FLUXANALYZER: A GRAPHICAL INTERFACE FOR STOICHIOMETRIC AND QUANTITATIVE ANALYSIS OF METABOLIC NETWORKS

S. Klamt* A. Kremling* E.D. Gilles*

* Max-Planck-Institute for Dynamics of Complex Technical
Systems, Leipziger Strasse 44, 39120 Magdeburg
klamt@mpi-magdeburg.mpg.de

Abstract: Metabolic Flux Analysis has turned out to be a powerful approach for studying and engineering metabolic networks. Nevertheless, there seems to be a lack of a comprehensive software platform where all computational routines for metabolite balancing and stoichiometric analysis can be used in a convenient way. Therefore, the *FluxAnalyzer*, a programming package for MATLAB, was developed. The interactive software allows a menu-driven use of various procedures within a graphical user interface, where the metabolic networks under investigation and the results of calculations are visualized in flux maps. Arbitrary networks can be defined symbolically. The *FluxAnalyzer* is used in current projects at our institute and its application for flux analysis of cell cultures in an industrial process is planned. *Copyright © 2001 IFAC*

Keywords: metabolic flux analysis, metabolite balancing, stoichiometric networks, stoichiometric analysis, metabolic engineering, flux maps, graphical user interface

1. INTRODUCTION

Metabolic Flux Analysis (MFA) has become a widely used approach for studying the properties and capabilities of metabolic networks in microorganisms (Stephanopoulos et al., 1998), (Nielsen, 1998), (Varma et al., 1993). Stoichiometric studies are useful for an identifcation of qualitative properties in biochemical reaction systems, e.g. the detection of metabolic constraints or subnetworks. Furthermore, measurements of uptake and excretion rates of a cell culture in steady state allow the determination of stationary metabolic flux distributions by using metabolite balancing techniques. Additional quantitative information useful for the determination of flux distributions can be obtained by isotopic tracer experiments (Marx et al., 1996), an approach not considered here.

MFA is a tool for studying the cell metabolism in biotechnological processes, where the goal is to optimize the product yield, for example by redirecting fluxes using genetic manipulations (Stephanopoulos et al., 1998), (Nielsen, 1998). Besides, MFA may be a suitable method for obtaining an impression of how an organism balances its metabolic network under certain environmental conditions.

Various mathematical techniques are involved in performing MFA such as algebraic routines, optimization strategies and balancing techniques, which will be briefly summarized in section 2. With the *FluxAnalyzer*, a programming package was developed, which allows a symbolic formulation of networks and which contains a complete set of routines for stoichiometric analysis and metabolite balancing. In contrast to other programs for MFA the *FluxAnalyzer* permits the convenient input of data and the display of the output of calculations in graphical representations of the metabolic network (flux maps). The embedded procedures can be selected from a menu.

Therefore, the *FluxAnalyzer* allows users which are not completely familiar with the mathematical background of the MFA to perform flux analysis. The features of the *FluxAnalyzer* and the integration of diverse functions in the application are described in detail in section 3.

2. PRINCIPLES OF STOICHIOMETRIC AND METABOLIC FLUX ANALYSIS

MFA uses steady state assumptions for analyzing metabolic networks. Thus dynamic studies of the cell metabolism are not possible. On the other hand, it does not require kinetic parameters of enzymes, which are often not available and/or have in vivo/in vitro discrepancies. In the following the principles of MFA and the derivations of linear constraints are briefly described. A detailed discussion can be found elsewhere, e.g. (Stephanopoulos *et al.*, 1998), (Heijden *et al.*, 1994).

