A Three-Level Problem-Centric Strategy for Selecting NMR Precusors & Analytes.

Soumitra Ghosh[1], Ignacio E. Grossmann[1], Mohammad M. Ataai[2], and Michael M. Domach[1*]

[1] Department of Chemical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

[2] Department of Chemical & Petroleum Engineering
University of Pittsburgh
Pittsburgh, PA 15213

*Person to whom correspondence should be sent; md0q@andrew.cmu.edu

ABSTRACT

We have developed a sequential set of computational screens that may prove useful for evaluating analyte sets for their ability to accurately report on metabolic fluxes. The methodology is problem-centric in that the screens are used in the context of a particular metabolic engineering problem. That is, flux bounds and alternative flux routings are first identified for a particular problem, and then the information is used to inform the design of NMR experiments. After obtaining the flux bounds via MILP, analytes are first screened for whether the predicted NMR spectra associated with various analytes can differentiate between different extreme point (or linear combinations of extreme point) flux solutions. The second screen entails determining whether the analytes provide unique flux values or multiple flux solutions. Finally, the economics associated with using different analytes is considered in order to further refine the analyte selection process in terms of an overall utility index, where the index summarizes the cost-benefit attributes by quantifying benefit (contrast power) per cost (e.g. NMR instrument time required). We also demonstrate the use of an alternative strategy, the Analytical Hierarchy Process, for ranking analytes based on the individual experimentalist's-generated weights assigned for the relative value of flux scenario contrast, unique inversion of NMR data to fluxes, etc.

INTRODUCTION

^{13}C nuclear magnetic resonance (NMR) is amongst the tools employed in metabolic engineering for metabolic flux identification. Such a method is employed because extracellular measurements such as glucose uptake are insufficient to fully determine all the metabolic reaction rates. To obtain flux information, the NMR-determined labeling pattern of specific metabolites derived from a labeled precursor (*e.g.* 1-^{13}C glucose) in conjunction with the extracellular measurements are "inverted" in order to uncover the intracellular reaction rates.

A related, but more focused application is using spectroscopic data for scenario discrimination. Here, mathematical analysis has led to a particular mutation strategy that will achieve a desirable flux distribution. For example, linear programming has indicated that attenuating the activity of pyruvate kinase will diminish acetate formation from glucose (Phalakornkule *et al*, 2000; Lee *et al,* 2000). However, two different flux distributions can reconcile the mutational block and acid-reduction outcome. The NMR spectra associated with different analytes that can help to discriminate between fluxes in subnetworks can be predicted. The extent to which the observed and predicted spectra concur will then either validate or refute the hypothesis that a particular flux distribution is actually used by the cells.

Several problems can arise that confound the use of NMR data for flux identification. When inverting NMR data to fluxes using isotopomer mapping matrices (Schmidt *et al*, 1997 a, b), a non-linear optimization problem can result that is highly non-convex due to the bi-linear and tri-linear terms present in the isotopomer balance equations. Consequently, local versus global solutions can be found, especially when conventional gradient-based optimizers are used. In an attempt to overcome the problem of local solutions, we have recently described the use of a two-step strategy (Ghosh *et al,* 2004). First, a Mixed Integer Linear Programming (MILP) formulation is used to provide lower and upper bounds on individual fluxes for the NMR data-to-fluxes problem. Apart from their utility for analyzing NMR data, MILP solutions are useful to obtain because they provide the flux bounds for the trafficking alternatives that equally fulfill an objective or a set of extracellular measurements. Thus, such solutions not only indicate potentially different targets for metabolic engineering, the tightened flux bounds enable the inverse problem for spectra-to-fluxes to be solved by using a deterministic nonlinear global

optimization algorithm of the branch-and-bound type (BARON, Sahinidis *et al,* 2002, Ghosh *et al,* 2004).

Another related problem is that more than one flux distribution might lead to an analyte exhibiting a similar NMR spectrum; hence, the data-to-fluxes problem could, in turn, generate alternative solutions that are not strictly local solutions. Intuitively, one would expect that as more data is accumulated and used (*e.g.* spectra from more analytes) the less likely that alternate solutions would arise. Overall, when a hydrolyzed protein or cell extract sample is used, a finite set of metabolites or amino acids are available that can provide unique flux solutions, and it is desirable to identify those analytes in advance.

In an attempt to address the above problems, we have developed a sequential set of computational screens that may prove useful for evaluating analyte sets for their ability to accurately report on metabolic fluxes. The methodology is problem-centric in that the screens are used in the context of a particular metabolic engineering problem. That is, flux bounds and alternative flux routings are first identified using METABOLOGICA (Zhu et al, 2003) for a particular problem, and then the information is used to inform the design of NMR experiments.

The general outline of the screens is as follows. After obtaining the flux bounds via MILP, analytes are first screened for whether the predicted NMR spectra associated with various analytes can differentiate between different extreme point (or linear combinations of extreme point) flux solutions. The second screen entails determining whether the analytes provide unique flux values or multiple flux solutions. The screening method employed is mathematically different than MILP, but the outcome is similar. Our MILP formulation enumerates all feasible extreme point flux solutions that meet constraints and an objective (Phalakornkule *et al*, 2001). The computation performed for analyte selection similarly enumerates all flux distributions that can yield within a tolerance a particular NMR spectrum from a metabolite (or a set of metabolites). Finally, the economics associated with using different analytes is considered in order to further refine the analyte selection process.

In this paper, a "test" metabolic network will be first described that will enable the illustration of how flux scenarios are enumerated, and how NMR calculations are performed. The

aforementioned screens will then be illustrated in the context of the "test" network where the attributes of different analyte sets are calculated. The computational screening strategies will also be applied to the more practical case of a larger *E. coli* network where pyruvate kinase activity has been deleted to inhibit acetate by-product formation. Economics will be addressed in two ways. For the "test" problem, the intensity of analyte sample preparation will be qualitatively weighed against NMR attributes. For the larger problem, abundance data will be used to in conjunction with screening results to infer how much NMR resource time and cost would be needed to acquire a usable spectrum. Finally, to illustrate an alternate means of using the quantitative results from the screens, the Analytic Hierarchy Process (AHP; Saaty 2000) is applied to the larger problem in order to rank order potential NMR analytes.

ABSTRACTED TEST NETWORK: FLUX BOUNDS, NMR CALCULATIONS, & SCREENS

The "test" network is shown in Figure 1a, where r represents a net flux (*e.g.* mmol/g cell h). The model is based on the abstracted representation of central metabolism described by Forbes *et al.* (2001). We assume, as posed by the originators, that the extracellular fluxes r_1, r_2, and r_9 are known (*i.e.* directly measured). The values are $\{r_1, r_2, r_9\} = \{1, 0.3, 0.4\}$.

Determination of Flux Bounds

To find the flux bounds, the network in Figure 1a can be formulated as a linear programming problem given by Equation (1) by considering the flux balances around each node in the network, and the three additional constraints imposed by measurements of extracellular fluxes. As noted earlier, such bounds greatly enable the operation of global optimization methods such as BARON when used to invert NMR data to fluxes. A dummy objective of $Z = 0$ is provided so that any feasible solution to the linear program is accepted.

$$Min \;\; Z = 0 \qquad (1a)$$

Subject to

$$r_1 + r_2 = r_3 + r_4 \qquad (1b)$$

$$r_4 + r_7 = r_5 \qquad (1c)$$

$$r_5 = r_7 + r_6 \qquad (1d)$$

$$r_6 + r_8 = r_5 \qquad (1e)$$

$$r_9 = r_8 + r_{10} \qquad (1f)$$

$$r_1 = 1 \qquad (1g)$$

$$r_2 = 0.3 \qquad (1h)$$

$$r_9 = 0.4 \qquad (1i)$$

$$r_j \geq 0 \qquad \text{for all } j = 1,2, \ldots, 10 \qquad (1j)$$

One characteristic of metabolic networks is that multiple feasible flux scenarios can arise. Multiplicity arises when the number of experimental observations is less than the number of unknowns. Another source is redundancy, which can manifest when, for example, multiple reactions can contribute to providing a set amount of ATP or NADPH. In order to resolve the existence of multiple flux distributions, a recursive MILP algorithm (Phalakornkule *et al.,* 2001) or a Depth-First-Search (DFS) method (Zhu *et al.,* 2003) can be employed, and either method

yields the same four flux distributions shown in Table 1. These distributions correspond to the extreme points of the feasible region for the metabolic flux solutions.

NMR Data-to-Fluxes Calculation Method

The actual fluxes within the "test" network can be a linear combination of the four solutions shown in Table 1. To determine the actual intracellular flux values, a ^{13}C-labeled precursor is added to label the metabolites within the network. The relative abundance of the 2^n isotopomers for a n-carbon analyte reflects which reaction paths and the relative rates that yielded the isotopomers from the labeled precursor.

To briefly describe the NMR experimentation and calculations in this context, we assume that 1-^{13}C **G** is fed to the system, where **G** denotes a carbon source such as glucose. A sample of labeled intracellular metabolites is then analyzed after attaining isotopic steady state. By integrating the group of NMR peaks, the Isotopomer Distribution Vectors (IDV) of various metabolites such as **A**, **E**, **C,** etc. can be obtained (Schmidt *et al,* 1997a,b).

