they influence the labeling distribution, meaning that net fluxes should be complemented by their respective exchange fluxes (Wiechert and de Graaf, 1997). Both net fluxes and exchange fluxes can be further constrained by imposing linear equality and inequality constraints derived from expert knowledge, e.g. concentration-induced flux directions (cf. Supplementary Fig. S2.2). These additional modeling assumptions can be introduced in a dedicated window with a register of defined fluxes. Here, the user specifies (in)equality constraints in terms of linear sums of fluxes. When the user is defining these assumptions, a pop-up window opens showing a register of all fluxes contained in the network. The (in)equalities are parsed in real time. Immediate reporting of typing errors prevents users from entering reaction names that do not exist. Expressive error messages point out inconsistencies with the network and typing errors (cf. Supplementary Fig. S2.3).

2.3 CAMs: status quo

CAMs must be specified to simulate CLEs. To this end, the fate of every single carbon atom in every reaction step of the network must be unambiguously defined (cf. Fig. 1B and Supplementary Figs S2.4–6). This information must be absolutely correct and consistent. Unfortunately, the correctness of the transitions is difficult to verify manually in both the textual representation of CAMs (Wiechert *et al.*, 1999) and in the mapping matrix representation (Schmidt *et al.*, 1997). Thus, it is almost impossible to proofread large models using the current state of the art.

Some databases like ARM (Arita *et al.*, 2006), carbon-fate maps (Mu *et al.*, 2007), the MetaCyc database family and publications (e.g. Ravikirithi *et al.*, 2011) supply computationally generated and partially manually curated information on CAMs. Nevertheless, some general fundamental problems have yet to be solved.

Firstly, there is still a lack of reliable CAM information with high coverage. Although progress was recently made in heuristics-based CAM generation (Latendresse *et al.*, 2012), unknown reaction mechanisms inducing unusual mappings pose an inherent problem for all automated approaches. Specifically, enzymatically indistinguishable ‘C-symmetric’ atoms require the formulation of alternative CAMs in ^{13}C -MFA models, which is still an unsolved problem (Zhou and Nakhleh, 2012).

Secondly, there is a lack of ^{13}C -MFA standards for CAMs and common tools for their data management, rendering model reuse, sharing and merging almost impossible. The underlying reason for this is that CAMs can only be correctly interpreted when the numbering of the carbon atoms is unambiguously stipulated and when this setting is clearly documented. Although IUPAC/IUBMB commissions provide detailed recommendations, these guidelines are invariably not obeyed in practice. In addition, the widely applied MDL®, MOL or CML files are not standardized with regard to atom numbers and, hence, cannot be used as standard (cf. Supplementary Fig. S2.7). As a consequence, up to now, ^{13}C -MFA models can only be exchanged on a personal basis.

2.4 Canonical CAMs and supportive automatisms

The most reliable way to address this limitation while unlocking available CAM information for fluxomics is (i) to trust in the human expert’s modeling and proofreading capabilities, and (ii) to keep the atom numbering issue away from the user. This requires graphical front-end tools that gather and display all of the necessary information in an intuitive visualization, allow for quick correction and the straightforward specification of new CAMs. If a unique numbering of the carbon atoms is additionally guaranteed in the back-end, this effort pays off immediately because once a correct model with curated CAMs of the central metabolism has been developed, further models can be quickly derived. Notably, model reuse and assembly benefits enormously from cut, copy and paste functionalities.

This principle has been applied for the first time by implementing CAMs that are independent of human-imposed numbering schemes. To this end, we used the IUPAC International Chemical Identifier (InChI™ ver. 1.04, www.iupac.org/home/publications/e-resources/inchi.html). The InChI Trust provides open software to transform a given structural formula into an

InChI string. A unique atom sequence is obtained in the process. Referring CAMs to this reference numbering ensures that two humans drawing the same molecule or loading its MOL or CML files will always get the same carbon atom numbering. InChI’s AuxInfo layer contains the information required to revert the conversion procedure (cf. Supplementary Fig. S2.7).

By using molecular structure files, the human user directly obtains a graphical representation of each molecule along with a standardized atom numbering. The open-source Java library CDK (Chemistry Development Kit ver. 1.4, Steinbeck *et al.*, 2003) is used for InChI string generation and the handling of diverse data formats. The molecules are imported and visualized by the Structural Formula plug-in. The limitation of InChI to correctly handle generic structures, such as Markush structures with generic R groups, is algorithmically solved within the plug-in.