The first step in MFA is the construction of the metabolic network containing the most important metabolites and reactions between them :

- m metabolites (nodes) with concentrations $c = (c_1 \ldots c_m)$
- q reactions with net rates $r = (r_1 \ldots r_q)$ in $[mmol/(g$ dry weight $\cdot h)]$; this includes a cumulative biomass synthesis with growth rate μ
- $T = \{t_{ij}\}_{i=1 \ldots m, j=1 \ldots q}$: stoichiometric matrix; t_{ij}: stoichiometric coefficient of metabolite i in reaction j

The dynamics of the concentration c_i of metabolite i is described given by

$$\frac{dc_i(t)}{dt} = \sum_{j=1}^{q} t_{ij} r_j(t) - \mu(t) c_i \qquad (1)$$

The reaction rates $r_j(t)$ depend on the kinetic properties of the participating enzymes, on the concentrations of metabolites and on environmental parameters such as temperature or pH. The enzymes are generally subject to complex regulations by inhibition and activation. The assumption of steady state ($\mu(t)$, $c(t)$ and $r(t)$ are constant) and the neglection of the dilution term due to usually low concentrations of intermediates leads to a homogeneous system of m linear equations, without any time-varying or enzymatic parameters:

$$0 = \frac{dc_i(t)}{dt} = \sum_{j=1}^{q} t_{ij} r_j \qquad (2)$$

This can be written in matrix notation:

$$0 = Tr \qquad (3)$$

Thus, the problem of determining reaction rates is reduced to solving a system of linear equations.

All linearly independent solutions of eq. (3) are contained in the kernel matrix K of T:

$$0 = TK \qquad (4)$$

The null space can be investigated without any further informations. For example, elementary flux modes may be computed to decompose the network into minimal parts which fulfil the steady state condition (3) and represent essential structural features of the biochemical network (Schuster and Hilgetag, 1994).

As the number of reactions (q) is usually greater than the number of metabolites (m), the system is highly underdetermined (with $q - m$ degrees of freedom). Measured rates or physiological assumptions can be used to reduce the possible solution space (which is a linear combination of vectors contained in the Matrix K). For this purpose the vector of reaction rates r is partitioned in the known part, r_b, and in the part of unknown fluxes, r_n. In the same way matrix T is partitioned in T_b and T_n:

- $r = (r_b, r_n)$
- $r_b = (r_{b_1} \ldots r_{b_k})$: k known/measured rates
- $r_n = (r_{n_1} \ldots r_{n_u})$: u unknown rates
- $T = (T_b, T_n)$; $T_b : m \times k$; $T_n : m \times u$

Equation (3) can now be formulated as

$$0 = Tr = T_b r_b + T_n r_n$$

$$\boxed{T_n r_n = -T_b r_b} \qquad (5)$$

Eq. (5) is the fundamental relation in MFA. The aim is the determination of preferably all unknown rates in r_n and the use of possible redundant information for detection of measurement or modelling errors. Whether and to what extent one may benefit from using the information of r_b depends on its composition. Rates of r_n may be calculable or non calculable, rates of r_b may contain redundant information or not. Analysis of the matrices T_n and T_b is necessary for these classifications. The complete system in eq. (5) can be classified by two properties:

- **Determinacy**
  - *underdetermined*: $rank(T_n) < u$ (not all rates are calculable)
  - *determined*: $rank(T_n) = u$ (all rates are calculable)
- **Redundancy**
  - *redundant*: $rank(T_n) < m$ (some rates are redundant)
  - *not redundant*: $rank(T_n) = m$ (no rate contains redundant information)

A special case in this context arises if the system is *exactly (fully) determined*, that is, it is determined and not redundant and thus can be

solved directly via the inverse of T_n (because the latter one exists). Unfortunately, this is a very rare case. Otherwise the Penrose pseudo inverse must be used for calculating the calculable rates and balancing (i.e. adapting the balanceable rates to obtain a consistent system) the balanceable rates, respectively. During balancing of redundant systems a consistency check may be performed. Optimization strategies (e.g. in the sense of optimal production capabilities) are a special method, where the degrees of freedom are reduced by optimizing an objective function (like 'Maximize product yield!').

3. *FLUXANALYZER*: FEATURES AND REALISATION

The software tool *FluxAnalyzer* provides an easy and convenient way for performing Metabolic Flux Analysis in arbitrary biochemical reaction systems. The package was developed using the programming language of MATLAB, where several required computational routines, particularly from linear algebra, are available. The various procedures associated with MFA are embedded in an interactive graphical user interface where calculated flux distributions or other computed results are visualized. The most important features of the *FluxAnalyzer* are described in the following.

