Software applications toward quantitative metabolic flux analysis and modeling

Thomas Dandekar, Astrid Fieselmann, Saman Majeed and Zeeshan Ahmed

Submitted: 5th April 2012; Received (in revised form): 7th September 2012

Abstract
Metabolites and their pathways are central for adaptation and survival. Metabolic modeling elucidates in silico all the possible flux pathways (flux balance analysis, FBA) and predicts the actual fluxes under a given situation, further refinement of these models is possible by including experimental isotopologue data. In this review, we initially introduce the key theoretical concepts and different analysis steps in the modeling process before comparing flux calculation and metabolite analysis programs such as C13, BioOpt, COBRA toolbox, Metatool, efntool, FiatFlux, ReMatch, VANTED, iMAT and YANA. Their respective strengths and limitations are discussed and compared to alternative software. While data analysis of metabolites, calculation of metabolic fluxes, pathways and their condition-specific changes are all possible, we highlight the considerations that need to be taken into account before deciding on a specific software. Current challenges in the field include the computation of large-scale networks (in elementary mode analysis), regulatory interactions and detailed kinetics, and these are discussed in the light of powerful new approaches.

Keywords: metabolism; flux modes; metabolites; metabolic network

THEORETICAL BACKGROUND: FLUXES, NETWORKS AND METABOLITES
Metabolic flux analysis is currently a key method in metabolic modeling, its application has provided important insights into metabolism and adaptation of different organisms [1, 2]. Metabolic modeling is a broad field and involves a number of specific tasks. Here, we describe metabolite flux modeling tools and also the incorporation of isotopologue data. While textbooks describe the central metabolic pathways such as glycolysis in form of linear and distinct pathways, it is clear that such linear pathways are embedded in comprehensive metabolic networks. Therefore, the design of specific algorithms is required to calculate all possible pathways in the context of a complex network.

Before discussing and comparing individual software solutions, we briefly introduce some of the theoretical concepts underlying such modeling work: (i) First the network of participating metabolites and enzymes has to be established, this is referred to as the metabolic reconstruction. (ii) Then the network can be analyzed in terms of structure, resources and adaptability. (iii) Next, a more dynamic step is modeling of metabolic fluxes according to the enzyme repertoire as well as the modeling of the metabolic adaptation to different external constraints. (iv) The incorporation of experimental data such as isotopologue data for validation and refinement of the calculated models and fluxes is important. (v) Finally, detailed dynamic analyses of enzymes, such as metabolic control coefficients or detailed modeling of metabolic subnetworks including...
differential equations is possible. Each of these steps has specific challenges (Table 1).

### Metabolic reconstruction

The list of enzymes involved in a metabolic model is often established from annotated genome data. Different annotation frameworks and sequence analysis techniques can be used for this step, but are not discussed here since they form a subject of their own. Complete biochemical pathway databases such as the KEGG database provide a useful source of information [3] and software containing tools such as KEGGbrowser [4] are able to import such data and quickly setup a specific model for the pathways of interest. Alternatively, systems biology markup language (SBML) and other source files for specific reconstructed metabolic networks are available from different repositories and individual researchers.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td><strong>Metabolic reconstruction</strong>&lt;br&gt;Key challenge: Find the exact set of enzymes and reactions involved in your metabolic network.</td>
</tr>
<tr>
<td>2.</td>
<td><strong>Analyze the network structure</strong>&lt;br&gt;Key challenges: where are well connected hub enzymes and metabolites, what about currency metabolites, centrality, side pathways, futile cycles, central pathways, intermediary metabolism, reversible and irreversible reactions, committed steps, spontaneous reactions, sources, drains, etc.</td>
</tr>
<tr>
<td>3.</td>
<td><strong>Flux balance- and constraint-based modeling</strong>&lt;br&gt;Key challenges: How does the set of flux balanced pathways look like? How many elementary modes are possible in your system? How many extreme pathways? Calculate a convex basis vector.</td>
</tr>
<tr>
<td>4.</td>
<td><strong>Verifying flux predictions: isotopologue data</strong>&lt;br&gt;Key challenge: Which of the predicted fluxes is real and how strong is it?</td>
</tr>
<tr>
<td>5.</td>
<td><strong>Detailed modeling of metabolism and its kinetics</strong>&lt;br&gt;Key challenges: How does regulation work in detail? Use detailed differential equations to model fluxes and enzyme kinetics as well as metabolites.</td>
</tr>
</tbody>
</table>
Analysis of network structure
Metabolic networks have specific structures. These reflect the various tasks that the specific metabolic network has to fulfill: Key metabolites have to be produced from the inputs, others, such as vitamins and essential amino acids have to be imported. Anabolism converts single molecules into polymers (such as proteins, carbohydrates, lipids, nucleic acids and the compound molecules derived from these) as well as intermediary substances, signaling molecules and hormones. Furthermore, energy generation and degradation is important. In addition to the structure of these networks, their topology can also be considered and both can be linked [5]: Structural analysis typically takes into account the stoichiometric coefficients, while in elucidating the topological properties, such as centrality and connectivity, the stoichiometric coefficients are not used.

Analyzing in detail the structure of a metabolic network not only reveals information on the key pathways, but also reveals organism specific differences. In addition it provides a large amount of information on the direct construction of the metabolic network such as well-connected hub metabolites, different subnetworks, currency metabolites, alternative pathways and futile cycles. Futile cycles use in general two different enzymes for the forward- and the reverse reaction and can be determined using elementary flux modes (EFMs [6], see below). They convert a substrate to a product and back again. This can be used for energy generation (e.g. in brown fat) but in most cases allow a switch between a catabolic (e.g. glycolysis) and an anabolic (e.g. gluconeogenesis) condition. Hence, the involved enzymes are well regulated and the identification of potential futile cycles reveals potential regulatory switches and regulated enzymes in the system.

Flux balance- and constraint-based modeling
Flux balance- and constraint-based modeling (CBM) calculate different pathways in a network to determine dynamic properties from the network graph with its nodes and edges [step (ii)]. For compound A with concentration [A] and its change (d/dt) over time, one can write the differential equation:

\[
\frac{d[A]}{dt} = f ([A], [X], k_i)
\]

where [X] are the concentrations of other compounds and \(k_i\) the kinetic constants. FBA reduces the computational difficulty of the differential equations by relying on the steady state assumption that \(d[A]/dt = 0\) at all times and for all metabolites. This considerably simplifies the problem to that of the flux balance within the system, for instance the rate of consumption of a metabolite \(v_{\text{cons}}\) should be balanced by the rate of its production \(v_{\text{prod}}\), so

\[
v_{\text{cons}} - v_{\text{prod}} = 0
\]

(2)