Obtaining the fluxes from the IDV data is achieved by solving the *Inverse Problem*, which is stated in Equation 2. The objective function chosen (Z; Eqn. 2a) is the Euclidean norm for the error (*i.e.*, a least-squares criterion). The constraints are comprised primarily of the Isotopomer Mapping Matrix (IMM) equations constructed for each intracellular metabolite. There are 10 net fluxes, $[r_1, r_2, ..., r_{10}]$ in the system, and as noted earlier, the reaction O ↔ E is reversible. Equations (2b-2f) are the isotopomer balances. Also, each of the net fluxes is a convex combination of the four MILP/DFS solutions as required by Equations (2g)-(2i).

To account for the reversible reaction (O ↔ E), an exchange coefficient variable, $\varepsilon_8 = (v_{11}/v_8)$, could be introduced. As an alternative, the reaction rates are split into 10 forward fluxes and one backward flux, $[v_1, v_2, ..., v_{11}]$, and ε_8 is later extracted from the calculations. By adopting this formulation, further bi-linearity can be prevented from being introduced into the problem. In order to account for the fact that $0 \le \varepsilon_8 \le 1$, the inequality constraint, Equation (2j), is provided. This constraint is based on noting that $\varepsilon_8 = (v_{11}/v_8)$, and $v_8 \ge v_{11}$ follows from $r_8 \ge 0$ (Eqn. 1a) and $r_8 = v_8 - v_{11}$. Finally, Equations (2k) and (2l) provide the appropriate bounds to the flux variables to aid the branch-and-reduce algorithm for global optimization.

$$\min \quad Z = \sum_{i=1}^{2} (IDV_A - IDV_A^{\exp t})^2 + \sum_{i=1}^{4} (IDV_E - IDV_E^{\exp t})^2 \qquad (2a)$$

$$s.t. \quad (v_4 + v_3) IDV_F = v_2 IMM_{S \to F} IDV_S + v_1 IMM_{G \to F} IDV_G \qquad (2b)$$

$$v_5 IDV_A = v_4 IMM_{F \to A} IDV_F + v_7 IMM_{C \to A} IDV_C \qquad (2c)$$

$$(v_7 + v_6) IDV_C = v_5 IMM_{A \to C} IDV_A \otimes IMM_{O \to C} IDV_O \qquad (2d)$$

$$(v_5 + v_{11}) IDV_O = v_6 IMM_{C \to O} IDV_C + v_8 IMM_{E \to O} IDV_E \qquad (2e)$$

$$(v_8 + v_{10}) IDV_E = v_{11} IMM_{O \to E} IDV_O + v_9 IMM_{Q \to E} IDV_Q \qquad (2f)$$

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1 \qquad (2g)$$

$$v_k = \alpha_1 r_{1k} + \alpha_2 r_{2k} + \alpha_3 r_{3k} + \alpha_4 r_{4k}$$
$$\forall k \in irreversib\ le\ reactions \qquad (2h)$$

$$v_8 - v_{11} = \alpha_1 r_{1_8} + \alpha_2 r_{2_8} + \alpha_3 r_{3_8} + \alpha_4 r_{4_8} \qquad (2i)$$

$$v_8 \geq v_{11} \qquad (2j)$$

$$0 \leq IDV_{M_i} \leq 1 \qquad \forall M \in metabolite\ s, \forall i = 1,2,..., length_{IDV_M} \qquad (2k)$$

$$\min\{r_{1k}, r_{2k}, r_{3k}, r_{4k}\} \leq v_k \leq \max\{r_{1k}, r_{2k}, r_{3k}, r_{4k}\}$$
$$\forall k \in irreversib\ le\ reactions \qquad (2l)$$

Analyte Screen 1: How Do Analyte Spectra Contrast Over Entire Feasible Flux Space?

First, available precursor labeling and analyte options need to be considered. For the "test" problem, the only possible labeling scenario for **G** in Figure 1(a, b) is a 1-^{13}C label, because a label in any other position would be immediately lost in the two products of reaction r_4. An inspection of the network suggests that the analyte candidates are **A, C, O, E,** and **P**. Note that **E** and **P** would have equivalent NMR spectra because they are the same molecule, but only differ in intracellular versus extracellular localization.

One indication of how well the analytes would perform over the entire flux space would be to establish how well they can distinguish between the four extreme point solutions obtained from the MILP formulation (Table 1). We will henceforth refer to the ability to distinguish between flux distributions as "contrast power." To determine contrast power, the IDVs from the four alternatives are first determined. One is chosen as a reference, and the spectrum associated with the reference is determined. Then, the angular differences between the spectra determined from the other three IDVs with the spectrum reference are computed. The average angular difference (in radians) between the different extreme point flux scenarios with the exchange coefficients set at 0.5 are compared in Figure 2. **C** provides the best contrast power over the entire flux space, while **E** is offers the worst. Note that an angle difference is not presented for **A**. **A** possesses

only one atom; hence, the angle difference between the vectorized spectra of **A** will always be zero. Such an analyte omission will generally not arise in less abstracted problems because one carbon analytes are not prevalent (CO_2 is the major one carbon species) and the lack of line splitting structure for small carbon compounds limits the information they provide.

The discrepancy in the analytes' contrast power can be understood by examining Table 1. Note that r_8 is the flux connecting **E** to the rest of the network (Fig. 1a), and $r_8 = 0$ in both Solutions II and III of the MILP (Table 1). This zero flux detaches the sub-network of **Q**, **E,** and **P** from the main metabolic network; hence, the ^{13}C label will never be able to make it to **E**. Additionally, **E** is infused with unlabeled carbon from **Q** via r_9. **E** will thus tend to be not highly enriched with ^{13}C, which reduces the magnitude of its average difference.

Analyte Screen 2: Do Analytes Invert Differently to Fluxes?

The second screen entails determining if the labeling information possessed by an analyte reverts to a unique flux distribution. Failure to revert to the correct flux distribution would indicate that other analytes need to be included in the set, or some analytes should be omitted based on the allowable flux distribution(s) and available precursor labelings. The varying ability to revert to the correct flux distribution will henceforth be referred to as "resolving power."

For the "test" network and a particular problem, variable resolving power can be demonstrated by first assuming that a mutational strategy is hypothesized to lead to a particular flux distribution $\check{\mathbf{R}}$. An example of a "hypothesized" flux distribution is shown in Table 2; it is a convex combination of the four extreme point solutions shown in Table 1. The experimental design task thus entails selecting one or more analytes that can provide information that confirms or refutes the hypothesis that $\check{\mathbf{R}}$ is indeed the flux distribution operating within the cell.

The hypothesized flux distribution was used to generate each analyte's IDV, and each analyte or set was subjected to the Inverse Problem (Eqn. 2). Table 2 shows the fluxes identified when each analyte or a combination of them was used. These results indicate that **A** and **E** (or even **E** by itself) can generate an accurate flux map, which also is in agreement with Forbes et al (2001). Thus, **E** is a very good choice for an NMR analyte. In contrast, the flux solution provided from **C** is vastly different from the hypothesized solution, which indicates that **C** has low resolving

power. The existence of such a solution suggests that an NMR experiment based on analyzing **C** may not help to confirm or refute the hypothesis because the spectrum of **C** is consistent with $\check{\mathbf{R}}$ or another much different flux solution. The "uneven" performance of analytes indicates the need for a more structured framework that can better screen the metabolites so that accurate flux information is obtained. A more structured version of a screen for resolving power is described next.

From the results of Table 2, analyte **C** can be perceived to provide a "local" solution to the inverse problem. An operational way of viewing this is that there might exist more than one flux distribution (or "clone"), which would all give the same IDV of **C** (within the limits of the prescribed tolerance). Detecting and enumerating such clones potentially provides a means for ascertaining how "tight" the correspondence is between an analyte's spectrum and the flux distribution.

This problem can be considered further with the aid of Figure 3. The function $y = f(x)$ has a distinct local minimum at P. Q is the global minimum here, but the dotted rectangular regions around both Q and R would qualify as global minima within the tolerance shown in Figure 3. This is akin to saying that all these clonal abscissas would define flux distributions that would give the same IDV or NMR spectra, as defined by the value of $y = y_0 \pm \delta$. Thus, to probe for clonal solutions, we modified the global optimization procedure (BARON, Sahinidis *et al,* 2002) so that instead of determining *the* global optima, a set of solutions is obtained that differ from the global solution within a small tolerance.

Traditionally, BARON implements a deterministic global optimization procedure that relies on a spatial branch-and-bound method that is applied to structured problems that involve non-convex functions such as bi-linear, tri-linear, and concave terms. The basic idea of the method consists of successively partitioning the solution space using a tree enumeration search, where the lower and upper bounds of the continuous variables define the space. At each node of this tree, a lower bound on the global optimum of the objective function is computed using valid underestimators of the non-convex functions. An upper bound to the global optimum is computed by solving the original non-linear problem from a starting point that lies in the current partition. Nodes whose lower bound lies below the upper bound are further partitioned, while nodes whose lower bound

exceeds the upper bound are fathomed. The search terminates when the lower and upper bounds of the objective function lie within a small tolerance. The effectiveness of this global optimization search relies largely on the quality of the lower bound for the global optimum, which in turn relies on how tight the lower and upper bounds are for the variables. Although BARON uses a reduction strategy in an attempt to improve the variable bounds, it is very important to provide tight values to reduce the computational cost of the search, which otherwise may be substantial. Thus, having predetermined flux bounds from either a MILP or DFS solution is beneficial.