In Supplementary Figure S2.8, we demonstrate how the KEGG RPAIR database containing atom correspondences between substrate and product metabolite pairs can be used to generate canonical CAMs. Owing to its universality, our novel approach is readily applicable to other CAM resources such as the BioCyc database. If the atom correspondences are not known, the user interactively specifies or edits the atom transitions by choosing the source carbon atom marker (colored circle) and dragging it toward its companion atom, which then changes its color from gray to the color of the source atom (cf. Supplementary Figs. S2.4–6). The actual carbon atom numbering of the reactants plays no role in this procedure and is only optionally displayed for the sake of integrity. It is not always desirable to map all atoms of a molecule. For instance, coenzyme A (CoA) synthesis is usually not regarded in ^{13}C -MFA network models. Thus, pruned CAMs include only the core carbon skeleton of compounds like acetyl CoA or succinyl CoA (cf. Supplementary Material Fig. S2.5). Consequently, only the balances of the traced carbon atoms are checked by the Atomic Layer Modeling plug-in.

Importantly, the user can also specify alternative mappings, so-called ‘variants’, for reactions with C-symmetric reactants (cf. Supplementary Fig. S2.6). This obviates the need for the modeler to explicitly formulate symmetric reactions.

All in all, the graphical specification and more vitally the proofreading have been turned into convenient procedures. Biochemical experts are familiar with the resulting representation and can now, for the first time, check as many as 100 reactions in a short time. We evaluated the visual CAM specification with a large carbon atom transition network for yeast metabolism containing 845 metabolites and 1100 reactions (826 thereof carrying manually specified CAMs). Visual proofreading of the network structure and CAMs took less than one working day. Other studies have showed similar figures.

3 SETTING THE NETWORK’S INS AND OUTS

The next task is to specify labeled substrates as an additional information source to effectively narrow down the number of possible flux solutions. Administering cells with ^{13}C -labeled substrates imprints specific labeling patterns on the intracellular metabolites. The goal is to end up with only one possible