Methods for FBA balance all metabolites within a cellular system by a suitable combination of enzymes: For each internal metabolite, the same amount is consumed as is produced. The stable and balanced metabolite flux achieved by such a combination of enzymes is called a flux mode. The representation of the above equation can be generalized to any similar biological network or tackled in a more powerful way by using matrices. The stoichiometric matrix \(S\) denotes for each internal metabolite \(m_i\) (rows in the stoichiometric matrix) whether it is consumed (− sign) or produced (+ sign) by a certain reaction \(r_i\) (columns in the matrix) or not occurring in that reaction (zeroes in the matrix), producing from the molar quantities consumed and produced for each metabolite in each reaction matrix:

\[
S = \begin{pmatrix}
-1 & 0 & -1 & m_1 \\
0 & 0 & 1 & m_2 \\
0 & 2 & -1 & m_3 \\
1 & -1 & 0 & m_4 \\
\end{pmatrix}
\]

In the simple example given, the first enzymatic reaction \(r_1\) produces the internal metabolite four \((m_1)\), consumes \(m_1\) and leaves all other metabolites unchanged. The second enzyme \(r_2\) produces two molecules of \(m_3\) and consumes one molecule of metabolite \(m_4\), again leaving all others unchanged. By linear combination of reactions the balancing condition above (Equation 2) is solved for all internal metabolites, keeping track of the enzymes involved and solving the stoichiometric matrix so that the stationary condition is fulfilled for all metabolites, i.e.

\[
S \cdot v = 0
\]

(3)

In general, flux balance (or steady-state constraint) is an essential condition of CBM. Thermodynamic constraints can furthermore reduce the solution space. The concept of elementary modes (EMs) was introduced to extend the property of
non-decomposability of the extreme rays of a flux cone encompassing all permissible fluxes to networks with reversible fluxes. In contrast, FBA constraint-based approaches (COBRA toolbox, classical and dynamic FBA, iMAT) do not enumerate the individual pathways, as they obtain a general model indicating the flux strengths of each reaction.

Thus, FBA and EFMs are two main approaches of CBM that consider balanced fluxes (steady-state, mass conservation). In addition, other approaches add extra constraints to include experimental information in order to obtain or verify the resulting metabolic fluxes ([7], for a detailed review).

An EM is a mode that cannot be decomposed further without creating an imbalance between the internal metabolites. Biochemical knowledge helps to determine irreversible and reversible enzyme reactions. Twice as many modes are possible changing a reaction from irreversible to reversible, e.g. if the irreversible reaction occurred in all EMs and now the same number of modes is recreated including the enzyme working in the opposite direction. However, in general the effect of adding modes by incorporating reversible reactions is not so drastic [8].

External metabolites provide sources and drains in and out of the system and therefore do not need to be balanced within the system. However, one can consider how the flux modes of the system convert different external source metabolites into products draining from the system.

In practice, many such flux modes are possible for a moderate sized set of enzymes (30–50) and it can be further distinguished whether the enzyme combination can not be reduced or split further without losing the ability to balance all used internal metabolites, which would then be an EM [9].

A convex basis is a specific set of EMs sufficient to reproduce all other modes by linear combination [10]. It describes all steady-state flux distributions or flux pathways in the network. It is not unique and is constituted by a minimal generating set of EMs [11]. The basis vectors that lie at the edge of a high-dimensional cone have been called ‘extreme pathways’ since they represent extreme states of the network. Extreme pathways form a unique basis set for describing the flux pathways and provide a systemic definition for the topological features of the metabolic networks [12].

Modifications to and extensions of metabolic flux analysis have continuously been proposed [13, 14] including software packages such as Metatool [15], Classical and Dynamic FBA [16]. For larger metabolic networks, there is the challenge of combinatorial explosion yielding very large numbers of modes to calculate. To address this, the Palsson group changed their algorithm to sampling over the full network by the uniform sampling approach [17, 18]. Further options are provided by the COBRA toolbox. Elementary flux patterns [13] calculate modes in a subnetwork and extend them over the rest of the network. Centler et al. [19] calculate flux modes consecutively in a parallelized, highly efficient routine. It is difficult to compare these alternatives in a fair manner, however, these approaches are certainly highly advantageous in large networks where calculation of all EMs is no longer possible.

Computing EMs can be considered equivalent to extreme ray enumeration of polyhedral cones. Instead, bit pattern trees distinguish extreme rays from normal (composite) vectors and the efimtool [20] features a new recursive enumeration strategy with bit pattern trees for adjacent rays (as ancestors of extreme rays). Its parallelized fast rank updating method includes a residue arithmetic method for matrix rank computations to circumvent potential numerical instability problems. Currently, this should be the fastest method to calculate EMs as shown in calculations on genome scale models for Escherichia coli and Helicobacter pylori.

Flux analysis only describes the scenario of all possible metabolic routes. To determine which of the different pathways is used under a given situation, a best fit according to experimental data on enzyme activity or any other surrogate indicator is required. Such software (e.g. YANA [21] or ReMatch [22]) evaluates the pathway strengths for the different modes and calculates the resulting activities for individual enzymes. This minimizes the error in the observed experimental data of the enzyme activity for the complete metabolic system under consideration. Surrogate indicators for enzyme activity are proteome data (which ignores difference in enzyme regulation and individual activity) or gene expression data (which in addition ignore different protein synthesis rates and protein stability). This first estimate can be further refined by other data, e.g. external metabolite concentrations [21]. Sufficient data points are necessary as otherwise there is not enough information to infer all pathway fluxes correctly.
CBM [23, 24] may include other constraints, e.g. a growth equation which primary compounds (amino acids, nucleic acids, sugars ...) are required in which stoichiometry for optimal biomass growth to calculate optimal fluxes \( v \):

\[
\text{max}(v_{\text{biomass}})
\]

This calculation is performed under conditions where the stoichiometric matrix condition has to be satisfied (Equation 3). For example in determining the optimal yield of citrate (as desired product) from glucose as input in *Aspergillus* would be another constraint-based pathway optimization problem [25]. Other objective functions can also be optimized, e.g. maximizing ATP production or minimizing \( \text{O}_2 \) consumption. Furthermore, genetic mutations can be modeled and their essentiality for growth predicted, this approach has additional applications in biotechnology and biomedicine [26].

**Verifying flux predictions: isotopologue data**

Isotopologue data provide strong support to compare and validate modeled fluxes by experimental data. Such methods follow the course of a labeled substance in the metabolic network. For example, an amino acid with certain carbon atoms labeled by a heavy, non-natural isotope such as \( ^{13}\text{C} \). During metabolic processing, the labeled substance (e.g. \( ^{13}\text{C} \)-labeled glucose) is oxidized and further compounds are created while the substance is metabolized, these can be identified since they carry the \( ^{13}\text{C} \) label. To trace metabolic processing in this way is non-trivial but yields important information such as which of two potential possible pathways and enzymes are used for the conversion of the metabolite into its products. A number of software packages have been developed to assist in the processing of such isotopologue data, such as FiatFlux and C13. Here we will focus only on \( ^{13}\text{C} \) data, however, analogous software enables the estimation of nucleotide or lipid concentrations or treats \( ^{15}\text{N} \) data [27]. The traced metabolic route is a validation for the calculated flux modes and is helpful in the investigation of the activity of the different modes under specific metabolic conditions.