Symbolic definition of the metabolic network

All metabolites are defined by a symbol, all reactions are defined using the symbols of the defined metabolites and the stoichiometries of the reactions. An example is given for the oxidation of malate (note, that only the balanced metabolites are contained in the symbolic equations):

$$Mal::OxA \qquad Mal = OxA + NADH$$
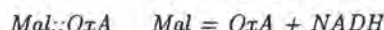
The symbolic formulation of the metabolic network is used for the internal generation of the stoichiometric matrix. Besides, for every declared reaction some parameters can be defined, such as reversibility or variances of measurements.

Network scenarios

Several network scenarios (with different stoichiometries) of one or more organisms can be managed by the network manager (Fig. 1).

Flux maps

The flux maps are the center of the *FluxAnalyzer*. They contain textboxes which allow the input of values (known reaction rates or parameters of reactions, biomass composition) as well as the output of calculations directly on the associated biochemical pathways in the visualized network. An arbitrary number of flux maps may be created
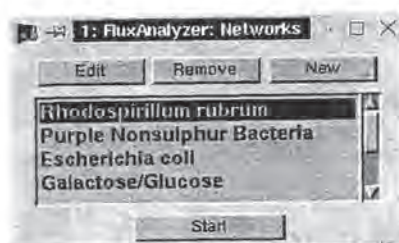

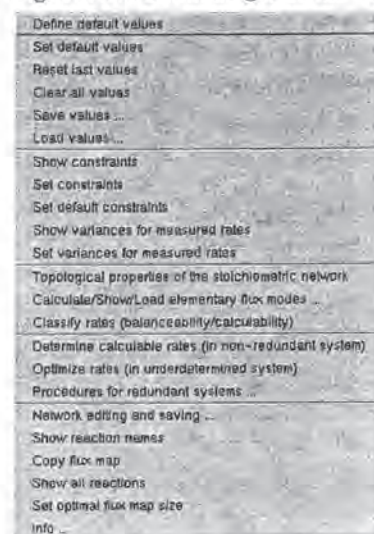Fig. 1. Network manager of the *FluxAnalyzer*

Fig. 2. Pull-down menu of the *FluxAnalyzer*

for every network scenario. The user is free to design the flux maps in his own way by using arbitrary graphic programs (e.g. xfig) or other sources of graphical network representations. An automatic creation of flux maps directly in the *FluxAnalyzer* by using the symbolic formulation of the network was intentionally avoided, because in most cases this would not lead to representations desired by the user. Besides, comments may be added in the network by this way. After the flux maps are created, the user can easily place editable or non-editable textboxes to the associated pathways in the flux maps. These preparatory works have to be done only once for every network. The flux maps and textboxes are then imported every time the network scenario is started in the network manager. In the example (Fig. 3 and 4) two flux maps exist: one for the catabolic and one for the anabolic network. In the anabolic network the biomass composition may be entered which is used to calculate the effluxes of catabolites in the biomass synthesis. Note that every rate in the anabolic network depends here only on the growth rate and the biomass composition. For this reason the associated textboxes were declared as non-editable by the user.

Menu items

All actions (described below) can easily be performed via menu-items from a pull-down menu

(Fig. 2) provided in the first flux map (item 'MFA' in Fig. 3). The user has not to be completely aware of mathematical details of the used algorithms.

Topology analysis

Useful information about topological properties may be determined: the degrees of freedom, a priori redundancies, true branchpoint metabolites (participating at least in 3 reactions), non-true branchpoint and dead-end metabolites.

Elementary flux modes

Elementary flux modes reflect special topological features of metabolic networks and are useful for several stoichiometric studies, especially for finding maximal product yield. The *FluxAnalyzer* allows the calculation of all elementary modes (with or without the reversibility condition). They can be displayed in the flux maps or saved/loaded.

Classification of rates

Before flux distributions are calculated for a defined set of known (and entered) rates, the classification of all rates in the categories calculable/non-calculable (unknown rates) or balanceable/non-balanceable (known rates, cf. section 2), respectively, may be determined and displayed by distinct colors.