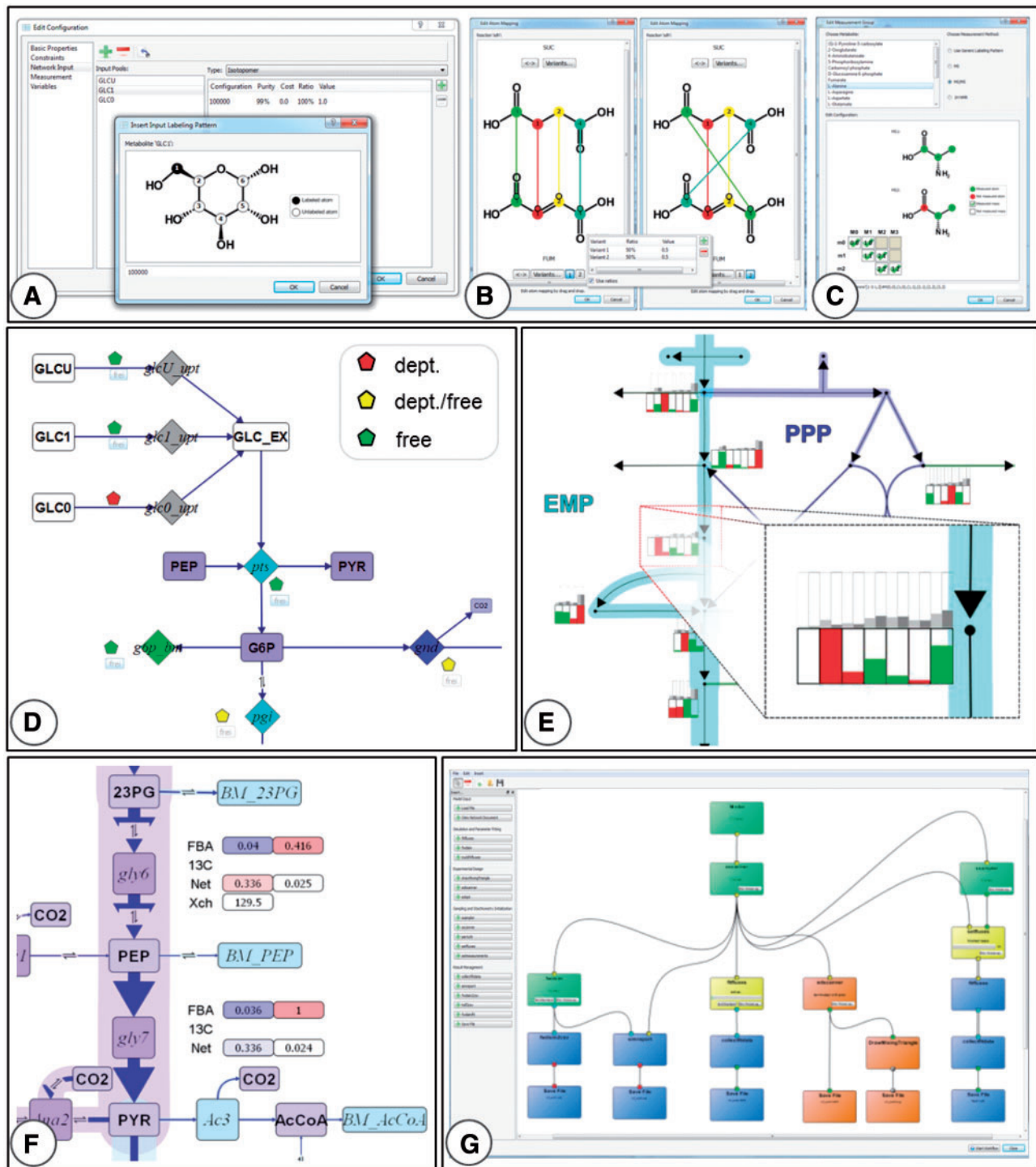


Fig. 1. Visual elements of graphical ^{13}C -MFA workflows. (A) User interfaces for the graphical composition of input labeling species. (B) Graphical modeling of CAMs. (C) Graphical specification of MS/MS labeling measurements. (D) Network annotation with qualitative and quantitative data: free fluxes. (E) Differential view of simulated and measured mass lanes from seven datasets (red—negative, green—positive) with measurement standard deviations (gray bars) and (F) comparison of flux maps derived from FBA and ^{13}C -MFA (blue—low, white—medium, red—high fluxes). The color coding provides a visual aid for identifying network-wide comparable flux values. (G) With the Omix workflow manager ^{13}C -MFA workflows can be easily constructed, configured, saved, run and shared. Enlarged versions of the figures are found in the supplement

remaining flux distribution that is consistent with these labeling patterns.

3.1 Isotopic substrate specification

It is common practice in ^{13}C -MFA to administer mixtures of differently labeled (^{13}C) and unlabeled (^{12}C) substrates to obtain maximum information from the CLE. Importantly, isotopically labeled substrate species always contain a small proportion of impure substances (commercially available substrates have typically a purity of <99 atom%). This means that each compound in a mixture is itself a mixture of differently labeled isotopomers. Owing to their impact on the flux calculation, it is essential to precisely define the input labeling impurities according to the manufacturer's specification.

As shown in (Wiechert *et al.*, 2001), mixture compositions can be conveniently modeled by including multiple confluent substrate uptake steps into the network. Each influx carries the fraction of one of the (impure) labeled species provided in the substrate mixture. All of these input fluxes are then blended in one single mixture pool, and the sum of all influxes constitutes the total carbon uptake rate. The influx ratio determines the proportion of the mixture species. In this way, input substrate compositions can be intuitively modeled as an integral part of the metabolic network structure.

In the graphical ^{13}C -MFA workflow, a list is generated of all input compounds found in the model. For each compound, the modeler specifies the associated input isotopomer visually by individually selecting the carbon atoms and interactively switching between labeled and unlabeled states (cf. Fig. 1A and Supplementary Fig. S2.9). Finally, the isotopic purity of each labeled species is entered into the user interface.

3.2 Measurement (error) configuration

One of the most difficult tasks in ^{13}C -MFA is identifying the precise relation between various measured signals and the corresponding labeling fractions of intracellular metabolite pools. The reason for this is the enormous heterogeneity of mass spectrometric (MS) and nuclear magnetic resonance (NMR) analytical platforms and protocols published in the literature as well as the continually advancing techniques. Recent developments in this field are described, for example, by Fan *et al.* (2012) and Nagana Gowda and Raftery (2014). Several other papers are concerned with the determination of labeling patterns for special equipment (cf. Supplementary Fig. S2.10–12 for more references).

A universal modeling framework must be flexible enough to model all present and expected future measurement configurations. A matrix representation describing a group of simultaneously measured signals as linear combinations of the analyte's isotopomer fractions provides a generally applicable approach (Choi and Antoniewicz, 2011 and Wiechert and de Graaf, 1997).

Although measurement matrices are universal, they are inconvenient for manual specification. For example, a metabolite with nine carbon atoms, such as the amino acid L-phenylalanine, has $2^9 = 512$ isotopomers, which induces a measurement matrix with as many columns. The generation of the corresponding measurement matrices should be hidden from the user and automated in such a way that the corresponding matrices are automatically

generated from a structure- or fragment-related input while remaining consistent with the carbon atom network (cf. Fig. 1C).