Finally, no matter which of these preparatory steps is the utilized, the graphical representation of the described pathways is again important and some of the discussed software is capable of producing these types of visualizations.

**Detailed modeling of metabolism and its kinetics**

Each metabolic reaction follows its own kinetics, the most well-known one is Michaelis–Menten kinetics. However, regulatory input, allosteric effects of different metabolites as well as protein interactions and direct regulation modulate the kinetics of a metabolic reaction as well as fine tuned effects on the enzymes structure (e.g. induced fit, cooperative interactions). Starting from individual reactions, different types of metabolism may be modeled in detail on this kinetic level. Different key aspects such as the metabolic control by key enzymes have been studied [25]. However, to obtain useful metabolic models on this level of dynamic resolution requires detailed information on kinetic constants, interactions and regulatory parameters that are only available for well-studied regions of the metabolic map (e.g. cyclic nucleotide metabolism in platelets [28]). Typically, differential equations capture individual parameters and kinetics and powerful software exists for this purpose. As this is again a topic for a review in its own right, we will mention only briefly in the next paragraph suitable software for this most detailed analysis step in metabolic modeling.

**APPLYING THE OPTIMAL MODELING SOFTWARE**

For each of the steps (i) to (v) mentioned in the Section ‘Theoretical Background’, suitable software improves calculations and drastically speeds up the analysis time. This is essential for large-scale networks. Our recommended software covers steps (i) to (iv) and is introduced accordingly.

(i) Metabolic reconstruction requires the correct compilation of involved enzymes and reactions for the organism and metabolism of interest. Useful software for this step often builds on the large biochemical database KEGG, for instance the KEGGconverter [29]. From the recommended software, the YANA suite of programs has the software tool KEGGbrowser [4]. Alternatives are the provided and well-curated full-scale metabolic networks supplied by iMAT [30]. Furthermore, Metannogen [31] allows the browsing and annotation of existing networks in SBML.

(ii) After metabolic reconstruction an analysis of the network in terms of structure, resources and
adaptability is possible. This can be investigated with the Systems Biology Toolbox for MATLAB [32]. This offers systems biologists a powerful, open and user extendable environment, in which models of biological systems can be created and it includes the basic tools required to analyze network structure such as centrality and connectivity. From our selection of versatile tools, network structure can be easily analyzed and visualized using VANTED [33] or YANAsquare [4]. Both allow different overviews of the complete metabolic network structure as well as zooming in on networks of interest. Furthermore, most metabolic modeling software discussed below includes useful and easy understandable visualization options. Thus, iMAT visualizes the ready-made complete metabolic network in an intuitive way and the COBRA toolbox offers customized visualization of metabolic maps to enable structure analysis.

(iii) FBA as a core task to calculate the constraint-dependent (see ‘Theoretical Background’ section) permissible flux modes in a given network can be achieved by software such as Metatool [11] or BioOpt [34]. The software Classical and dynamical FBA [16] was developed further into the COBRA Toolbox [35], which is useful in a broad and versatile manner. It is particularly useful in CBM including besides experimental data further constraint conditions such as growth equations. iMAT [30] is particularly rapid and easy to use if gene expression data or protein expression data are used as constraints on the large basis of custom-made metabolic models on different organisms. Furthermore, a rapid fit of metabolic flux distributions to data points from gene and protein expression can be achieved and visualized by YANAsquare [4] and, for large-scale networks and big data sets by YANAvergence [36].

(iv) Flux verification by direct measurement such as isotopologue data is achieved by applying FiatFlux [37]. $^{13}$C-constrained flux analysis is furthermore supported by C13 [38], Openflux [39] and ReMatch [22].

(v) For detailed dynamic analyses of enzymes such as metabolic control coefficients or detailed modeling of metabolic subnetworks with differential equations software applications include the BioMet Toolbox [34], CellNetAnalyzer (formerly FluxAnalyzer) [40], COPASI [41], Jarnac [42], Pathway Analyser [43], TinkerCell [44], WebCell [45] and SCAMP [46]. Furthermore, the metabolic modeling tool (MMT) [47] has a relational database specifically for the analysis of rapid sampling experiments. The tool allows construction of complex pathway models with information stored in the relational database. MMT contains parameter fitting and simulation algorithms for the resulting system of ordinary differential equations and the calculation of explicit sensitivity functions. A good application example is the detailed analysis of *E. coli* glucose pulse experiments.

The choice of the software depends on which of the five steps is considered (Table 1). Each software has strengths in specific areas of analysis, and this also depends on experience with using the software. Another aspect is the type and amount of assistance the user expects from the software. In particular, there are large differences in the amount of visual output generated which we will briefly discuss. For the calculations, strictly speaking, this is not necessary and the same applies to all other sorts of result presentation, e.g. detailed listings or specific outputs generated by different software. The importance of user-familiarity with the software is not to be underestimated. Finally, one could think of combining different software for an optimal path of calculation, for instance starting from metabolic reconstruction, calculating EMs and only then placing specific constraints on the results including validation and comparison with experimental data, before evaluating the results in the end.

However, as visualization can happen at any step and the detail of the metabolic network analysis would follow more closely the five steps detailed earlier, we have decided to present the software in the above order of the five steps. We will make clear for the more versatile programs where they have additional capabilities relevant to other steps of analysis.

The tests and data provided by the different applications often support each other for subsequent analysis steps. Typical tasks include the reconstruction of the glycolysis and pentose phosphate pathway in the organism of choice [10], mapping isotopologue data onto these pathways [48], calculating EMs and resulting fluxes according to the data and further visualization and comparison of results [36]. Their application generally aims for biological insights,
e.g. mutations in any of the central metabolic pathways with effects on survival in macrophages such as mutations in the aldolase gene in *Listeria* [49]. Each of the software mentioned in the following (web links summarized in Table 2) aims to cover several steps of this analysis chain. However, in its details, each package provides different advantages and limitations and excels in different steps of the metabolic modeling analysis chain (Table 3).

This makes it difficult to perform a fair comparison of the software when selecting and applying a specific test data set to each platform. However, we think that the individual descriptions following will at least allow the reader to make an educated choice depending on the analysis step he or she is interested in (Supplementary Data illustrates the software discussed).

**BioOpt**

BioOpt is developed in C++ and uses integer linear programming principles to perform FBA. Using input reactions, constraints, external metabolites and an objective function, it calculates all internal mass balance fluxes, reduced costs and shadow prices depending on the constraints and objectives (maximize, minimize) defined by the user. BioOpt’s linear programming identifies the best set of gene deletions for a given objective function value, such as growth rate of an organism according to a given growth equation. It has been successfully applied to genome scale metabolic models, e.g. of *A. oryzae*. It implements an exhaustive combinatorial search for gene deletions and includes a basic sensitivity analysis [34]. Results include the calculated objective value, transformations and achieved reactions, internal fluxes as well as according to the constraints set cost, metabolites and shadow price. BioOpt is able to use Metatool 4.3 as a third party tool, to compute the null space matrix, EMs and other structural properties of biochemical reaction networks (convex basis, subsets and stoichiometric matrix) by directly passing the input to Metatool.