In order to obtain multiple solutions to the Inverse Problem with the IDV of the metabolite **C** in the objective function, we changed the options file in BARON to obtain the *K*-best solutions of the model. When implemented, the (up to) *K*-best solutions found during every branch-and-bound iteration are maintained, and the nodes provably worse than the K^{th} best solution are deleted (Sahinidis *et al,* 2002). For $K = 3$, the "clonal" solutions listed in Table 3 were obtained. Solution I lies the furthest from the desired flux distribution. The 2^{nd} and the 3^{rd} solutions are reasonably close to the correct fluxes. A similar analysis was performed on **A** and **E,** and it was observed that no such alternative solutions exist within the given tolerance. The optimization was performed on a Pentium IV, 1.8 GHz machine with 512 MB RAM. The network model had 51 variables and 47 constraints, and took ~1 minute to solve with BARON.

Analyte Screen 3: Economic Considerations

After establishing the resolving and contrast powers of candidate analytes, the next screen that is applied is economics. Here, one may be able to determine, for example, what minimal outlay will yield satisfactory results. The economics of analyte selection is driven, in part, by the material and labor costs associated with their isolation. Thus, at the minimum, analytes can be ranked in coarse (low, medium, high) expense categories that are associated with the complexity of their isolation. How much NMR facility time is required is also a comparative economic factor. The time component will be considered in the next case study.

While the test problem is an abstracted view of metabolism, it presents some options for analyte isolation that can be potentially economically distinguished, and then combined with their contrast and resolving power attributes in order to devise different cost-benefit scenarios.

Metabolites **C, E,** and **O** would probably require at least a cytosolic extraction. Alternately, **O** can be viewed to be the carbon skeleton for aspartate; hence, NMR analysis of protein hydrolysis products would provide a source of information on how **O** is labeled. Thus, cell concentration, cell lysis, protein hydrolysis, chromatography, sample clean up (e.g. metal chelation), and concentration (e.g. lyophilization) would all be required.

In contrast, **P** is excreted in the growth medium; hence, a simpler cell-liquid separation (e.g. centrifugation or rapid filtration) followed by cleanup (e.g. metal chelation) and concentration (e.g. lyophilization) would be required. **P** would be especially useful if a defined minimal medium (e.g. glucose plus inorganic salts) was used to grow the cells, and significant consumption of the labeled glucose occurred. Fewer signals would appear in the sample spectrum that originates from naturally-labeled and abundant carbon compounds (e.g. amino acids in nutrient broth) as well as unconsumed labeled glucose. The absence of such signals would lessen the potential for confounding the interpretation of the line splitting pattern of **P**.

E and thus **P** invert reliably to the correct flux distribution. However, **C** and **O** perform significantly better when the contrast power over the entire flux space is considered. Thus, one low expense, but effective strategy that would simultaneously maximize flux resolution and contrast power would be to use **P** (found in cell-free growth medium; high resolving power) and **C** (cytosolic extraction; high contrast power). To envision implementing the strategy, the actual flux distribution would lie between the extreme points. In a sequential approach, the data from **C** in an objective function would drive the solution towards the correct flux distribution or an alternate that fits **C**'s data well. Data from **P** would then be used to determine if the correct solution has been found. Using the data from **C** and **P** simultaneously would then be expected to locate the correct solution in one step, which has been observed (not shown).

LARGER METABOLIC ENGINEERING NETWORK PROBLEM

Network & Flux Solutions to Contrast by NMR

The methodologies described in the prior section have been used to investigate a metabolic engineering problem in *E. coli* that involves a larger metabolic network. The network is shown in Figure 4. Through prior experimentation, it has been proposed that removing the enzyme pyruvate kinase, which catalyzes *PEP → pyruvate*, could eliminate acid by-product formation

(Goel *et al*, 1995) while preserving the wild-type's rapid growth rate. Subsequent network analysis and mutation (Lee *et al*, 2000; Zhu *et al*, 2001) confirmed the efficacy of the mutational strategy. More recently, Siddiquee *et al* (2004) have also studied the metabolic regulation of *E. coli* lacking a functional *pykF* gene. They also observed low flux ratios through the lactate- and acetate-forming pathways, which corroborates the efficacy of the mutation strategy.

Interestingly, network analysis indicates that for similar glucose uptake, there are two feasible extreme point flux distributions where the pyruvate kinase-catalyzed flux is zero and acid production is nil. The bounds of the flux space differ mainly by the flux ratios at the G6P branch point (Ghosh *et al*, 2004). In one case, the flux to the hexose monophosphate pathway (G6P→ Ribulose 5P) is greater than G6P isomerization to F6P, and the other extreme point exhibits the opposite partitioning at the branch point. The analyte screening methodology illustrated for the test problem was applied to the problem of designing an NMR experiment that would resolve the fluxes in a pyruvate kinase-deficient mutant. This particular problem of flux scenario differentiation is representative of a practical problem, and the larger size of the network was envisioned to provide useful information on how fast computations can be performed for less abstracted network problems.

In order to implement the analyte screening methodology, NMR data was generated using a linear combination of $\alpha_1 = 0.4$ and $\alpha_2 = 0.6$ of the fluxes in the two solutions. We assumed that resolving a "middle of the road" case (i.e. one extreme solution does not dominate) would be reasonably challenging. Relaxing the "middle of the road" assumption would be straightforward and would entail repeating the analysis for several different values of α to explore the utility of different analytes.

Analyte Set

The analysis limited the scope to cytosolic or protein hydrolysis-derived amino acids, which are commonly used in NMR experimentation. The computational methods, however, are not limited to evaluating amino acid analytes. The criteria used to select candidate amino acid analytes for the purpose of demonstrating the computational methods were (1) abundance and (2) metabolic origin. Abundance was considered a virtue because as will be discussed in more detail later, the more abundant an analyte is, the less sample preparation effort and usage of NMR facility time

that is required, which lowers the cost of experimentation. Metabolic origin considers whether all the analytes are derived from different metabolites within mainstream metabolism or if some of the analytes report on the same metabolite. We attempted to keep the example set tractable while yet insuring that the analytes are derived from different and important metabolic nodes and with some redundancy. Within the Discussion section, we comment further on origin of the example analytes in terms of their "connectivity to metabolic nodes."

The cytosolic candidates chosen were the three most abundant amino acids: glutamate, alanine, and glycine (Tempest & Meers, 1970). The later amino acid analytes are synthesized from the central metabolites, α-ketoglutarate, pyruvate, and 3-phosphoglycerate, respectively. For brevity, the top four prevalent amino acids that can be derived from protein hydrolysis were chosen for demonstration purposes. They were alanine, glutamate/glutamine, aspartate/asparagine, and glycine. The latter four amino acids are synthesized from pyruvate, α-ketoglutarate, oxaloacetate, and 3-phosphoglycerate, respectively.

Based on metabolic origin, alanine and glycine are somewhat redundant. Glycine is derived from 3-phosphoglycerate, which maps carbon-to-carbon into alanine's precursor, pyruvate. Thus, both alanine and glycine are derived from similarly labeled 3-carbon intermediates in glycolysis. However, the spectra and isotopomer possibilities differ for a two (glycine) versus a three-carbon compound (alanine). Overall, the analyte set can be viewed as covering the top three abundant amino acids in the cytosol, and all three are diverse in metabolic origin. For the top four abundant amino acids in proteins, three are diverse in origin and two (alanine and glycine) both report on the labeling of 3-carbon compounds in glycolysis. Thus, the screening results can be viewed as a report on the overall utility of different analytes as well as a means for differentiating between two analytes that report on the labeling of similarly labeled glycolytic metabolites. There is also overlap between the cytosol- and protein hydrolysis-derived analyte candidates. However, the relative abundances and thus signal strengths of each analyte will differ in a cytosol- versus a protein hydrolysis-derived sample. As will be seen later, abundance difference can affect the utility of an analyte. Finally, other amino acids could be included in the computational screens that follow such as the much less abundant aromatic amino acids that are synthesized from erythrose-4-phosphate.

Analyte Screen 1: How Do Analyte Spectra Contrast Over Entire Feasible Flux Space?

Unlike the "test" problem, different precursor (i.e. glucose) labelings are possible. Quantitatively, *detectable* contrast is provided by high ^{13}C enrichment (i.e. significant NMR line intensity) and how the allowable fluxes variably enrich different carbons in the analyte. Thus, glucose labeling was considered in parallel with extreme point comparisons when determining the contrast powers of the analytes. Singly (from 1 through 6) and uniformly labeled glucose were evaluated. The NMR spectra for the candidate analytes were simulated in case of the two extreme point solutions (corresponding to $\alpha_1 = 1$ and $\alpha_1 = 0$).

For 1-^{13}C and 2-^{13}C labeling of glucose, it was found that the extreme point spectra were notably different, while the other labels provided very little contrast. For example, glutamate spectra (three internal carbons) are shown in Figure 5 for when 1-^{13}C, 2-^{13}C, or U-^{13}C labeled glucose is used. When 1-^{13}C and 2-^{13}C labeled glucose is used, the spectra corresponding to the extreme point solutions can easily be distinguished by eye. The same occurs for the other analytes (not shown). The difference between glutamate spectra is difficult to discern by eye when U-^{13}C labeled glucose is used (Fig. 5), as is also the case for other analytes (not shown). Difficult to differentiate glutamate spectra also arise in both of the extreme cases when 3, 4, 5, or 6-^{13}C labeled glucose is used (not shown).