The general approach taken in our visual ^{13}C -MFA workflow is to provide an interactive canvas where the modeler conveniently specifies the measurements. After one of the intracellular metabolites has been chosen from a drop-down menu, the measurement technique is selected (cf. Supplementary Fig. S2.11). Here, the modeler chooses between four different measurement types: ' ^1H -NMR' (positional ^{13}C enrichments), 'MS' (mass isotopomers), 'MS/MS' (positional mass isotopomers) and 'Generic' for all cases where the aforementioned measurement techniques do not provide sufficient flexibility. All measurement types are exemplified for L-phenylalanine in Supplementary Fig. S2.12. In addition to labeling patterns, the modeler likewise specifies extracellular fluxes, i.e. specific rates associated with substrate consumption, biomass and product synthesis. Each single measurement has to be accompanied by its estimated standard deviation, which depends on experimental, analytical and device-specific factors.

4 NETWORK ANALYSIS

Graphical tools and algorithms help us to understand important characteristics of the formulated metabolic labeling system and thus play an important role in the plausibility assessment, testing and proofreading of network models. Conventionally, any network analysis starts with inspecting the stoichiometric matrix containing the full set of stoichiometric coefficients of all reactants in the network. The set of all minimal pathways through the network is characterized by the elementary flux mode (Terzer and Stelling, 2008), and phenotypes can be predicted with flux balance analysis (FBA, Orth *et al.*, 2010). We have developed Omix plug-ins for all of these analyses (cf. Supplementary Table S1.2). Additionally, the Klamt group has provided an Omix plug-in for the CellNetAnalyzer toolbox, facilitating various further network analysis functions (Klamt *et al.*, 2007). In the following, we concentrate on the advanced analysis of isotope labeling systems.

4.1 Picking free net fluxes

The set of stoichiometric equations in union with the equality constraints reduces the dimension of the network model's flux space. The rank of the corresponding extended stoichiometric matrix determines the dimensionality of the feasible flux space. It proved convenient to represent the degrees of freedom in the stoichiometric space using so-called 'free fluxes' (Wiechert and de Graaf, 1997). A set of free fluxes is a distinguished non-unique subset of all network fluxes. When it has been assigned to certain values, it can uniquely determine all other fluxes and thus the complete flux map. Usually, free fluxes are 'prominent' fluxes like the entry fluxes of the glycolysis or the citric acid cycle. Manual selection eases some of the following tasks because guesses for the initial fluxes are intuitively interpretable.

Unfortunately, not every combination of the right number of fluxes gives a proper set of free fluxes (Wiechert and de Graaf, 1997). For large networks, it is almost impossible to guess a valid set of independent fluxes by intuition. Various matrix factorization algorithms (e.g. singular value decomposition or Gaussian

elimination) are capable of determining free flux sets automatically. However, the trade-off is that the proposed flux constellations are seldom human-interpretable.

To avoid these problems and to allow a convenient and rapid user-driven selection of free fluxes, we invented an interactive recursive graphical algorithm. From the user viewpoint, the Omix Free Flux plug-in displays a clickable ‘traffic light’ symbol in the vicinity of each flux node in the network visualization (cf. Fig. 1D and Supplementary Fig. S2.13). This symbol is green (the flux is selected as a free flux), yellow (the flux can still be chosen as free flux) or red (the flux can no longer be selected because of its stoichiometry-induced dependency on the free fluxes already selected). After each selection, the dependencies are automatically recalculated by recursive Gaussian elimination and the traffic light colors are updated. If a free flux selection turns out to be disadvantageous, it can be simply redefined with another mouse click on the traffic light symbol. Our experience with large-scale networks has shown that this visual explorative procedure always yields a convenient result for the selection of free fluxes within a short time frame of only a few minutes.

4.2 Tracing the paths of labeled atoms

Recently, a universal path-tracing theory was developed for CAM networks that quantitatively explains the accumulation of labeled molecular species in each metabolite for a given input labeling composition (Wiechert *et al.*, 2013). It was found that metabolic cycles induce an infinite number of reaction sequences contributing to the label enrichment of a single metabolite pool. Path tracing of (un)labeled atoms and molecule fragments with a finite number of steps already enables a qualitative simulation of the CLE. This means that by forward tracking the fate of labeled input species and the backward tracing of measured species to their origins, CLEs can be played through.

For this reason, tracing the possible paths of isotopically labeled atoms is an invaluable visual aid for proofreading and identifying basic effects such as the cycling of carbon atoms (Chance *et al.*, 1983) or the unexpected emergence/dilution of labeling suggesting alternative routes (Creek *et al.*, 2012). For inexperienced users, the tracing of labeled molecules is an essential tool to understand the underlying complexity of CLEs and provides more confidence for evaluating the significance of observed effects. As a result, label path tracing has been frequently applied in the literature using manually drawn two-dimensional pictures (e.g. Klapa *et al.*, 1999).