**C13**

C13 was initially developed during the PhD course ‘Linear and non-linear optimization methods’ at the

### Table 2: Software web links

<table>
<thead>
<tr>
<th>Description</th>
<th>Web link</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioLayout Express 3D</td>
<td><a href="http://www.biolayout.org/">http://www.biolayout.org/</a></td>
</tr>
<tr>
<td>COBRA Toolbox</td>
<td><a href="http://opencobra.sourceforge.net/openCOBRA/Welcome.html">http://opencobra.sourceforge.net/openCOBRA/Welcome.html</a></td>
</tr>
<tr>
<td>C (programming language)</td>
<td><a href="http://www.cprogramming.com/">http://www.cprogramming.com/</a></td>
</tr>
<tr>
<td>Description of algorithm for computing elementary flux modes</td>
<td><a href="http://pinguin.biologie.uni-jena.de/bioinformatik/networks/metatool/algorithm.pdf">http://pinguin.biologie.uni-jena.de/bioinformatik/networks/metatool/algorithm.pdf</a></td>
</tr>
<tr>
<td>efmtool</td>
<td><a href="http://www.csb.ethz.ch/tools/efmtool">http://www.csb.ethz.ch/tools/efmtool</a></td>
</tr>
<tr>
<td>Flux Balance Optimization</td>
<td><a href="http://arep.med.harvard.edu/gmc/fba.html">http://arep.med.harvard.edu/gmc/fba.html</a></td>
</tr>
<tr>
<td>i-MAT</td>
<td><a href="http://www.cs.technion.ac.il/~tomersh/methods.html">http://www.cs.technion.ac.il/~tomersh/methods.html</a></td>
</tr>
<tr>
<td>Metatool</td>
<td><a href="http://pinguin.biologie.uni-jena.de/bioinformatik/networks/metatool.html">http://pinguin.biologie.uni-jena.de/bioinformatik/networks/metatool.html</a></td>
</tr>
<tr>
<td>Metatool 5.0 for GNU octave and MATLAB</td>
<td><a href="http://pinguin.biologie.uni-jena.de/bioinformatik/networks/metatool5.0.html">http://pinguin.biologie.uni-jena.de/bioinformatik/networks/metatool5.0.html</a></td>
</tr>
<tr>
<td>NetCDF</td>
<td><a href="http://www.unidata.ucar.edu/software/netcdf/">http://www.unidata.ucar.edu/software/netcdf/</a></td>
</tr>
<tr>
<td>Octave</td>
<td><a href="http://www.gnu.org/software/octave/">http://www.gnu.org/software/octave/</a></td>
</tr>
<tr>
<td>ReMatch</td>
<td><a href="http://www.cs.helsinki.fi/group/sysfys/software/rematch/">http://www.cs.helsinki.fi/group/sysfys/software/rematch/</a></td>
</tr>
<tr>
<td>Sample metabolic networks and input files for Metatool</td>
<td><a href="http://pinguin.biologie.uni-jena.de/bioinformatik/networks/metatool5.0/ecoli_networks.html">http://pinguin.biologie.uni-jena.de/bioinformatik/networks/metatool5.0/ecoli_networks.html</a></td>
</tr>
<tr>
<td>MATLAB</td>
<td><a href="http://www.mathworks.com/">http://www.mathworks.com/</a></td>
</tr>
<tr>
<td>VANTED</td>
<td><a href="http://vanted.ipk-gatersleben.de/index.php?file=doc0.html">http://vanted.ipk-gatersleben.de/index.php?file=doc0.html</a></td>
</tr>
<tr>
<td>YANASquare</td>
<td><a href="http://www.tr34.uni-wuerzburg.de/computations/yanasquare/">http://www.tr34.uni-wuerzburg.de/computations/yanasquare/</a></td>
</tr>
<tr>
<td>YANA</td>
<td><a href="http://yana.bioapps.biozentrum.uni-wuerzburg.de">http://yana.bioapps.biozentrum.uni-wuerzburg.de</a></td>
</tr>
<tr>
<td>YANAvvergence</td>
<td><a href="http://www.biozentrum.uni-wuerzburg.de/bioinformatik/">http://www.biozentrum.uni-wuerzburg.de/bioinformatik/</a></td>
</tr>
</tbody>
</table>
### Table 3: Software specifics and advantages

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioOpt</td>
<td>Calculates all internal mass balance fluxes, reduced costs and shadow prices&lt;br&gt;Identifies best set of gene deletions for given objective function value&lt;br&gt;Implements exhaustive combinatorial search for combinations of gene deletions and over expression of fluxes and basic sensitivity analysis&lt;br&gt;Capable of using a third party tool (METATOOL 4.3)</td>
</tr>
<tr>
<td>BioLayout Express3D</td>
<td>Imports data from various standard graph formats&lt;br&gt;Handles up to 500 to 1000 expression arrays&lt;br&gt;Cluster of graphs of up to 30000 nodes and 2-3 million edges&lt;br&gt;Construct network graphs&lt;br&gt;Converts a 2D representation to a 3D graph&lt;br&gt;Calculates correlation matrix&lt;br&gt;View and navigate results in the 3D interface&lt;br&gt;Alters aesthetic characteristics of the 2D and 3D interfaces&lt;br&gt;Cluster Graph using the MCL, by defining the granularity of the clustering by setting high and low inflation values&lt;br&gt;Identifies genes of interest and export list of selected genes for further use&lt;br&gt;Mine selected genes for overrepresentation of classes, the number of calculations necessary to perform this task inevitably makes this process slow&lt;br&gt;Works with other data formats and edit networks using GraphML input file&lt;br&gt;Select nodes (by a number of ways) for the editing of their properties, e.g. activation (A), inhibition (I), translocation (T) or logic nodes&lt;br&gt;Render pathways in 2D and 3D modes</td>
</tr>
<tr>
<td>CI3</td>
<td>Estimate fluxes satisfying stoichiometric constraints&lt;br&gt;Resolve limited enrichments by isotope balances around carbon atoms&lt;br&gt;Computes deviation between fluxes and between fractional labeling</td>
</tr>
<tr>
<td>COBRA toolbox</td>
<td>Constraint-based modeling&lt;br&gt;Constraint-based reconstruction including network gap filling&lt;br&gt;Simulation and analysis of phenotypes&lt;br&gt;Genome-scale models and -omics-guided analysis&lt;br&gt;CI3 analysis&lt;br&gt;Metabolic engineering</td>
</tr>
<tr>
<td>efmtool</td>
<td>Provides very fast (in fact fastest currently known) elementary mode calculation&lt;br&gt;Introduces bit pattern trees&lt;br&gt;Rank updating method well suited for parallel computation</td>
</tr>
<tr>
<td>FiatFlux</td>
<td>User friendly and first publicly available software for flux ratio analysis&lt;br&gt;Computes metabolic flux ratios exclusively from MS data in the RATIO module&lt;br&gt;Estimates net carbon fluxes within a comprehensive model of metabolite balances from measured extracellular fluxes, previously determined flux ratios and biomass requirements.&lt;br&gt;Estimates error using flux ratio from $^{13}$C labeling&lt;br&gt;Estimates flux using measured extracellular rates.</td>
</tr>
<tr>
<td>iMAT</td>
<td>Ready made metabolic networks for more then 160 organisms&lt;br&gt;Integrates transcriptomic and proteomic data with genome-scale metabolic network models&lt;br&gt;Optimized fit of gene expression or proteomic data to identify stronger and weaker fluxes</td>
</tr>
<tr>
<td>Metatool</td>
<td>Gives numbers of internal metabolites and reactions&lt;br&gt;Parses reaction equations and translates them into a stoichiometric matrix&lt;br&gt;Identifies kernel or null space&lt;br&gt;Provides fast and simple elementary flux mode calculations&lt;br&gt;Tackles middle-large reaction systems&lt;br&gt;Integrates with other third party tools&lt;br&gt;Runs using GNU Octave and MATLAB environments&lt;br&gt;Capable of computing structural invariants like conservation relations, enzyme subsets and fits a power law to the connectivity distribution of metabolites.</td>
</tr>
<tr>
<td>ReMatch</td>
<td>First web-based tool capable of metabolic network model construction, store and sharing&lt;br&gt;Integrates carbon mappings for $^{13}$C metabolic flux analysis&lt;br&gt;Allows combining user developed models from several comprehensive metabolic data resources into a common repository for metabolic network models&lt;br&gt;Generates stoichiometric matrix and visualizations&lt;br&gt;Resolves conflicts between the nomenclature and augmented reactions with carbon mappings (if available) in model and existing information in database, fully/semi automatic matching is performed among user given reactions and reactions stored in database.&lt;br&gt;Visualizes the metabolic network&lt;br&gt;Exports in use model in SBML, $^{13}$C-FLUX or stoichiometric matrix formats&lt;br&gt;Shares exported model by declaring it public in ReMatch platform</td>
</tr>
</tbody>
</table>