In order to quantify the above observations, the spectra were vectorized as was previously described. The results are shown in Table 4. As far as comparing extreme point spectra is concerned, using 1-^{13}C labeled glucose endows glutamate, alanine, and aspartate with better contrast, while using 2-^{13}C labeled glucose provides better contrast for glycine. Using U-^{13}C labeled glucose, however, was found to provide nil contrast between the two extreme points for all the analytes considered.

Label selection can be viewed from another vantage point. The *spectral strength* can be defined as the sum of the line intensities present in the analyte's NMR spectrum. From this standpoint, U-^{13}C glucose produces the best spectra in both the extreme cases ($\alpha_1 = 0$ and $\alpha_1 = 1$). This result follows intuitively from the fact that U-^{13}C glucose yields the lowest fraction of unlabeled analyte; hence, greater total NMR signal will be yielded per quantity of analyte. Moreover, more analyte carbons will become labeled despite the activity of CO_2 evolving reactions that expel ^{13}C

from the cell. For example, when 1-^{13}C glucose is used, the reaction 1-^{13}C Glucose 6P → Ribulose 5P + $^{13}CO_2$ removes ^{13}C from the system that would otherwise label amino acid analytes. Thus, a trade-off exists between spectrum contrast and strength.

One-way to resolve the trade-off would be to use a mixture of 1-^{13}C and U-^{13}C labeling (i.e. the winning labels in the two categories). The effect of label mixing can be quantified by computing the analytes' IDVs when 1-^{13}C or U-^{13}C glucose is used, and then taking an average of the IDVs that is weighted according to the proportions of 1-^{13}C and U-^{13}C glucose. Based on the weighted IDVs and line splitting rules, a spectrum can be calculated. Table 5 summarizes the results when different ratios of 1-^{13}C:U-^{13}C labeling are used. The distinction between extreme points is better when there is a higher fraction of 1-^{13}C label in the mixture (e.g. 0.2135 with 75% 1-^{13}C as opposed to 0.0473 with 25% 1-^{13}C in the case of glutamate). However, the spectral strength decreases as the fraction of 1-^{13}C labeling grows.

Based on the results in Table 5, one way to select the glucose labeling is to maximize contrast. Another way is to use "enough" 1-^{13}C glucose to provide "sufficient" contrast, and add U-^{13}C glucose to provide more strong and thus usable spectral features. Using glutamate as an example, the different results obtained when the former and latter approaches are used are shown in Figures 5 (top) and 6, respectively. Using 1-^{13}C glucose (Fig. 5 top) results in a simple spectrum with a few strong features that significantly change between extreme points. Using a 25/75 mixture of U-^{13}C and 1-^{13}C glucose results in a larger number of strong features that significantly change between extreme points. Both approaches could work in terms of yielding fluxes, but the approach that provides more "varying and above the noise data points" could be argued to provide to better flux results. For example, the effect of random noise on the ratio of the intensity of strong and weaker lines could skew the results more when there are few line features to analyze.

Interestingly, Fischer *et al* (2004) used U-^{13}C and 1-^{13}C labeled glucose in two separate experiments and concluded that a comparable 20/80 mixture of U-^{13}C/1-^{13}C glucose allowed one to resolve the metabolic network of *Escherichia coli* MG1655 strain with high resolution. It is

also interesting to note that 1-^{13}C and U-^{13}C are the least expensive labelings of glucose.[1] Based on the results (Figs. 5 (top) and 6) and prior experimental work (Fischer *et al*, 2004), the subsequent analyses focused on refining the selection of analytes when a 75/25 mixture of 1-^{13}C/U-^{13}C labeled glucose is used.

Note that using a 25/75 mixture of U-^{13}C/1-^{13}C glucose will provide high contrast power and spectral strength for three of the four analyte candidates. Glycine was "sacrificed" to be suboptimal because the chosen labeling for glucose can (1) provide high spectral strength and contrast power for the majority of the candidate analytes and (2) the subset majority reports on the labeling of *different* metabolites (i.e. pyruvate, α-ketoglutarate, and oxaloacetate). Overall, the experimental design used here emphasizes the reporting ability of analytes derived from different metabolites, while retaining some contrast power and spectral strength for an analyte that provides redundant information on the labeling of 3-carbon compounds in glycolysis. Similar computations could enable the implementation of an alternative experimental design (Mollney et al, 1999).

Analyte Screen 2: Do Analytes Invert Differently to Fluxes?

All the candidate molecules were subjected to the "Multiple Flux Distribution from NMR data test." The spectra from the signal molecules all inverted to a $\alpha_1 = 0.4$ and $\alpha_2 = 0.6$ combination of the extreme point flux distributions, when labeling is provided by a 75/25 mixture of 1-^{13}C/U-^{13}C label of glucose. Multiple flux distributions do not arise in this case upon the inversion of NMR data because even though the *E. coli* problem is significantly larger than the prior example, several extracellular measurements are available. The degrees of freedom is reduced in the NLP of the inverse problem, which guides the solution to the optimal α's.

Analyte Screen 3: Economics-Related Overall Utility Index

While all the candidate analytes were found to yield spectra that reliably invert to fluxes, the analytes differ in contrast power. The analytes also differ in abundance. Low abundance and/or spectral strength is problematic because signal to noise will be low, and thus little information for flux resolution will be provided unless effort is expended for sample concentration and/or many free induction decays are collected. Both add to the cost of experimentation. Long-term

[1] Stable Isotopes 1997-98, Cambridge Laboratories Inc.

data acquisition can be especially costly because signal to noise increases only with the square-root of the number of free induction decays. Throughput through an NMR facility is also lowered.

To introduce the relative cost of NMR experimentation time, we assume that the inverse of NMR time depends on the product of analyte abundance and spectral strength. Analyte abundance is now introduced as opposed to using only the spectral strength because the spectral strength was based on an analyte's IDV. An IDV records the mole fractions of isotopomers, which must sum to one. Thus, the spectra (e.g. Figs. 5 top and 6) and spectral strengths (Table 5) computed from the information in IDVs are all on the same concentration basis. Therefore, scaling further by relative analyte abundance attempts to further distinguish the analytes based on the magnitude scale the experimentalist would actually "see." An overall utility index can then be defined as data quality (i.e. contrast power) per experiment cost (measure of NMR instrument time requirement). This index is provided by the product of the three terms: (1) contrast power of an analyte, (2) spectral strength, and (3) an analyte's relative abundance.

The alternative to using linear scaling to develop an overall utility index is to contrast analtyes on the basis of similar signal to noise ratio (SNR). That is, one can compare analytes by constrast power/time (i.e. cost) to achieve similar SNR. As before, one would expects that as abundance and/or spectral strength decrease, more time (i.e. money) will be required to acquire a usable spectrum. However, the statistics of signal averaging introduces a nonlinearity between SNR and time allotted to spectrum acquisition. Here, SNR depends on the square-root of the number of free induction decays. Because the number of free induction decays is proportional to total instrument time, an alternative version of the utility index is (contrast power)* {spectral strength * relative abundance)^2. When either version of the overall utility index is used, the same rank order for the analytes results.

The abundance of amino acids in the cytosol of *E. coli* (Tempest & Meers, 1970) and within proteins (Lehninger) is shown in Table 6. Using the aforementioned utility index, the results in Table 6 indicate that glutamate emerges as the best choice for an analyte when NMR analytes are obtained via cytosolic extraction. Alanine is ranked second followed by glycine. In case of

protein hydrolysis-derived analytes, aspartate emerges as the best followed in rank order by glutamate, alanine, and glycine.

Alternative Analyte Selection Using the Analytic Hierarchy Process

The previously described manner of ranking analytes entailed posing a multifactor objective function, and then attempting to maximize the function. Situations can conceivably arise, however, where different experimentalist- or problem-dependent values are placed on the various aspects of resolving fluxes from NMR data. For example, NMR facilities may be scant so minimizing facility usage is a priority and sufficient versus maximal flux resolution is acceptable. In this case, one could pose a different objective function or focus entirely on relative analyte abundance and spectral strength. An alternative approach that retains all information while allowing different relative emphases to be introduced is provided by the Analytic Hierarchy Process (AHP; Saaty, 2000). This section illustrates the use of the AHP for the purpose of ranking NMR analytes.

AHP is based on the innate human ability to use information, experience, and/or individual preferences to estimate relative magnitudes through paired comparisons. These comparisons are used to construct ratio scales on a variety of dimensions, both tangible and non-tangible. Thus, the method proceeds from judgments on comparisons with respect to dominance, which is the generic term for expressing importance, preference or likelihood, of a property which they have in common, to their numerical representation according to the strength of that dominance and then derives a ratio scale. Arranging these dimensions in a hierarchic or network structure allows a systematic procedure to organize our basic reasoning and intuition by breaking down a problem into its smaller constituent parts. The AHP thus leads from simple pairwise comparison judgments to the priorities in the hierarchy.

In this problem, the following criterions[2] have been considered while ranking the NMR analytes:

(i) Extreme point comparison (ExP)

(ii) Abundance data (Abun) based on both cytosolic content (cc) and frequency in proteins (fP)

[2] Multiple Global Optima Test (MGOT) has not been included as a criterion in AHP, because all the candidates successfully passed this test. In the event that the analytes perform differently in this test, MGOT should be included as a criterion with the appropriate weights.

(iii) Sum * abundance data.

In the first step, the criterions are compared in pairs, *e.g.* sum*abundance data is considered 7 times more important than ExP and 9 times more important than Abun. Once all the pair wise comparisons in a group are completed, a scale of relative priorities is derived from them. This process is repeated for all groups on all levels (*i.e.* the analytes).