The CumoVis plug-in provides a tailored tool for label tracing that serves the needs of both, the practitioner and the theorist. CumoVis displays not only single paths of isotopic tracers but also the complete isotope labeling network (cf. Supplementary Fig. S2.14), which complements the reaction-wise proofreading described in Section 2.4. To facilitate a better understanding, three-dimensional visualization techniques are applied (Droste *et al.*, 2008). As an extension of its first version, the second-generation CumoVis plug-in fully exploits the benefits of this visualization in a wealth of 3D views (cf. Supplementary Movies). The views show freely rotatable and zoomable large-scale metabolic and isotope labeling networks from every angle, allowing the user to browse the network structure and to filter

out subgraphs for closer inspection. Important new features are that tracers can be followed both forward and backward through the network starting from any specified atom or molecule fragment, represented by its chemical structure, up to a specified path length. All label paths can be conveniently investigated one by one in a fully animated fashion.

5 COMPUTATIONAL TASKS

The subsequent computational tasks constitute the backbone of any ^{13}C -MFA workflow and thus provide typical functionalities of fluxomics software tools. In this article, we feature the 13CFLUX2 simulation framework (Weitzel *et al.*, 2013). In principle, it is also possible to interface other ^{13}C -MFA software systems from the Omix environment (cf. Supplementary Table S1.4). We have developed several tailored visual interfaces to pass the input data to the respective computational tools, check the validity of the 13CFLUX2 model file, launch calculations via Web services on distributed computers and, finally, aggregate and visually present the results to the modeler.

5.1 CLE simulation

Based on the CAM network model with specified input substrate labeling, constraints and, optionally, measurements, the forward simulation of a CLE represents the central computational module of any ^{13}C -MFA workflow. ‘Forward simulation’ is the assumption of some feasible flux distribution—i.e. values are assigned to the free fluxes fulfilling all specified constraints—followed by the computing of the steady-state isotope labeling distribution of the metabolic intermediates. The simulation tool is of great value for exploring the influence of different fluxes and input labels on the outcome of the CLE. In this way, for example, sensitive measurement signals can be identified. The calculated simulation results are deposited in 13CFLUX2-tailored XML-based output formats and stored for subsequent analysis or directly used to visualize informative flux maps (cf. Fig. 1E).

5.2 Experimental design

Owing to a limited number of available measurements and the ubiquitous measurement errors in extracellular fluxes and labeling data, it can happen that not all fluxes in a metabolic network are reasonably well determined. Fortunately, a good prediction of the expected quality of the flux map can be calculated before any experiment is carried out (Möllney *et al.*, 1999). Thus, the optimal experimental design (OED) methodology has become a routine tool for designing CLEs with a wide range of applications (Crown *et al.*, 2012; Millard *et al.*, 2014; Nargund and Sriram, 2013; Weitzel *et al.*, 2013). OED algorithms compute a maximally informative input substrate mixture targeted to the specific research aim. All OED variants are based on statistical quality measures evaluating how well the fluxes are determined by the data. By assigning user-specified weights to the fluxes, the design can be focused on specific regions of the network. The mathematical details of various design measures and ‘alphabetical’ optimality criteria can be found in the classical literature (Pázman, 1986).

In ^{13}C -MFA studies, seldom more than three differently labeled substrate species—typically ^{12}C -, $1\text{-}^{13}\text{C}$ - and $\text{U-}^{13}\text{C}$ -labeled glucose—are fed to the cells. In the triplicate case, the space of possible substrate mixtures can be conveniently presented by a mixture triangle (Möllney *et al.*, 1999), which is a statistical quality measure for each input composition (cf. Supplementary Fig. S2.15). From this mixture triangle, optimal or suboptimal input composites can be directly read off. It is just as important that the robustness of an experimental design with respect to variations in the input mixture can be easily judged visually.

Nowadays, a variety of specifically labeled substrate species can be synthesized. Thus, for explorative purposes, higher-dimensional OEDs with more than three labeled substrate species may be investigated with the 13CFLUX2 program. Instead of directly visualizing the results of such OEDs, the modeler can download them as HDF5 files for post-processing purposes.

5.3 Flux estimation and statistical assessment

After the well-designed CLE has been carried out, the central step in the ^{13}C -MFA workflow is the estimation of the *in vivo* fluxes from the data. Measurements are typically made available as a table of extracellular measured fluxes and MS/(MS)- or NMR-based labeling enrichments. Every measured value must be accompanied by a standard deviation characterizing its uncertainty. All measurement/standard deviation pairs can be either arranged in groups or as independent single measurements. Measurement values and standard deviations can be conveniently copied from prepared spreadsheets and pasted in the measurement specification dialogue as described in Section 3.2 (cf. Supplementary Fig. S2.11).