Calculation routines for all network classes

Procedures necessary to handle redundant/not redundant and determined/not determined systems are available. Thus, all calculable rates can be determined and all balanceble rates can be balanced in a given scenario. For balancing of rates with redundant information one may use (and enter) known variances of measurements of reaction rates. Several methods (least squares, weighted balancing, gross measurement error detection) are implemented for redundant systems.

Optimization routine

Constraints and coefficients of an objective function for an arbitrary optimization problem can be defined in the flux maps. Then the optimization can be started from the menu and the results are presented in the flux maps.

Linear functions

In a defined scenario the linear dependency of an unknown rate from a known one can be presented as a graph or in form of an equation.

Documentation

Saving/loading of data sets of a scenario (including reaction rates and biomass composition) and saving/printing of flux maps is possible.

4. APPLICATIONS AND DISCUSSION

At present the *FluxAnalyzer* is used for studying the capabilities and metabolic constraints of the metabolism in anoxygenic photosynthetic bacteria (see flux maps in Fig. 3 and 4). For determination of flux distributions under distinct growth conditions, continuous fermentations of these bacteria are performed in laboratory scale bioreactors, in order to measure the uptake and excretion rates of steady state cultures.

Recently the BAYER AG in Leverkusen, Germany, decided to use the *FluxAnalyzer* for investigations of metabolic fluxes in cell cultures of an industrial process.

To our opinion the *FluxAnalyzer* contains all functions and procedures necessary for stoichiometric analysis and determination of flux distributions (by metabolite balancing) in biochemical reaction networks. They are arranged in a convenient graphical user interface and this allows an efficient use of these methods. The provision of opportunities for balancing isotopomers or for performing metabolic control analysis could be possible extensions.

5. REFERENCES

Heijden, R.T.J.M., B. Romein, J.J. Heijnen, C. Hellinga and K.Ch.A.M. Luyben (1994). Linear constraint relations in biochemical reaction systems: I. classification of the calculability and the balanceability of conversion rates. *Biotechnology and Bioengineering* **43**, 3–10.

Marx, A., A.A. Graaf, W. Wiechert, L. Eggeling and H. Sahm (1996). Determination of the fluxes in the central metabolism of *Corynbacterium glutamaticum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. *Biotechnology and Bioengineering* **49**, 111–129.

Nielsen, J. (1998). Metabolic engineering: Techniques for analysis of targets for genetic mainipulations. *Biotechnology and Bioengineering* **58(2-3)**, 125–132.

Schuster, S. and C. Hilgetag (1994). On elementary flux modes in biochemical reaction systems in steady state. *Journal of Biological Systems* **2**, 165–182.

Stephanopoulos, G.N., A.A. Aristidou and J. Nielsen (1998). *Metabolic Engineering. Principles and Methodologies.*. Academic Press. San Diego.