Next, each criterion is assigned a weight relative to a node in the next higher level, *e.g.* both cytosolic content and frequency in proteins are considered to contribute equally to the net abundance criterion. Since each of these nodes carries only its priority of the unit goal, the derived scale is suitably transformed through multiplication by the weights of the criteria so that each alternative receives its portion of the unit goal.

Two separate hierarchical analyses have been considered, one considering frequency of amino acids in proteins (Table 7) while the other takes into account cytosolic content (Table 8). The extreme point criterion is, however, the same in both cases.

Thus, in the final tabulation, the decision-maker judges the importance of each criterion in pair-wise comparisons and the outcome is a prioritized ranking of each alternative. Tables 7 and 8 show the performance of each NMR analyte according to the abovementioned criteria. Aspartate ranks first when considering frequency in proteins while glutamate ranks first when abundance in cytosolic content is taken into account. This ranking is consistent with the results of our structured methodology used in the previous sections.

It is important to note, however, that the relative weights are subjective and are at the discretion of the experimenter. Hence, changing their values might lead to a change in overall ranking of the candidates. In this particular instance, however, glutamate and aspartate perform significantly well in their respective categories, so, they are unlikely to be ‘dethroned’ if the weights are changed.

The advantage of the AHP is that the structure of the problem represented in the hierarchy can be easily extended to include more comparison criteria and/or more metabolites. Further, the

judgment process is so simple that the decision makers are in command of the problem as they see it.

Computational Characteristics for Reversion

Glutamate was found to provide high contrast and resolving power. To conclude, we report on the computational characteristics of inverting a glutamate spectrum to fluxes using the methods described in this paper. A least-square like objective function (akin to Eqn. 2a) was used to reconcile the NMR spectra of glutamate to the fluxes. When BARON was used for the optimization, the fluxes shown in Figure 7 were obtained, where the values of some important fluxes are indicated numerically while those with zero flux values have been removed. The problem had 910 variables and 897 constraints, and took ~30 minutes to solve.

DISCUSSION

A structured and problem-centric framework for screening metabolites for their potential use as NMR analytes for flux identification has been described. The first step entails probing the contrast power of each analyte candidate over the whole feasible flux space. Having MILP solutions in-hand facilitates this step. Such solutions provide the flux bounds a network can attain when varying amounts of extracellular data exist. Moreover, the labeling of a precursor such as glucose can be considered in this step such that contrast and the desired data type (e.g. few strong lines vs. many strong spectral lines) are obtained. Next, the analytes are evaluated on the ability of their NMR spectra to invert to a unique flux distribution. Here, it is of interest to determine in advance whether an analyte's NMR spectrum could revert to clonal solutions that are characterized by highly similar values of an objective function. Determining this prospect in advance can facilitate how analyte sets are constructed as well as illuminate what uncertainties may lie within an experimental approach.

Some examples of how economics can be introduced to generate cost-benefit scenarios were also presented. One example sought to determine an effective yet a minimal and easily isolated analyte set. Another example first found that the least expensive forms of labeled glucose provided significant contrast and resolving power, and then analytes could be ranked with an overall index that considered the NMR facility time requirement. Finally, as an alternative to posing an arbitrary objective function, we presented how a relative value-based scheme, the Analytical Hierarchy Process, can also be used to rank analytes.

When these screens were applied to the case of an *E. coli* mutant system, glutamate emerged as a highly useful NMR analyte in that it can provide high flux resolution per unit of experimental cost. This result is interesting from numerous standpoints. First, glutamate has desirable NMR properties. Glutamate's three central carbons have equivalent spin-lattice relaxation times and Nuclear Overhauser Enhancement effects. Thus, signal magnitudes from each of the three internal carbons are directly comparable to one another because saturation affects each nucleus equally. Secondly, glutamate is readily isolated by cytosolic extraction due to its high abundance. Glutamate is also a prevalent amino acid residue in proteins. Overall, glutamate provides a strong and easily interpreted signal. For the particular *E. coli* problem examined, glutamate NMR data also inverted reliably to the correct fluxes.

The candidate analytes were chosen based on high abundance in the cytosol or within proteins, and metabolic origin. Abundance was deemed desirable because less strenuous sample preparation would be required to obtain analyte spectra the exhibited significant signal to noise. Concerning origin, the candidate set attempted to cover most of the amino acid families as a starting point for demonstration. Also, an analyte such as glutamate is often qualitatively viewed as lying "deep" within metabolism with the result that ample flux history is "recorded" as the ^{13}C nuclei within labeled glucose traverse the branching, converging, and molecular rearranging reactions of glycolysis, the Krebs cycle, and other pathways. The simple reasoning applied to origin could exclude some useful analytes and "deep" is a subjective notion.

The recent application of the "Kevin Bacon Test" (Fell & Wagner, 2000) to the structure of the *E. coli* core metabolism suggests some potentially interesting overlaps may exist between the test's conclusions and the NMR analyte selection problem. Fell and Wagner (2000) found that glutamate lies at the core of *E. coli* metabolism as indicated by the minimum separation distance between it and other metabolites. Alanine scored second in minimal separation distance.

Interestingly, glutamate and alanine are first and second in abundance in the *E. coli* cytosol, and within proteins they are the top two residues when glutamine is scored as equivalent to glutamate. Thus high abundance and minimal separation distance correlate. Here, the correlation to abundance may reflect the development of high and coordinated mass action potentials that can drive connected reactions (e.g. Majewski & Domach, 1989).

One speculative connection between "degree of separation" and NMR is given that glutamate and alanine are closely connected to all other metabolites, then the origin and thus the labeling of glutamate and alanine are highly representative of the network's reaction processes. Overall, the connected feature plus their abundance may underlie why glutamate and alanine score and perform well when used as NMR analytes. The connected nature of glutamate and alanine may also suggest that the set of analytes used for demonstration purposes is a reasonable starting point. Scrutinizing these possible connections further might entail editing out transaminase reactions from a "degree of separation analysis." Transaminase reactions transfer amino groups

between different carbon skeletons as opposed to altering or building new carbon skeletons, which shuffles ^{13}C and creates isotopomers.

Acknowledgment

This publication is based upon work supported, in part, by EPA Grant R82 9589 and NSF Grant BES-0224603 in cooperation with the Interagency Program in Metabolic Engineering. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the EPA or the National Science Foundation.

Appendix: Analytic Hierarchy Process

The Analytic Hierarchy Process is a framework of logic and problem-solving that provides a qualitative basis for organizing observations in order to systematically rank-order decisions. Specifically, the AHP is a multi-attribute modeling methodology which was first developed and applied by Saaty (). The process is based on the innate human ability to use information and experience to estimate relative magnitudes through paired comparisons. These comparisons are used to construct ratio scales on a variety of dimensions both tangible and intangible. Arranging these dimensions in a hierarchic or network structure allows a systematic procedure to organize our basic reasoning and intuition by breaking down a problem into its smaller constituent parts. The AHP thus leads from simple pairwise comparison judgments to the priorities in the hierarchy.

In the physical world, ratio scales are often employed to establish relations between different objects or phenomena. The AHP derives ratio scales of different magnitudes of a set of elements by making paired comparisons. When used in a process systems engineering context, AHP can be a powerful tool for comparing alternative design concepts.

Let us assume that we are faced with a problem where we wish to choose one out of 3 alternatives C_i (i = 1,2,3) on the basis of 4 criteria K_j (j = 1,…,4). Further, criteria K_2 and K_3 can have subcriteria k_{2j} (j = 1,2) and k_{3n} (n = 1, 2, 3) associated with them. The various elements of this decision problem are organized into a hierarchy shown in Figure A1.

Each level has multiple nodes with respect to which the alternatives in the next level are compared. The first step is to compare the elements in each level in pairs. The comparisons are made using judgments based on various subjective criteria to interpret data according to their contribution to the parent node in the level immediately above. When making a comparison, one chooses the smaller or lesser one of the pair as the unit and estimates the larger one as a multiple of that unit, based on the perceived intensity factor. As we shall see presently, each criterion (or subcriterion or alternative) is assigned a weight relative to a node in the next higher level. Since each of these nodes carries only its priority of the unit goal, the derived scale is suitably transformed through multiplication by the weights of the criteria so that each alternative receives its portion of the unit goal.