The unknown *in vivo* fluxes are calculated from a parameter fitting procedure where the difference between real and simulated data is minimized, e.g. by means of a non-linear weighted least-squares functional (Draper and Smith, 1998). In contrast to the simulation step, the process of fitting CLE data depends on various factors and its running time therefore cannot be predicted a priori (Dalman *et al.*, 2013). Hence, the modeler can monitor the progress of the iterative flux estimation procedure online (cf. Supplementary Fig. S2.17).

After finalization of the fitting procedure, the best flux fits are collected and transferred back to the local Omix modeling environment to visualize the resulting flux maps. The final step involves a companion uncertainty analysis (Dalman *et al.*, 2013; Wiechert *et al.*, 1997), which assesses the reliability of the calculated flux estimates. Confidence intervals provide a range for each flux, which contains the true flux value up to a certain degree of confidence. By using OVL, reaction/metabolite nodes of the flux maps are easily decorated with annotation boxes that display all relevant values numerically, size- or color-coded (Fig. 1F).

5.4 Putting pieces together: workflow composition

We have described various computational modules requiring different input configurations, data specifications and dependencies that must be fixed and maintained by the scientist. Controlling the tools manually is a must for learning and training purposes and is adequate for one single fluxome study. For routine analysis of large-scale datasets, however, this practice is not

effective. Particularly in eScience, scientific workflows have emerged as a useful new paradigm to describe and automate routine computational analysis tasks (Ebert *et al.*, 2012 and Gil *et al.*, 2007). By stringing existing 13CFLUX2 modules together, tailored ^{13}C -MFA evaluation pipelines can be assembled, ranging from initial model building to final flux-map visualization. In turn, these workflows act as computational protocols for subsequent reuse, verification and adaption purposes.

The Omix 13CFLUX2 Launcher plug-in provides a lightweight scientific workflow manager with a graphical editor to ease the tasks of designing, manipulating, executing and monitoring universal fluxomics analysis (cf. Fig. 1G). As a starting point, all network, atom mapping, constraint, measurement and input substrate-mixture data are assumed to be available either as Omix or 13CFLUX2 model files. Graphical primitives representing the workflow's atomic components on a high-level are provided, representing fully configurable front-ends to the aforementioned tools. These components can be manually selected and linked by the modeler to set up comprehensive ^{13}C -MFA evaluation pipelines.

Supplementary Figure S2.16 shows the Omix workflow manager canvas with an explorative ^{13}C -MFA workflow assembled by drag-and-drop from the tool list. The workflow composition is guided by the colors of the modules' I/O nodes, which indicate their interoperability. All workflow tasks are implemented as remote Web services where independent workflow branches run in parallel. To enable an easy online survey of the computation progress, advancing steps are colored yellow, completed tasks are displayed in green and, should unexpected failures emerge, the box turns red and the reasons for interruptions can be inspected in detail. Once all computational tasks have been completed, the workflow manager automatically downloads and saves the simulation results for further processing.

6 DISCUSSION

The presented comprehensive plug-in suite for the Omix visualization software complements the recently published 13CFLUX2 software and graphically supports the complete well-established ^{13}C -MFA workflow. The new 'look' and 'feel' enables experimentalists who are not familiar with programming or scripting to use the ^{13}C -MFA computational machinery immediately. The classical command-line orientation of ^{13}C -MFA tools has thus become history, although certain special requirements still might require a scripted solution. At the time of writing, 27 Omix plug-ins can be orchestrated to tailored evaluation workflows (cf. Supplementary Fig. S1.1 and Supplementary Table S1.2).

These plug-ins have been extensively tested and used in several ^{13}C -MFA studies with toy examples as well as with realistically complex real-life examples. As a typical use case, the developed suite is used to analyze a variety of *Corynebacterium glutamicum* strains for a broad range of dilution rates and bioreactor working volumes (unpublished data). Once a reference network model was set up, >20 datasets were quickly processed and visually evaluated with no overhead. Resulting flux maps were visualized with a tailored OVL script (e.g. Fig. 1E), making it easy for the experimentalist to comparatively assess and characterize the CLEs according to their biological, analytical and experimental

reproducibility. With that an in-depth understanding of experimental and biological robustness of *C.glutamicum* is conveyed. In similar studies, the graphical workflows was found to strongly reduce the development time for flux models and ease serial, yet interactive, fluxome analyses. Equally important, the graphical look-and-feel makes it much easier for ^{13}C -MFA newcomers to get started. Therewith, our graphical suite delivers productivity-enhancing tools that accelerate ^{13}C -MFA learning, execution and dissimulation processes.