(continued)
Center for Microbial Biotechnology, BioCentrum-DTU in 2002. It is a useful tool for the quantification of in vivo metabolic fluxes given carbon labeling experiments. In C13, FBA (see ‘Theoretical Background’ section) is based on stationary carbon isotope labeling experiments, using fractional enrichment data. The mathematical framework adopted for C13 in MATLAB was developed by Wiechert and de Graaf [38].

Inputs to the program include the measurements of extracellular rates (e.g. substrate uptake rates, product formation rates, drain fluxes into biomass) and fractional enrichment data, which also consider the labeling state of the substrates. Furthermore, a metabolic model file is required (see ‘Theoretical Background’ section) comprising all relevant biochemical conversions, reversible and irreversible enzymes and/or transport processes and the fate of all carbon atoms throughout the metabolic network.

C13 thus performs the metabolic flux analysis by implementing the concept of isotopic labeling (of carbon atoms), analyses fluxes and fractional labeling and obtains the required results by iterative error minimization. First an estimate of the fluxes satisfying the stoichiometric constraints is done. Next limited enrichments are calculated according to isotope balances around carbon atoms. Then the deviation between fluxes and fractional labeling is calculated. These three steps are repeated until the deviation is below a certain threshold.

Specific lower and upper constraint boundaries (see ‘Theoretical Background’ section) on net and exchange fluxes can be defined as well as substrate uptake rates.

The COBRA toolbox
This toolbox is a MATLAB package developed by one of the leaders in the field of flux balance [16], Palsson and co-workers, and a research alliance studying systems biology of entero pathogens. Models are exported in SBML format. It is particularly strong in constraint-based reconstruction and analysis methods (hence the name COBRA) to simulate, analyze and predict a variety of metabolic phenotypes using, e.g. genome-scale models. Recent new functions [50] allow (i) network gap filling, (ii) $^{13}\text{C}$ analysis, (iii) metabolic engineering, (iv) omics-guided analysis and (v) visualization. The performance as well as the documentation has been improved further. The philosophy behind the approach acknowledges the challenge to model an organism at genome-scale. It focuses on physicochemical constraints to define the set of feasible states for a biological network in a given condition based on current knowledge (see also ‘Theoretical Background’ section; the step 5, regulation and detailed kinetics in subnetworks is not the focus of the COBRA toolbox). These include compartmentalization, mass conservation, molecular crowding and reversible/irreversible enzyme reactions as well as transcriptome data that are integrated as constraints or an objective function to be optimized such as a growth equation [51]. Depending on the algorithm COBRA does not provide a unique solution but a reduced set of solutions for well-founded biological hypothesis generation, e.g. in calculation of minimization of metabolic adjustment.

A static optimization approach divides a batch time into time intervals and solves an instantaneous optimization problem at the beginning of each

---

Table 3: Continued

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VANTED</td>
<td>A platform-independent Java-based software tool that provides a framework for visualization of experimental (-omics) data and statistical analysis. Allows user to integrate complex structured data sets and connect several values to one single network element by presenting them as, e.g. line- or bar-charts. Supported input and output network formats are, e.g. GML, SBML and Pajek.NET. Provides the visualization of flux distributions from various sources with recently developed add-ons such as FluxMap. Provides model set-up and modification with the choice to do a KEGG import for a rapid retrieval of biological networks. Takes input and produces output in SBML or Metatool format. Performs Internal elementary calculation using Metatool. Calculates flux distributions for each elementary mode activity set by a genetic algorithm and visualizes them by different edge thicknesses.</td>
</tr>
</tbody>
</table>
interval by the application of repeated linear programming to obtain flux distribution at a particular time instant. Furthermore, a dynamic optimization approach considers time profiles of fluxes and metabolite levels, using non-linear programming. COBRA excels when complex constraints have to be taken into account.

As an open project and powerful package, COBRA has attracted attention from many investigators and has been applied to many studies including a repository for community contributed modules. It is advantageous that a number of models are available from the repository, e.g. different E. coli genotypes and metabolic pathway models for the full core metabolism including glycolysis, pentose phosphate pathway, TCA cycle and respiration as well as further pathways. These should be imported carefully before the MATLAB scripts of the package are applied. Particularly useful for becoming acquainted quickly with COBRA is a suite of test scripts covering the first four steps of analysis mentioned in the Theoretical Background that can be used to learn the core functionality of the toolbox and validate its results.

FiatFlux

FiatFlux deals with the problem that in vivo metabolite fluxes are inferred indirectly from measured quantities in 13C experiments. To allow such complex analyses also for non-expert groups, a conceptual simplification is introduced, the analytical determination of metabolic flux ratios exclusively from mass spectrometry (MS) data, which then are used to estimate absolute in vivo fluxes. FiatFlux is user-friendly and MATLAB based. It is one of the first publicly available software for flux ratio analysis in labeling mixtures, including various substrates. ‘The software is preconfigured to derive flux ratios and absolute in vivo fluxes from 13C-glucose experiments and GC-MS analysis of amino acids from a variety of microorganisms.’