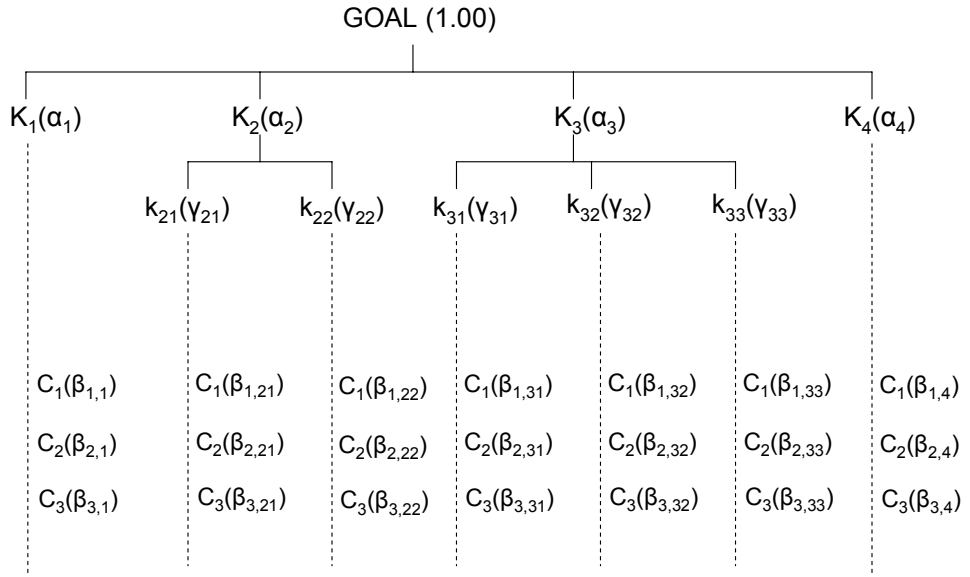
Table A1 shows the judgments for comparing the criteria of the second level (K_j's) with respect to the goal. The table comprises of the pairwise comparison matrix (*A*, say) where the number in the i^{th} row and the j^{th} column gives the relative importance of K_i as compared with K_j. The matrix *A* is square and also reciprocal (*i.e.* $a_{ij} = 1/a_{ji}$). The diagonal entries will be 1 because of the self-comparison of the criteria. Since half of the remainder are reciprocals by virtue of the inverted comparisons, 6 (or more generally, *n(n-1)/2*, where *n* is the number of elements being compared) independent comparisons are needed. In order to express dominance of one criterion over another, the authors have proposed the numbers 3,5,7 and 9, corresponding to the verbal judgments "moderately more", "strongly more", "very strongly more", and "extremely more", respectively. The numbers 2, 4, 6 and 8 are used when a compromise is in order. Clearly, the choice of a linear scale and the maximum value is somewhat arbitrary but this is simply a reflection of the general problem of quantifying preferences in multi-attribute decision models (Dubois, 200x). It is quite reasonable to use a bounded positive evaluation scale, *i.e.* no infinite preferences are allowed. The elements of the matrix *A* are therefore finite and positive. The entries in the priorities column are the eigenvalues (λ) of the matrix. The use of eigenvalues consolidates the 16 relative intensity ratios of the matrix into 4 measures of intensity and also eliminates any numerical inconsistency that might arise due to the use of a linear scale in AHP. This new scale is called the *derived* scale. Saaty () has proposed that an easy way to get a good approximation to the priorities is to multiply the elements in each row together and take the n^{th} root, where *n* is the number of elements. The column of numbers thus obtained are then normalized by dividing each entry by the sum of all entries. Instead of giving K_1 over K_2 the intensity a_{12}, we use the values in the derived scale to form $\lambda_1 * a_{12}$. This procedure is carried out for all entries in the matrix.

The next step in AHP is to repeat the procedure for every criterion. The subcriteria under each criterion are compared as to their importance with respect to that criterion to derive their local priorities. Table A2 shows such as example where the subcriteria k_{3n} are compared with respect to the criterion K_3. The product of the local priorities and those of the parent criteria gives the global priorities of the subcriteria.

Finally, the three alternatives, $C_{1\text{-}3}$, are compared with respect to each subcriterion (*e.g.* Table A3), or criterion (as in the case of K_1 or K_4, which have no subcriteria) and weighted by the overall priority of the subcriteria. The sum of these products taken for each alternative is the overall priority of that alternative.

Thus, a simple comparison matrix for the subcriteria would look like Table A4.

The alternatives are then ranked 1-3 according to the decreasing values of the overall rank, in other words, the alternative with the highest overall rank is selected.



$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$$

$$\gamma_{21} + \gamma_{22} = \alpha_2 \; ; \quad \gamma_{31} + \gamma_{32} + \gamma_{33} = \alpha_3$$

$$\sum_{i=1}^{3} \beta_{i,k} = \alpha_k , k = \{1,4\} \; ; \sum_{i=1}^{3} \beta_{i,l} = \gamma_l , l = \{21,22,31,32,33\}$$

Figure A1: Overall schematic of the AHP methodology showing the different criteria and alternatives. Symbols in brackets indicate the priorities associated with each criterion or alternative. The set of equations given below the chart indicates the relationships between the priorities.

Table A1: Comparison of second level criteria. The 4x4 matrix *A* is reciprocal.

GOAL	K_1	K_2	K_3	K_4	Priorities
K_1	1	a_{12}	a_{13}	a_{14}	λ_1
K_2	$1/a_{12}$	1	a_{23}	a_{24}	λ_2

K_3	$1/a_{13}$	$1/a_{23}$	1	a_{34}	λ_3
K_4	$1/a_{14}$	$1/a_{24}$	$1/a_{34}$	1	λ_4

Table A2: Comparison of subcriteria k_{3n} with respect to the criteria K_3. The global priorities are obtained by multiplying the local priorities with the priority of the parent criteria.

$K_3(\alpha_3)$	k_{31}	k_{32}	k_{33}	Priorities	
				Local	Global
k_{31}	1	$c_{3,12}$	$c_{3,13}$	λ_1^3	$\lambda_1^3*\alpha_3 = \gamma_{31}$
k_{32}	$1/c_{3,12}$	1	$c_{3,23}$	λ_2^3	$\lambda_2^3*\alpha_3 = \gamma_{32}$
k_{33}	$1/c_{3,13}$	$1/c_{3,23}$	1	λ_3^3	$\lambda_3^3*\alpha_3 = \gamma_{33}$

Table A3: Comparison of alternatives $C_{1\text{-}3}$ with respect to subcriteria k_{33}.

$k_{33}(\gamma_{33})$	C_1	C_2	C_3	Priorities	
				Local	Global
C_1	1	c_{12}	c_{13}	λ_1^{33}	$\lambda_1^{33}*\gamma_{33}$
C_2	$1/c_{12}$	1	c_{23}	λ_2^{33}	$\lambda_2^{33}*\gamma_{33}$
C_3	$1/c_{13}$	$1/c_{23}$	1	λ_3^{33}	$\lambda_3^{33}*\gamma_{33}$

Table A4: Final overall ranking of the alternatives.

	K_1	K_2		K_3			K_4	Overall
		k_{21}	k_{22}	k_{31}	k_{32}	k_{33}		Rank
C_1	$\beta_{1,1}$	$\beta_{1,21}$	$\beta_{1,22}$	$\beta_{1,31}$	$\beta_{1,32}$	$\beta_{1,33}$	$\beta_{1,4}$	$\Sigma\beta_{1,i}$
C_2	$\beta_{2,1}$	$\beta_{2,21}$	$\beta_{2,22}$	$\beta_{2,31}$	$\beta_{2,32}$	$\beta_{2,33}$	$\beta_{2,4}$	$\Sigma\beta_{2,i}$
C_3	$\beta_{3,1}$	$\beta_{3,21}$	$\beta_{3,22}$	$\beta_{3,31}$	$\beta_{3,32}$	$\beta_{3,33}$	$\beta_{3,4}$	$\Sigma\beta_{3,i}$

Abbreviated example: For illustration purposes, let us assume that there are 4 objectives (which maybe either criteria or alternatives), X_1, X_2, X_3 and X_4. Using a 1-9 scale, we might arrive at a possible comparison matrix:

$$A = \begin{bmatrix} 1 & 1/5 & 1/3 & 1/7 \\ 5 & 1 & 3 & 5 \\ 3 & 1/3 & 1 & 3 \\ 7 & 1/5 & 1/3 & 1 \end{bmatrix}$$

The next step is to compute the priorities. Using Saaty's approximation, we compute the priorities vector by $\sqrt[4]{1 \times 1/5 \times 1/3 \times 1/7}$ for the 1st row, and so on, and finally normalizing. Thus, we get the priorities or weights: $w = [0.0579\ 0.5452\ 0.2438\ 0.1531]^T$

Note that, by construction, $\sum_{i=1}^{4} w_i = 1$.

The weights w would be used in summing the measures as required in the evaluation of the AHP.

References

Castro, CD., Koretsky AP, and Domach, MM.. (1999). "Performance trade-offs in in situ chemostat NMR." *Biotechnol Prog* **15**(2): 185-95.

Fell & Wagner …

Fischer, E, Zamboni, N, and Sauer, U. (2004) "High throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived ^{13}C constraints." *Anal. Biochem.* 325(2):308-316.

Forbes, NS, Clark, DS and Blanch, HW. (2001) Using isotopomer path tracing to quantify metabolic fluxes in pathway models containing reversible reactions. *Biotech.Bioengg.* Volume 74, No 3, pp 196-211.

Ghosh, S., Zhu, T, Grossmann, IE, Ataai, MM and Domach, MM. (2004). "Closing the loop between feasible flux scenario identification for construct evaluation and resolution of realized fluxes via NMR." *Comput. Chem. Engg. (in press)*

Goel, A, Lee, JW, Domach, MM and Ataai, MM. (1995). Suppressed acid formation by co-feeding glucose and citrate in *Bacillus* cultures: Emergence of pyruvate kinase as a potential metabolic engineering site. *Biotechnol. Prog.*, 11, pp 380-386.

Lee S, Phalakornkule C, Ataai MM, Domach MM and Grossmann, IE. (2000) Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput. Chem. Eng.* Volume 24, pp 711-716.

Majewski, RA and Domach, MM. (1989). Effect of Regulatory Mechanism on Hyperbolic Reaction Network Properties. *Biotechnol. Bioeng* . 36, pp 166-178.

Majewski, RA and Domach, MM. (1990). Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnol. Bioeng* .35, pp 732-738.

Mollney M, Wiechert W, Kownatzki D, de Graaf AA (1999) Bidirectional reaction steps in metabolic networks IV: Optimal design of isotopomer labeling experiments". *Biotechnol. Bioengg* 66(2) pp 86-103.