Apart from practical issues, the implementation of a complete graphical tool suite supporting ^{13}C -MFA is not just a simple WYSIWYG task. A collection of innovative ideas and algorithms were developed to exploit the potential of interactive graphical visualization. Non-trivial, domain-specific, yet universal, algorithmic solutions are the derivation of CAMs from information contained in databases, full copy/paste functionality and compatibility of CAMs, graphical proofreading and editing of CAMs, interactive determination of free flux sets, three-dimensional representation and exploration of carbon labeling networks, interactive path tracing, as well as the automatic generation and drawing of mixture diagrams (cf. Supplementary Table S1.3 for a complete list).

Although bare simulation times at present are in the range of milliseconds for single simulation runs, extensive analyses should preferably be distributed to dedicated computing resources. Purchase-on-demand compute models are an option for large-scale ^{13}C -MFA workflows (Dalman *et al.*, 2013). Future work will be dedicated to alleviating (post-)processing tasks on distributed HPC resources to meet the increasing need for real-time accessibility and the control of fluxomics workflows, which, undoubtedly, presents new challenges for our visual workflow.

ACKNOWLEDGEMENTS

The development of Omix plug-ins for fluxomics is driven by interactions and collaborations with the ^{13}C -MFA community and 13CFLUX2 user groups as the ultimate end users of the suite. On the other hand, work in related projects also inspired and influenced the presented developments. The authors would like to thank all direct and indirect contributors, in particular, Andreas Dietrich (FreeFluxes, FBA plug-ins), Tobias Vehrkamp (KEGG, Molecular Structure plug-ins), Michael Weitzel, Tolga Dalman and Robert Siegburg (13CFLUX2 and Web service assistance), Sebastian Niedenführ, Stephan Noack, as well as Elisabeth and Stephan Zelle for providing application studies and rigorous and time-intensive testing.

Funding: This work is partially supported by the European Commission project SysInBio (212766).

Conflict of interest: none declared.