Varma, A., B.W. Boesch and B.O. Palsson (1993). Biochemical production capabilities of *Escherichia coli*. *Biotechnology and Bioengineering* **42**, 59–73.
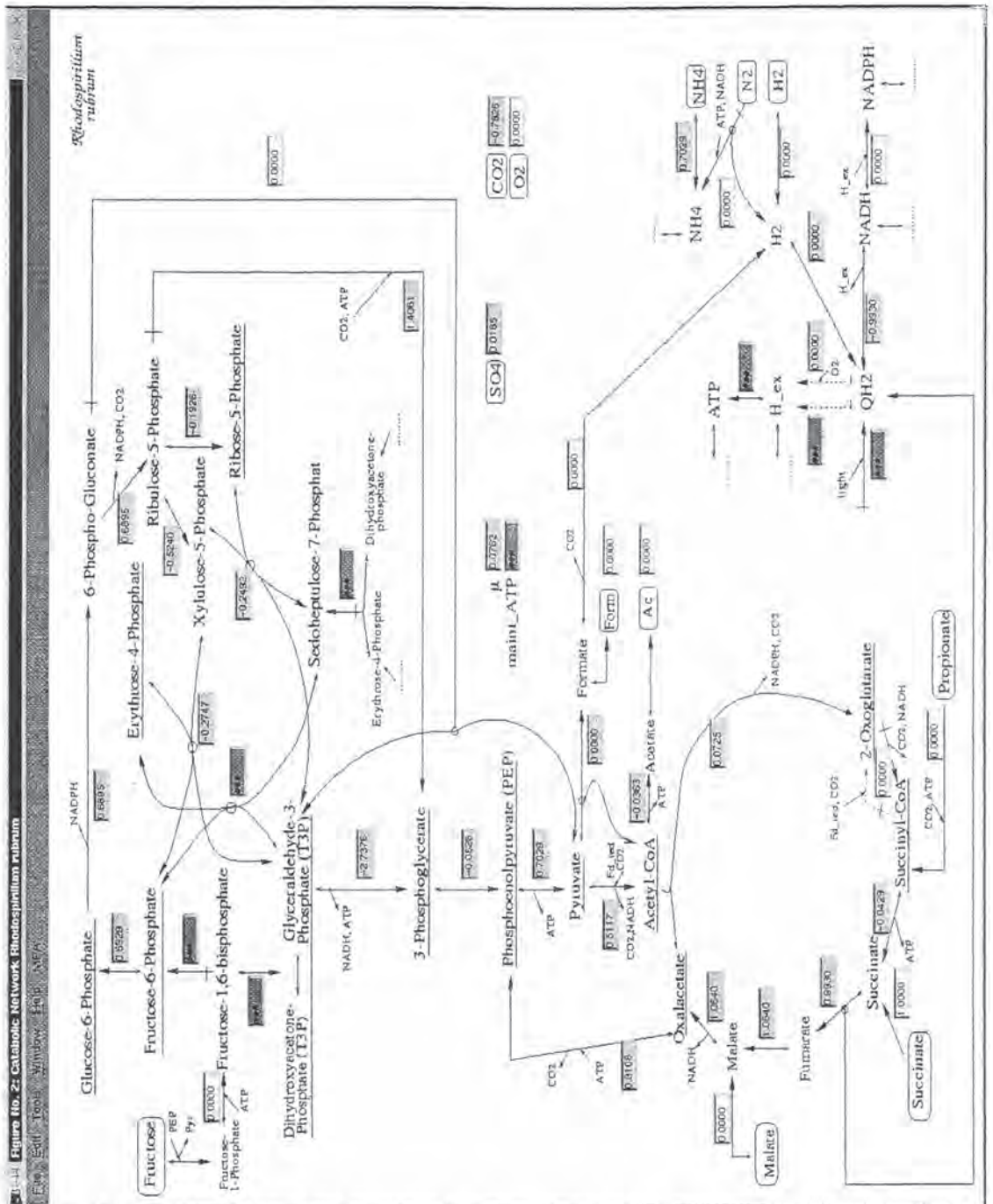
Fig. 3. Catabolic flux map of the photosynthetic bacteria *Rhodospirillum rubrum* used by the *FluxAnalyzer* (hardcopy). Every (partially cumulative) reaction has an associated editable textbox, where values (e.g. measured reaction rates) can be entered. Then calculations can be performed via the pulldown-menu 'MFA'. Results of calculations are presented as follows: Entered values have a green (here: white), calculated rates a blue (here: gray) and non-calculable rates a black (here: dark gray) background. Non-calculable rates are indicated by the value ###. No balanceable rates (would have a red background) occur