FiatFlux is divided into two different and independent modules. In the first module, ‘ratios of converging fluxes are automatically calculated from GC-MS-detected 13C-pattern in protein-bound amino acids. Predefined fragmentation patterns are automatically identified. Their statistical treatment is based on the comparison of redundant information in the MS spectra’. This RATIO module performs 13C-constrained flux analysis [52]. It works exclusively from MS data and estimates net carbon fluxes within a comprehensive model of metabolite balances from measured extracellular fluxes, previously determined flux ratios and biomass requirements [53].

In a second module, absolute intracellular fluxes are calculated by 13C-constrained FBA. Experimentally determined fluxes in and out of the cell are combined with the above flux ratios.

The overall internal work-flow of FiatFlux starts working with a textual input file (*.txt) consisting of reactions, ratios and biomass precursors. Reactions are unique identifiers, ratios are the equality/inequality constraints (‘=’, ‘>’, ‘<’) and biomass precursors are calculated according to growth rate dependent withdrawals of metabolites in μmol/gCDW. These parameters are inputs for FiatFlux Netto to compute flux using measured extracellular rates and to estimate errors using flux ratio from 13C-labeling [54]. Initially, mass balances and flux ratios are identified and then an appropriate method is selected. Furthermore, depending upon the active set of constraints and reactions, the input system can be undetermined, determined or overly constrained. In case of an undetermined system, possible fluxes are estimated and then the objective function is maximized. If the system is determined, manual analytical skills are needed for good results, and if the system is constrained then it is solved by best-fit optimization.

FiatFlux aims to be helpful for the non-specialist, however, the input procedure for FiatFlux is quite complex and third party tool dependent (which requires at least basic knowledge of MATLAB and NetCDF, the network common data format describing arrays of data).

In FiatFlux, ‘user supervision is necessary only when MS signals are low, saturated or overlapping’ [54]. This ‘affects the ion statistics of the corresponding fragment and results in relatively high residuals’ [54] after the first estimates of the mass isotopomer distribution vector. FiatFlux RATIO is not compatible with MS/MS product ion scans. Microsoft Windows is preferred as an operating system, because some minor problems were encountered using MATLAB’s graphic user interface with Linux.

Metatool

Metatool was originally programmed by Thomas Pfeiffer using the C programming language for EFM calculation and using the algorithm published by [15]. Early versions of Metatool are simple
stand-alone applications capable of taking the input and providing an output in a simple text file format. For efficient EFM calculations with the application of a new and fast algorithm for scripted input files from shared libraries, a new platform independent version of Metatool (5.0) has been developed that is capable of providing fast EFM calculations, tackling larger reaction systems, integrating with other third party tools and being able to run using GNU Octave and MATLAB environments. Along with all new features Metatool 5.0 is also capable of processing input files of previous Metatool versions.

In addition, flux mode calculations are covered with increasing performance by a number of related software (see also ‘Theoretical Background’ section and efmtool). To give further work for comparison, the Null-Space algorithm proposed by Urbanczik and Wagner [55] calculates elementary fluxes of chemical reaction systems based on stoichiometry matrices (see ‘Theoretical Background’ section) about 20 times faster than the previously used algorithm by Schuster et al. [15]. Correspondingly, this is used in the developed version of Metatool (5.0) [56]. It also internally distinguishes stoichiometric matrices and performs exact representation of numbers using integer data format. Moreover, Metatool (5.0) is also capable of computing structural invariants such as conservation relations, enzyme subsets and fits a power law to the connectivity distribution of metabolites. Conversion between DOS- and UNIX-based systems by unix2dos and dos2unix may sometimes pose a challenge for feeding in the last line of the input file. Finally, Octave and MATLAB are not 100% compatible with each other.

**iMAT**

iMAT integrates ‘transcriptomic and proteomic data with genome-scale metabolic network models’ to predict metabolic fluxes [30]. Based on earlier work, Shlomi et al. [57] provide a well-curated human metabolic network as well as application of mixed integer linear programming to find a steady-state flux distribution that satisfies stoichiometric and thermodynamic constraints, while maximizing the number of reactions whose activity is consistent with their expression state. An optimized fit of gene expression or proteomic data to identify stronger and weaker fluxes under a given condition in the calculated flux network is a powerful strategy found again in other software discussed here, e.g. YANAsquare [4] and COBRA toolbox [35]. As with the other recommended tools, detailed metabolic modeling studies will need expert work by the user (metabolic reconstruction, consideration of all data, detailed analysis of fluxes obtained, reactions, internal and external metabolites and their model implementation).

However, iMAT's choice of ready made metabolic networks is a big bonus, including common model organisms such as *E. coli* and *Saccharomyces cerevisiae* and an array of automatically reconstructed networks for 160 bacteria provide a quick start for any interested user [30]. Furthermore, the output is a map visualizing the metabolic flux state of the organism, showing the most likely predicted metabolic fluxes across its reactions. Thus, this is a very useful tool for rapid interpretation of gene expression or proteomic data in terms of the resulting fluxes in a variety of organisms. This allows the user to easily base his flux predictions for expression data on the available curated models.

**ReMatch**

ReMatch [22] is a user-friendly web-based tool (Table 1) capable of metabolic network model construction. It stores, shares and integrates carbon mappings for $^{13}$C-metabolic flux analysis. Furthermore, models from different users or several comprehensive metabolic data resources (KEGG, MetaCyc and CheBI) can be combined into a common repository for metabolic network models.

Stoichiometric network models for metabolic flux analysis are constructed based on the user input and the above three databases. In general, this is a tedious task. It is the first step in our analysis (see ‘Theoretical Background’ section) and ReMatch is particularly helpful as it contains a comprehensive metabolite name thesaurus as well as stored metabolic databases and wild cards (if the enzyme name is not known by the user). Next, ReMatch ‘augments the metabolic reactions of the model with carbon mappings to facilitate $^{13}$C metabolic flux analysis’ [22]. Again, it uses for this the internal database or mappings provided by the user with an easy to use tool. The constructed models are exported into $^{13}$C-FLUX and SBML file formats. Furthermore, visualizations of the network model can be generated (using the BMVis viewer which was developed using graph mining tools). Sharing, if desired, of such models is optionally made available to the other users of ReMatch by a common repository for metabolic network models with carbon mapping.
VANTED

VANTED is a platform-independent Java-based software tool that provides a framework for visualization of experimental (–omics) data and statistical analysis [33]. The user can integrate complex structured data sets (using an Excel input form specifying experiment and different –omics data) and connect several values to one single network element by presenting them as, e.g. line or bar charts. VANTED allows the mapping of measurement data from different experiments onto arbitrary networks, which can be edited with the built-in graph editing functions.

In addition to general graph editor functions such as node/edge selection, modification or deletion, an algorithm for the removal of node overlaps [13], and layout algorithms such as circular, tree-shaped and force-directed are available. As alternatives to network creation, networks may be loaded with the built-in importer from the KEGG Pathway database [34] or from the standard file formats GML [58], SBML [58] and Pajek.net [29].