Phalakornkule, C, Lee, S, Zhu, T, Koepsel, R, Ataai, MM., Grossmann, IE and Domach, MM. (2001). A MILP-based flux alternative generation and NMR experimental design strategy for Metabolic Engineering. *Metabolic Engineering* 3, pp 124-137.

Saaty, TL (2000). Fundamentals of Decision Making and Priority Theory. 2nd Ed. Pittsburgh, PA: RWS Publications.

Sahinidis, NV and Tawarmalani, M. (2002). GAMS/BARON 5.0 - Global Optimization of Mixed-Integer Nonlinear Programs.

Schmidt, K, Carlsen, M, Nielsen, J and Villadsen, J. (1997b). Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol. Bioeng.* 55, pp 831-840.

Schmidt, K, Carlsen, M., Nielsen, J and Villadsen, J. (1997a). Quantitative analysis of metabolic fluxes in *E. coli*, using 2-dimensional NMR spectroscopy and complete isotopomer models. *J. Biotechnol.*

Siddiquee, KA, Arauzo-Bravo, MJ and Shimuzu,K (2004). Effect of a pyruvate kinase (pykF-gene) knockout mutation on the control of gene expression and metabolic fluxes in *Escherichia coli*. *FEMS Microbiol. Lett.* Jun1; 235(1), pp 25-33.

Tempest, DW and Meers, JL. (1970). Influence of environment on the content and composition of microbial free amino acid pools. *J. Genl. Microbiol.*, 64, pp 171-185.

Zhu T., Phalakornkule C., Koepsel R.R., Domach M.M., and M.M. Ataai. (2001) Cell growth and by-product formation in a pyruvate kinase mutant of *E. coli*. *Biotechnol. Prog*. 17: pp 624-628.

Zhu, T, Phalakorkule, C, Ghosh, S, Grossmann IE, Koepsel RR, Ataai, MM, and Domach, MM. (2003). A metabolic network analysis and NMR experiment design tool with user interface-driven model construction for depth-first search analysis. *Metabolic Engineering* 5 pp. 74-85.

Table 1. Alternate Optima Obtained from the Recursive MILP Formulation

Fluxes	Solution I	Solution II	Solution III	Solution IV
r_1^*	1	1	1	1
r_2^*	0.3	0.3	0.3	0.3
r_3	1.3	0	1.3	0
r_4	0	1.3	0	1.3
r_5	0.4	1.3	0	1.7
r_6	0	1.3	0	1.3
r_7	0.4	0	0	0.4
r_8	0.4	0	0	0.4
r_9^*	0.4	0.4	0.4	0.4
r_{10}	0	0.4	0.4	0
*Denotes fixed fluxes.				

Table 2. NMR to Flux Inversion Results When Using Different Analytes

Analyte	v_3	v_4	v_5	v_6	v_7	v_8	v_{10}	v_{11}	ε_8
hypothesized	0.43	0.87	1.19	0.87	0.32	0.78	0.08	0.46	0.59
A & E	0.43	0.87	1.19	0.87	0.32	0.78	0.08	0.46	0.59
C	1.3	0	0	0	0	0	0.4	0	0.12
O	1.3	0	0	0	0	0	0.4	0	0.37
A	1.3	0	0	0	0	0	0.4	0	0.9
E	0.43	0.87	1.19	0.87	0.32	0.78	0.08	0.46	0.59

Table 3. Clonal Solutions Obtained with Metabolite C as the NMR Analyte

Fluxes	Actual	Multiple Global Solutions from C		
		Solution I	Solution II	Solution III
v_1	1	1	1	1
v_2	0.3	0.3	0.3	0.3
v_3	0.43	1.3	0.37	0.47
v_4	0.87	0	0.93	0.83
v_5	1.19	0	1.27	1.13
v_6	0.87	0	0.93	0.83
v_7	0.32	0	0.34	0.3
v_8	0.78	0	0.18	0.6
v_9	0.4	0.4	0.4	0.4
v_{10}	0.08	0.4	0.058	0.095
v_{11}	0.46	0	0.15	0.3

Table 4. Performance Comparison of Candidate NMR Analytes when 1-^{13}C, 2-^{13}C, or U-^{13}C Labeled Glucose is Used

Analytes	Angular Deviation of the Vectorized Extreme Point Spectra			sum(NMR)					
				1-13C		U-13C		2-13C	
	1-13C	U-13C	2-13C	alpha1=1	alpha1=0	alpha1=1	alpha1=0	alpha1=1	alpha1=0
glutamate	0.486	2.70E-06	0.1629	0.2193	0.3985	0.75	0.7476	0.2063	0.2669
alanine	0.0908	9.44E-05	0.0784	0.1403	0.2428	0.9999	0.9996	0.1889	0.2498
glycine	0.046	8.50E-04	0.0886	0.1403	0.2347	0.9999	0.9986	0.1908	0.2498
aspartate	0.5536	1.90E-06	0.1564	1.0034	0.5681	1.7497	1.7483	0.1998	0.2651

Table 5. Performance Comparison of Candidate NMR Analytes when Different Mixtures of 1-^{13}C/U-^{13}C Labeled Glucose are Used

Analytes	Angular Deviation			sum(NMR)					
	1-13C:U-13C			0.25:0.75		0.5:0.5		0.75:0.25	
	0.25:0.75	0.5:0.5	0.75:0.25	alpha1=0	alpha1=1	alpha1=0	alpha1=1	alpha1=0	alpha1=1
glutamate	0.0473	0.1091	0.2135	0.6943	0.713	0.5885	0.6331	0.4299	0.5083
alanine	0.0335	0.0939	0.1778	0.792	0.8101	0.5794	0.6156	0.3622	0.4161
glycine	0.0342	0.0958	0.1814	0.792	0.8082	0.5794	0.612	0.3622	0.411
aspartate	0.102	0.2292	0.2987	1.5803	1.5664	1.4002	1.3625	1.2078	1.1368

Table 6. Combination of NMR Resolution & Amino Acid Abundance of Candidate Analytes when 75/25 1-^{13}C/U-^{13}C Labeled Glucose is Used*

	1	2	3	4	1x3x4	2x3x4
Amino acids	pool-free content	relative frequency in proteins	extreme point NMR comparison	average sum of NMR peaks	combination based on cytosolic content	combination based on relative frequency in proteins
glutamate	72.5000	10.8100	0.2135	0.4691	7.2611	1.0827
alanine	17.5000	13.0200	0.1778	0.3892	1.2108	0.9009
glycine	3.7500	7.8100	0.1814	0.3866	0.2630	0.5477
aspartate	0.0000	9.9000	0.2987	1.1723	0.0000	3.4666

*Abundance data obtained from Tempest *et al*, 1970

Table 7. Implementation of the Analytic Hierarchy Process for the Selection of NMR Analytes (considering frequency of amino acids in proteins only)

	sum*abun fP	ExP	abun fP	Overall Priorities	Overall Rank
Analytes	0.772	0.1734	0.0545		
glutamate	0.1875	0.3650	0.2718	0.2229	2
alanine	0.1875	0.0795	0.2718	0.1733	3
aspartate	0.5625	0.3650	0.2718	0.5124	1
glycine	0.0625	0.0314	0.1017	0.0592	4

Table 8. Implementation of the Analytic Hierarchy Process for the Selection of NMR Analytes (considering abundance of amino acids in cytosol only)

	sum*abun cc	ExP	abun cc	Overall Priorities	Overall Rank
Analytes	0.772	0.1734	0.0545		
glutamate	0.6430	0.3650	0.4846	0.5861	1
alanine	0.2683	0.0795	0.3360	0.2392	2
aspartate	0.0372	0.3650	0.0448	0.0945	3
glycine	0.0515	0.0314	0.0448	0.0477	4

Figure Legends

Figure 1. Hypothetical metabolic network used to demonstrate NMR calculations and analyte screens. (a) The conversion of precursors **G**, **S**, and **Q** to excreted products **P** and **R** are shown as well as a cyclic series of reactions. (b) Shown are the metabolic fluxes obtained from the optimization using synthetic NMR data. The black circles indicate ^{13}C while the white ones are ^{12}C. The grey metabolites, **A** and **E** are the NMR analytes, which are amongst the analytes that can be used to successfully solve the inverse NMR → fluxes problem (Equation 2).

Figure 2. Performance of NMR signal molecules with respect to contrast power over the entire flux space. Analyte **C** performs the best.

Figure 3. Existence of "multiple" global minima or clones of a function $y = f(x)$. Point P is clearly a local minimum while Q is the global minimum. Within the prescribed tolerance, however, R is also a 'global' minimum.

Figure 4. The metabolic network of *E. coli*. The glycolytic, pentose phosphate pathway, malic enzyme, and Krebs cycle reactions are shown. The double-ended arrows represent potentially reversible reactions and the solid-headed arrows depict the net directions. Abbreviations: AcCoA, acetyl-CoA; CIT, citrate; DHAP, dihydroxy-acetone phosphate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; FDP, fructose-1,6-bisphosphate; G6P, glucose-6-phosphate; KG, α-ketoglutarate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PGP, phosphoglycerol phosphate; PG3, 3-phosphoglycerate; P2G, 2-phosphoglycerate; PYR, pyruvate; R5P, ribose-5-phosphate; Succ, succinyl-CoA; T3P, triose phosphate; 3PG, 3-phosphoglycerate; MAL, malate; S7P, sedoheptulose-7-phosphate; aa, amino acids; na, nucleic acids.