in this scenario because the network is not redundant. The values of the reaction rates are related to the direction indicated by the large arrow head. Underlined metabolites illustrate precursors.
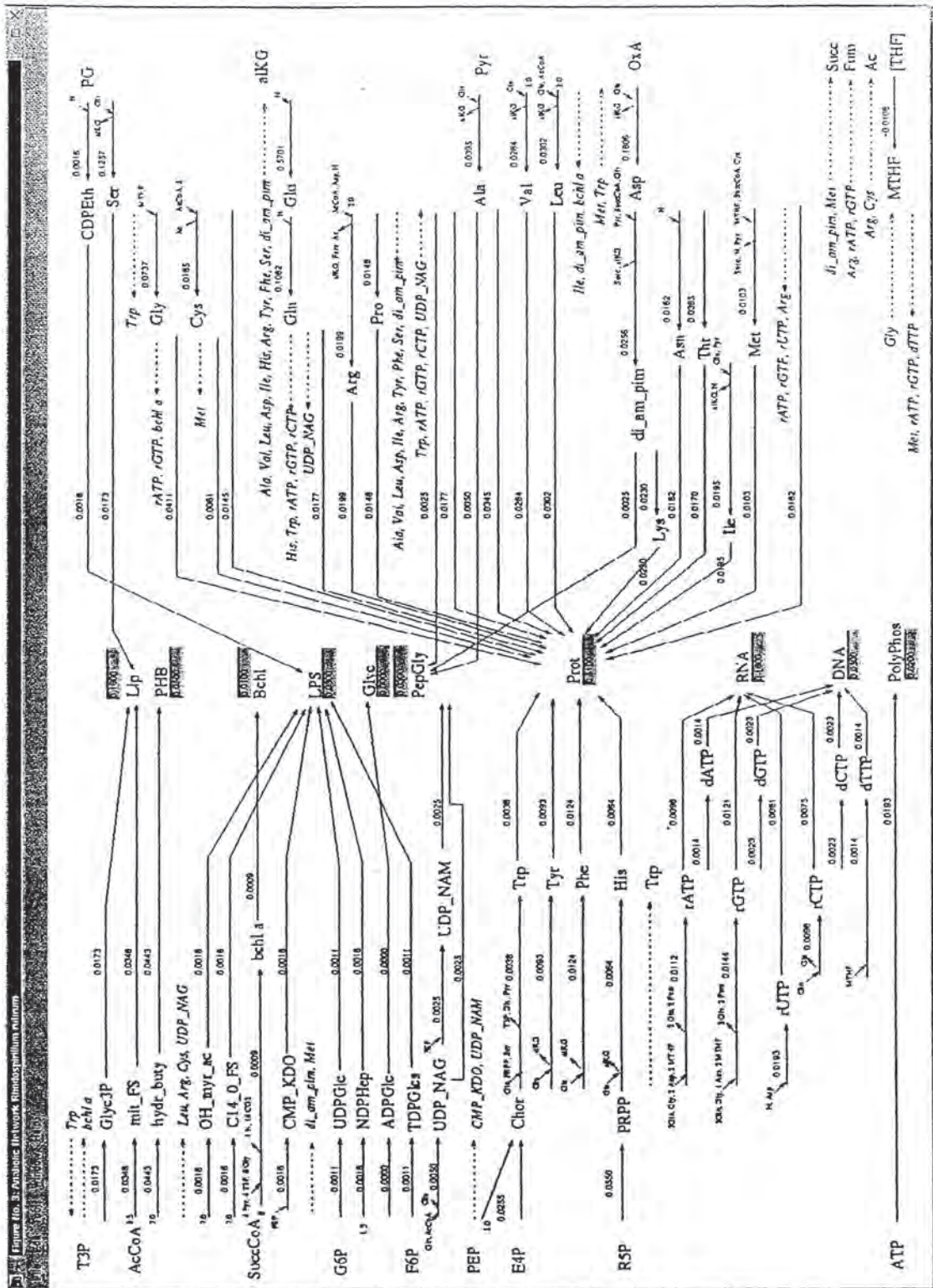
Fig. 4. Anabolic flux map of *R. rubrum* used by the *FluxAnalyzer* (hardcopy). In the center of the flux map the biomass composition (in [g/g dry weight]) can be entered. Starting points of biomass synthesis are precursors produced in the catabolic network. All anabolic reaction rates in this map depend only on the given biomass composition and on the given or calculated growth rate. Therefore textboxes of anabolic rates were configured as non-editable and thus anabolic rates may only be calculated and then presented at their corresponding arrow in the flux map.