REFERENCES

- Antoniewicz, M.R. *et al.* (2007) Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.*, **9**, 68–86.
- Arita, M. *et al.* (2006) Map editor for the atomic reconstruction of metabolism (ARM). In: Saito, K. *et al.* (eds) *Plant Metabolomics*. Springer-Verlag, Berlin Heidelberg, pp. 129–139.
- Chance, E.M. *et al.* (1983) Mathematical analysis of isotope labeling in the citric acid cycle with applications to ^{13}C NMR studies in perfused rat hearts. *J. Biol. Chem.*, **258**, 13785–13794.
- Choi, J. and Antoniewicz, M.R. (2011) Tandem mass spectrometry: a novel approach for metabolic flux analysis. *Metab. Eng.*, **13**, 225–233.
- Creek, D.J. *et al.* (2012) Stable isotope-assisted metabolomics for network-wide metabolic pathway elucidation. *Anal. Chem.*, **84**, 8442–8447.
- Crown, S.B. *et al.* (2012) Rational design of ^{13}C -labeling experiments for metabolic flux analysis in mammalian cells. *BMC Syst. Biol.*, **6**, 43.
- Dalman, T. *et al.* (2013) Cloud MapReduce for Monte Carlo bootstrap applied to metabolic flux analysis. *Futur. Gener. Comput. Syst.*, **29**, 582–590.
- Draper, N.R. and Smith, H. (1998) *Applied Regression Analysis*, 3rd edn. Wiley, New York, USA.
- Droste, P. *et al.* (2013) Omix - a visualization tool for metabolic networks with highest usability and customizability in focus. *Chemie Ing. Tech.*, **85**, 849–862.
- Droste, P. *et al.* (2011a) Semi-automatic drawing of metabolic networks. *Inf. Vis.*, **11**, 171–187.
- Droste, P. *et al.* (2008) Visual exploration of isotope labeling networks in 3D. *Bioprocess Biosyst. Eng.*, **31**, 227–239.
- Droste, P. *et al.* (2011b) Visualizing multi-omics data in metabolic networks with the software Omix: a case study. *Biosystems*, **105**, 154–161.
- Ebert, B.E. *et al.* (2012) Flux-P: automating metabolic flux analysis. *Metabolites*, **2**, 872–890.
- Fan, T.W.-M. *et al.* (2012) Stable isotope-resolved metabolomics and applications for drug development. *Pharmacol. Ther.*, **133**, 366–391.
- Gil, Y. *et al.* (2007) Examining the challenges of scientific workflows. *IEEE Comput.*, **40**, 24–32.
- Klamt, S. *et al.* (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol.*, **1**, 2.
- Klapa, M.I. *et al.* (1999) Metabolite and isotopomer balancing in the analysis of metabolic cycles: I. Theory. *Biotechnol. Bioeng.*, **62**, 375–391.
- Kruger, N.J. and Ratcliffe, R.G. (2009) Insights into plant metabolic networks from steady-state metabolic flux analysis. *Biochimie*, **91**, 697–702.
- Latendresse, M. *et al.* (2012) Accurate atom-mapping computation for biochemical reactions. *J. Chem. Inf. Model.*, **52**, 2970–2982.
- Millard, P. *et al.* (2014) IsoDesign: a software for optimizing the design of ^{13}C -metabolic flux analysis experiments. *Biotechnol. Bioeng.*, **111**, 202–208.
- Möllney, M. *et al.* (1999) Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments. *Biotechnol. Bioeng.*, **66**, 86–103.
- Mu, F. *et al.* (2007) Carbon-fate maps for metabolic reactions. *Bioinformatics*, **23**, 3193–3199.
- Nagana Gowda, G.A. and Raftery, D. (2014) Advances in NMR based metabolomics. In: Simó *et al.* (eds) *Fundamentals of Advanced Omics Technologies: From Genes to Metabolites, Comprehensive Analytical Chemistry*. Elsevier, Amsterdam, The Netherlands, pp. 187–211.
- Nargund, S. and Sriram, G. (2013) Designer labels for plant metabolism: statistical design of isotope labeling experiments for improved quantification of flux in complex plant metabolic networks. *Mol. Biosyst.*, **9**, 99–112.
- Orth, J.D. *et al.* (2010) What is flux balance analysis? *Nat. Biotechnol.*, **28**, 245–248.
- Pázman, A. (1986) *Foundations of Optimum Experimental Design*. Reidel, Dordrecht, The Netherlands.
- Ravikirithi, P. *et al.* (2011) Construction of an *E. coli* genome-scale atom mapping model for MFA calculations. *Biotechnol. Bioeng.*, **108**, 1372–1382.
- Sauer, U. (2006) Metabolic networks in motion: ^{13}C -based flux analysis. *Mol. Syst. Biol.*, **2**, 62.
- Schmidt, K. *et al.* (1997) Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol. Bioeng.*, **55**, 831–840.
- Steinbeck, C. *et al.* (2003) The Chemistry Development Kit (CDK): an open-source Java library for Chemo- and Bioinformatics. *J. Chem. Inf. Comput. Sci.*, **43**, 493–500.
- Stephanopoulos, G. (1999) Metabolic fluxes and metabolic engineering. *Metab. Eng.*, **1**, 1–11.
- Tang, Y.J. *et al.* (2009) Advances in analysis of microbial metabolic fluxes via ^{13}C isotopic labeling. *Mass Spectrom. Rev.*, **28**, 362–375.
- Terzer, M. and Stelling, J. (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, **24**, 2229–235.
- Weitzel, M. *et al.* (2013) 13CFLUX2 - high-performance software suite for ^{13}C -metabolic flux analysis. *Bioinformatics*, **29**, 143–145.
- Weitzel, M. *et al.* (2007) The topology of metabolic isotope labeling networks. *BMC Bioinformatics*, **8**, 315.
- Wiechert, W. (2001) ^{13}C metabolic flux analysis. *Metab. Eng.*, **3**, 195–206.

- Wiechert, W. et al. (2001) A universal framework for ^{13}C metabolic flux analysis. *Metab. Eng.*, **3**, 265–283.
- Wiechert, W. et al. (1997) Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis. *Biotechnol. Bioeng.*, **55**, 118–135.
- Wiechert, W. et al. (1999) Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol. Bioeng.*, **66**, 69–85.
- Wiechert, W. et al. (2013) Metabolic isotopomer labeling systems. Part III: path tracing. *Math. Biosci.*, **244**, 1–12.
- Wiechert, W. and de Graaf, A.A. (1997) Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. *Biotechnol. Bioeng.*, **55**, 101–117.
- Wiechert, W. and Nöh, K. (2013) Isotopically non-stationary metabolic flux analysis: complex yet highly informative. *Curr. Opin. Biotechnol.*, **24**, 979–986.
- van Winden, W. et al. (2002) Cumulative bondomers: a new concept in flux analysis from 2D [^{13}C , ^1H] COSY NMR data. *Biotechnol. Bioeng.*, **80**, 731–745.
- Winter, G. and Krämer, J.O. (2013) Fluxomics - connecting 'omics analysis and phenotypes. *Environ. Microbiol.*, **15**, 1901–1916.
- Zamboni, N. (2011) ^{13}C metabolic flux analysis in complex systems. *Curr. Opin. Biotechnol.*, **22**, 103–108.
- Zamboni, N. et al. (2009) ^{13}C -based metabolic flux analysis. *Nat. Protoc.*, **4**, 878–892.
- Zhou, W. and Nakhleh, L. (2012) Quantifying and assessing the effect of chemical symmetry in metabolic pathways. *J. Chem. Inf. Model.*, **52**, 2684–2696.