The added value of VANTED compared to standard network analysis is the integrated analysis of metabolic pathways, fluxes and flux distributions including their confidence limits. Networks can also be analyzed with the above standard network visualization tools such as Pajek.net [21, 29]. The data mapping procedure ‘will be done automatically if the substance name in the input form is equal to a target node label. Mapping procedures consider any synonym or identifier defined in the KEGG Ligand database [39, 40] or in the Swiss Institute of Bioinformatics Enzyme nomenclature database. Moreover, in the remaining cases where no automatic mapping is possible a user-defined mapping may be performed, in which a data subset for a measured substance is assigned manually to a node’ [33]. Furthermore, VANTED allows the automatic creation of new nodes for all measurement data subsets which do not map onto the given network. The mapping of data onto (optionally organism specific) KEGG pathways is facilitated by a function which counts the number of possible automatic mappings of the measured substances onto the list of KEGG pathways. VANTED easily processes time series metabolite data for instance from developing plant seeds.

A template-based flux data import ‘assigns flux values and optional quality parameters (e.g. the confidence interval) to biochemical reactions. It supports the discrimination between mass and substance fluxes, such as C- or N-fluxes. After import, flux data mapping and network-based visualization allows the interactive exploration of the data set. Various visualization options enable the user to adapt layout and metabolic network representation according to individual purposes’ [59]. FluxMap [60], as a new VANTED add on, allows exploration of flux distributions showing different edge thicknesses and provides several parameters to alter the appearance of a flux distribution (global edge thickness, arrowhead/-tail ratio, style of reactions nodes, colors show confidence limits). VANTED certainly fills a specific niche providing enhanced visualization to metabolic data and metabolism. However, it should be noted that in general visualization software is improving over time, an example is BioLayout Express 3D [61] as ‘an application that has been specifically designed for the integration, visualization and analysis of large network graphs derived from biological data’. As VANTED, it displays and clusters large graphs in two- and three-dimensional space. Furthermore, it renders graphs in a highly interactive format. However, its focus is the presentation of gene expression data. Thus, in contrast to a statistical approach to identifying genes of interest in a data set where the biological groupings and contrasts of interest need to be defined, the network paradigm applied in BioLayout Express 3D [61] presents the structure in the data irrespective of the question asked. Data will be included in the graph based purely on whether it correlates to other data above the defined threshold. This provides independent insight and prevents cluttering by too many data, but is inferior to VANTEDs capabilities as far as presentation of metabolic data is concerned.

YANA programs

The YANA metabolic software was written to complement EM calculations. Specifically, these programs can directly use output produced by Metatool for its calculations. However, the programs YANA and YANASquare have integrated EM calculation as stand-alone tools.

YANA is a platform-independent (written in Java), ‘dedicated toolbox for metabolic networks with a graphical user interface’ to calculate (integrating Metatool), edit (including support for the SBML format), visualize, centralize and compare EFMs [21]. In particular, proteomics or gene expression data provides a rough indication of various individual enzyme activities, however, in many cases the complete flux distribution in the network is not
known. YANA deals with this task by calculating the strength of flux modes and using available experimental data as constraints (see ‘Theoretical Background’ section, constraint-based modeling). In particular, according to constraints and available data from gene expression, protein expression or enzyme activity data the resulting fluxes are calculated and vice versa. A fast evolutionary algorithm is used for the prediction of EM activities with minimum error. There are alerts for inconsistent experimental data. Furthermore, there are specific algorithms to analyze network topology [see ‘Theoretical Background’ section, step (ii)]: a dissection algorithm, a centralization algorithm and an average diameter routine are available for such topological analyses. Furthermore, these can be used to simplify and analyze complex networks.

To better analyze larger metabolic networks, the Java-based software YANAsquare provides the EM specific visualization of biological networks by, e.g. distinguishing internal and external species with different node styles and colors. Like VANTED, it provides a model set-up and modification with the choice to use the KEGGbrowser routine for rapid retrieval of biological networks, starting from information supplied in KEGG format (including an inhouse database, local data) or directly from KEGG database (or similar remote data) and subsequent editing, visualization and simplification according to user-specific needs. Furthermore, it provides a network overview with a visualization routine and the YANAsquare editor. In particular, different options exist for graphical network layout while the strength of different flux modes according to the specific metabolic situation modeled can be visualized. Furthermore, YANAsquare offers network performance analysis besides calculation of flux modes such as target and robustness tests.

The fit of the experimental data to a metabolic network is also an interesting challenge, for instance Liang et al. [36] compared the genetic algorithm with a steepest descent strategy. The latter is particularly efficient in fitting data to larger metabolic networks (200–300 enzymes; [36]). If the wish is to illustrate flux distributions from EM calculations in the context of other -omics data or to compare them with other flux measurements we suggest to integrate the resulting flux distributions from YANAsquare in VANTED using its FluxMap module. By this, one can use the advantages of the specific software for an integrated visualization approach.

**TOWARD AN OPTIMAL ANALYSIS OF METABOLIC NETWORKS**

Considering typical metabolic modeling challenges (Table 1), how would we best tackle them with the software reviewed?

(i) Find the exact set of enzymes and reactions involved: Metabolic reconstruction requires first a set-up step to obtain an overview of pathways, enzymes and reactions involved. This can be tackled by the YANA suite of programs using its software tool KEGGbrowser [4]. This is easily achieved for the pathway or network of interest, it relies however on data supplied by KEGG or by the user in a KEGG-compatible format. It may prove faster to use the full genome-scale models supplied by iMAT [30] or those available from the COBRA toolbox [35]. Cross-linking and data compatibility is improved by standard formats used by the above software, in particular KEGG database format, SBML and XML format. However in both ready made and self constructed models, enzymes may be overlooked. There is lower risk for a complete miss in well curated models provided by iMAT and the COBRA toolbox, but the multiple roles of enzymes in different pathways, flexible use of substrates, the recognition of transporter specificities and presence present a challenge for any metabolic reconstruction.

(ii) Network structure: VANTED [33] provides structural analysis of metabolic networks containing experimental data and biological networks including the high quality visualization of complex data sets and graphs, disparate data and annotation of fluxes. In contrast, YANAsquare allows the rapid visualization of central enzymes (hubs) including the emanating different flux strengths and metabolites as well as the creation of metabolic maps (automatic setup as well as user-directed analysis and layout). iMAT allows structural analyses on the prepared genome-scale metabolic networks in an intuitive way including central pathways and intermediary metabolism, the latter being a particular bonus. The COBRA toolbox offers customized visualization of metabolic maps to allow structure analysis. This includes futile cycles, reversible and irreversible reactions. Thus these tools are helpful for step (ii), besides the already mentioned challenges this also includes identification
of sources and drains in the system. However, committed steps become only apparent by detailed analysis with any of the above tools including biochemical knowledge, which is also critical to account for spontaneous reactions.

(iii) Visualization of flux balanced pathways: BioOpt [34] and Metatool [15] are simple tools to efficiently perform FBA up to middle-sized networks. Metatool includes in higher versions also the calculation of a convex basis vector. For more sophisticated analysis, the COBRA toolbox [35] can even be used for metabolic behavior analysis and is particularly useful to satisfy and analyze complex constraint conditions such as growth equations. iMAT [30] offers rapid metabolic modeling based on constraints from gene expression data or protein expression data on the large basis of custom-made metabolic models for different organisms. After establishing the network or subnetwork of interest [maybe exploiting the KEGG browser for model set-up, see step (i)], rapid CBM based on different data sets is also achieved and visualized by YANAsquare [4]. However, for large-scale networks and CBM incorporating large data sets YANAvergence shows better performance [36] and for analysis of metabolite fluxes in high-complexity data sets the COBRA toolbox [35] is particularly powerful. The different YANA programs allow metabolic flux analysis of genome-scale networks and the incorporation of experimental data sets as constraints. This includes the prediction of metabolite concentrations if suitable data (e.g. measurements of external source and drain metabolites) are available. CBM is well implemented in the above three recommended software, however, with YANA programs as well as the above mentioned COBRA toolbox and iMAT accurate predictions of metabolite concentrations remain a challenge (hence step iv is important).

(iv) For processing of direct experimental data on metabolic fluxes such as isotopologue data, we recommend FiatFlux [37]. FiatFlux is best used, e.g. for flux ratio analysis. As a further tool, C13 can be used for metabolic flux analysis by stationary carbon isotope labeling on fractional enrichment data [38]. ReMatch [22] can be used for metabolic network model construction, storing, sharing and integrating carbon mappings for $^{13}$C-metabolic flux analysis over the web. These programs also allow the mapping of carbon atoms including manually editing and identification of metabolites and mass isotopomers in a user-friendly way. SBML model repositories are found for FiatFlux. Nevertheless, 4.1–4.7 in Table 1 stress that there is clearly space for further improvement.

(v) Detailed understanding of metabolic regulation becomes possible after performing the analysis steps (i) to (iv) with the above tools and visualizing the different data sets (e.g. with VANTED [33]). The comparison of gene expression data with calculated and experimentally verified metabolite fluxes allows the identification of regulated enzymes applying YANAvengersence [36]). On the other hand, this qualitative or even semi-quantitative understanding of regulation including identification of futile cycles (see recommended software above) and key regulatory enzymes (e.g. pace-maker enzymes) is, however, limited compared to a comprehensive quantitative analysis by a full kinetic description of the system, e.g. using a set of differential equations describing all enzyme kinetics in full detail. Furthermore, integration of disparate data types and best quality visualization remains a challenge for large-scale networks analyzed in this manner.

Integration of the results from the different steps, as calculated by different software tools is easily achieved for all tools within one toolbox, for instance regarding the COBRA toolbox or the different YANA programs. The latter are compatible among each other as well as with Metatool; Metatool is compatible across all its versions. However, more important are the use of well known standard formats such as SBML or XML, so that output generated from the software can be used by other software for further analysis. This applies in particular for YANA, Metatool (higher versions) and COBRA toolbox. Furthermore, all models available from iMAT are available either in TIGER or COBRA model format.

Scripting languages such as Taverna or Perl (in particular BioPerl) help further, if large data files have to be processed or even broader compatibility between all softwares is sought. All software discussed here (Metatool, ReMatch, FiatFlux, C13, BioOpt, COBRA toolbox, efmtool, VANTED,
iMAT and YANA programs) provide at least some sort of text format with tab delimitation. Such scripts are helpful for more complex task in processing output generated from the software (the COBRA toolbox has actually several powerful scripts already integrated for exactly such purposes). For example, establishing lists of re-occurring, highly used enzymes in the list of modes generated to identify critical enzymes involved in survival of *Listeria* under macrophage attack [49].

Apart from their specific limitations, the discussed software is helpful in many ways for metabolic flux analysis and visualization (Table 3). The advantages and limitations given for the software selection apply with modifications also for other software in the field (see other references) and highlight general strengths and weaknesses in current capabilities. Limitations change as new versions of software are continuously developed in this rapidly growing field. Furthermore, we want to stress that compatible metabolic modeling software can be advantageously combined to provide a more comprehensive perspective (e.g. different routines from the COBRA toolbox; Metatool and YANA programs; VANTED with add-on Flux Map). At each of the analysis steps (see ‘Theoretical Background’ section), visualization is advantageous. Some software such as VANTED and BioLayout Express 3D stress this aspect and offer attractive visualization options. However, most of the other modeling tools offer metabolic maps of some sort and then there is always the option to combine several programs to obtain an impressive overall workflow including good visualization options (e.g. COBRA toolbox, YANA programs).

However, there are also general limitations apparent from our comparison, e.g. the modeling of large-scale (‘genome-sized’) networks. These may be tackled using clever sampling procedures [17, 18] that are useful in predicting the outcomes of lethal knockout analysis. However, in nature there may occur fairly complex scenarios where the accuracy of these methods goes down (e.g. double knockout, partial enzyme inactivation, etc.). The efmtool with its rapid calculation of flux modes as well as several novel concepts on flux mode calculation obtain good results in large networks involving millions of EFMs (e.g. 26 million EFM in an *E. coli* model [20]).

To tackle genome-scale networks, powerful algorithms have been proposed for specific EFM set computations [62] as well as algorithms to calculate a minimal set of EFMs, the generating flux modes, for such genome-scale networks [63]. Furthermore, constraint-based approaches exist which also allow the analysis of gene knockouts and their effects without computing the complete set of EFMs [64].

Finally, the fifth step, a detailed modeling of regulatory interactions in metabolism, remains a challenge. This includes further complexities such as protein stability and RNA turnover [36]. A key challenge remains the detailed kinetics of biochemical reactions. However, it is only possible to model small enzyme networks in this detail since so many parameters are unknown and the calculations become very complex. There have been continuous efforts to master these challenges [25] and thus steady progress to meet these challenges better has been achieved in the last 10 years.

**CONCLUSION**

Metabolic flux analysis is complex and requires dedicated software. It is developing rapidly, offering a multitude of different solutions to the user. Due to limitations of space, we restrict detailed comparisons to a choice, which covers a broad range of the individual analysis steps in metabolic modeling. We opted for user-friendly programs with broad applicability and compare these to a background of alternatives. These have specific advantages and limitations as outlined. Typical general challenges remain such as modeling of complex networks and visualization of rich data sets. However, as the field is moving fast, the capabilities of metabolic modeling software are steadily improving.

**SUPPLEMENTARY DATA**

Supplementary data are available online at [http://bib.oxfordjournals.org/](http://bib.oxfordjournals.org/).

**Key Points**

- Metabolic data are complex and become available in increasing quantity.
- Software is required to process these data and to efficiently extract the maximal amount of information from each analysis step.
- Current challenges include the modeling and processing of genome-scale networks as well as visualization of the complex data.

**Acknowledgements**

We thank Dr S. Gorski and Dr U. Rapp-Galmiche for stylistic corrections and comments on the manuscript.
FUNDING

German Research Foundation DFG [grant Da 208/10-2 (modeling intracellular pathogens) and TR34/Z1 (Staphylococcus aureus modeling)].

References