Figure 5. NMR spectra of glutamate when derived from 1-^{13}C, 2-^{13}C, or U-^{13}C labeled glucose. Comparison of the extreme point spectra (corresponding to α_1 = 0 and 1) shows that 1-^{13}C glucose provides the highest contrast, followed by 2-^{13}C glucose, while using U-^{13}C glucose results in significantly more spectral lineforms, but hard to discern contrast.

Figure 6. NMR spectra of glutamate when derived from a mixture of 75/25 1-^{13}C and U-^{13}C glucose. Comparison of the extreme point spectra (corresponding to α_1 = 0 and 1) to that of

glutamate derived from 1-^{13}C glucose (see Fig. 5) shows that the mixture provides more lineforms that change between the extreme points.

Figure 7. Metabolic Flux Map of *E. coli*. The numbers associated with some important fluxes are indicated, while pathways with zero flux values have been removed from the network. The abbreviations for the species can be found from the legend for Figure 4.

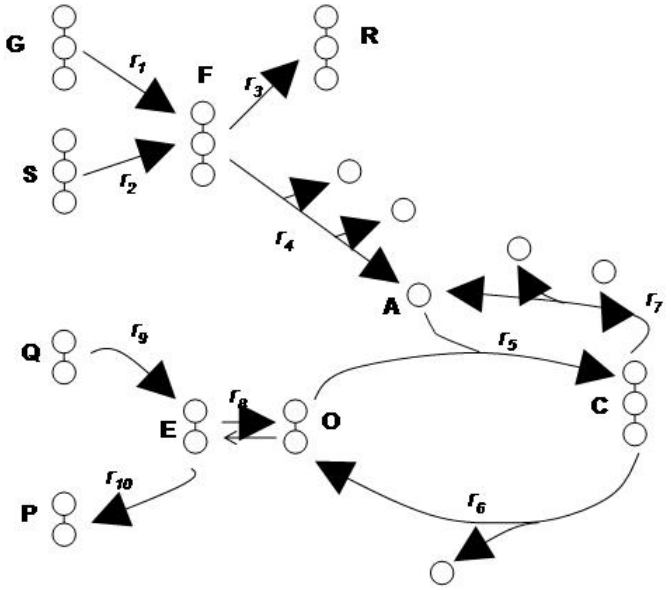


Figure 1a.

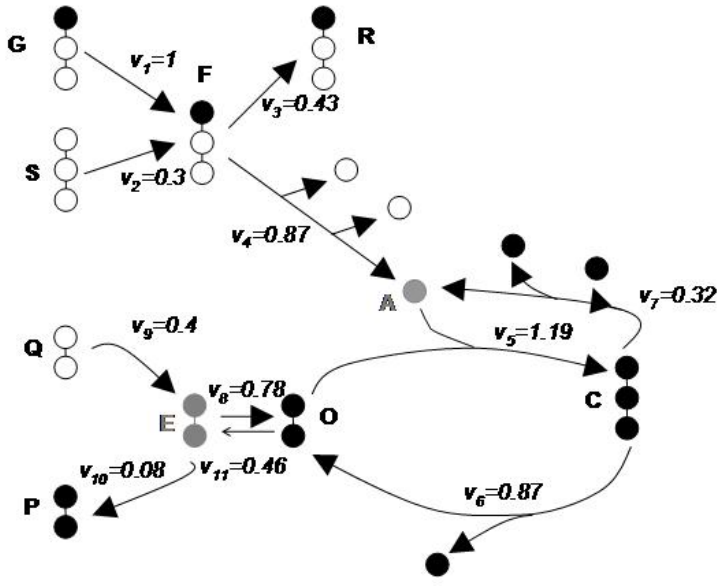


Figure 1b.

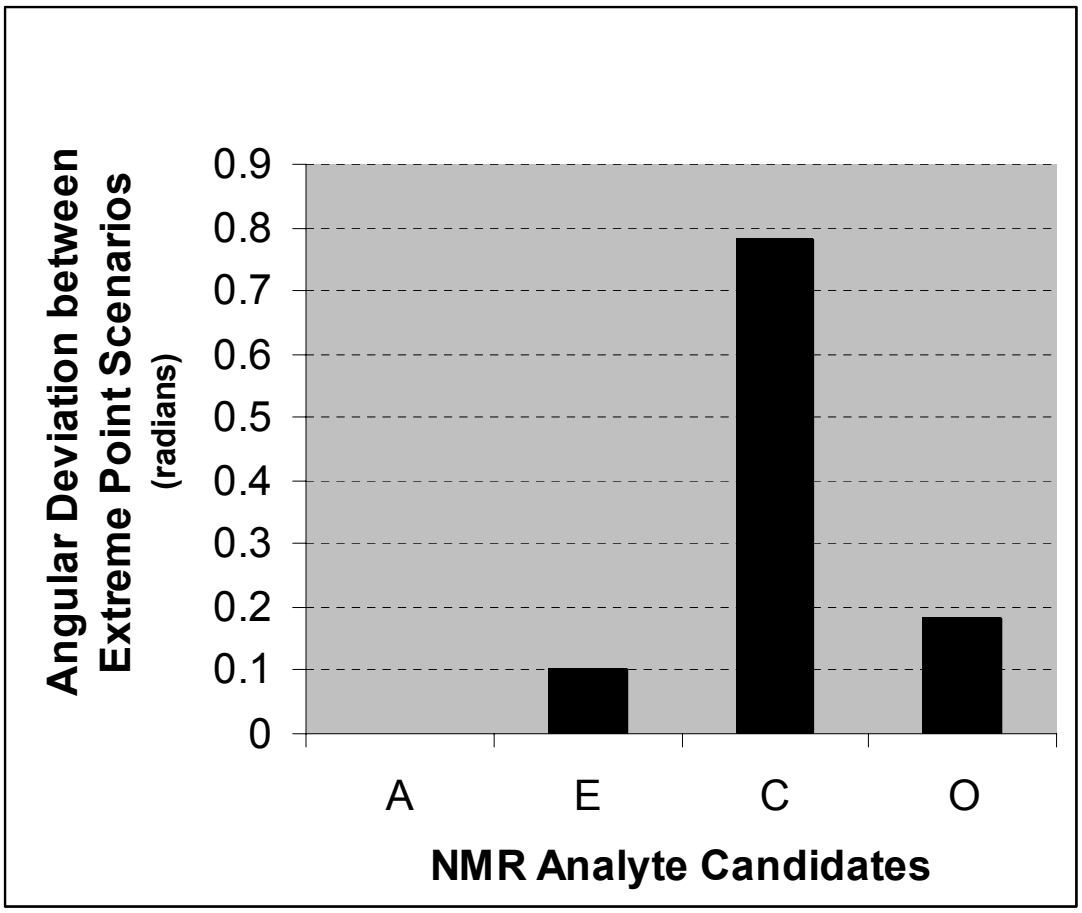


Figure 2.

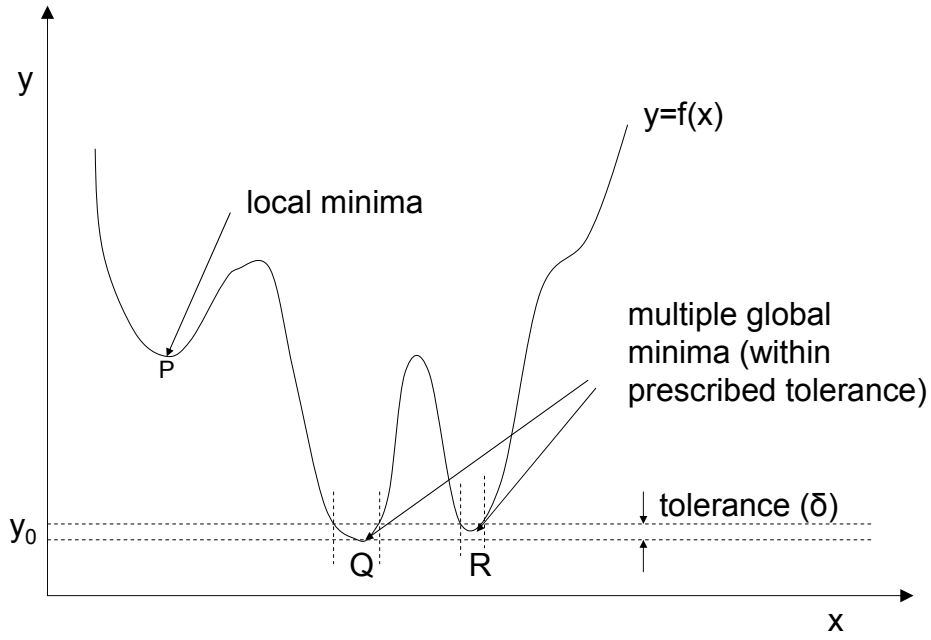


Figure 3.

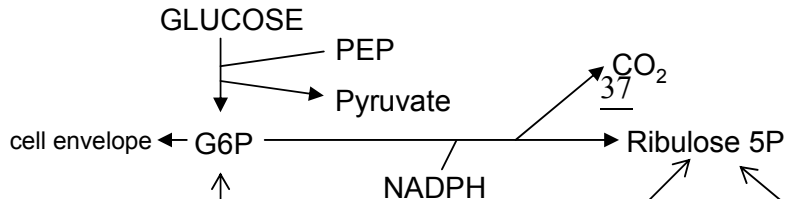


Figure 4

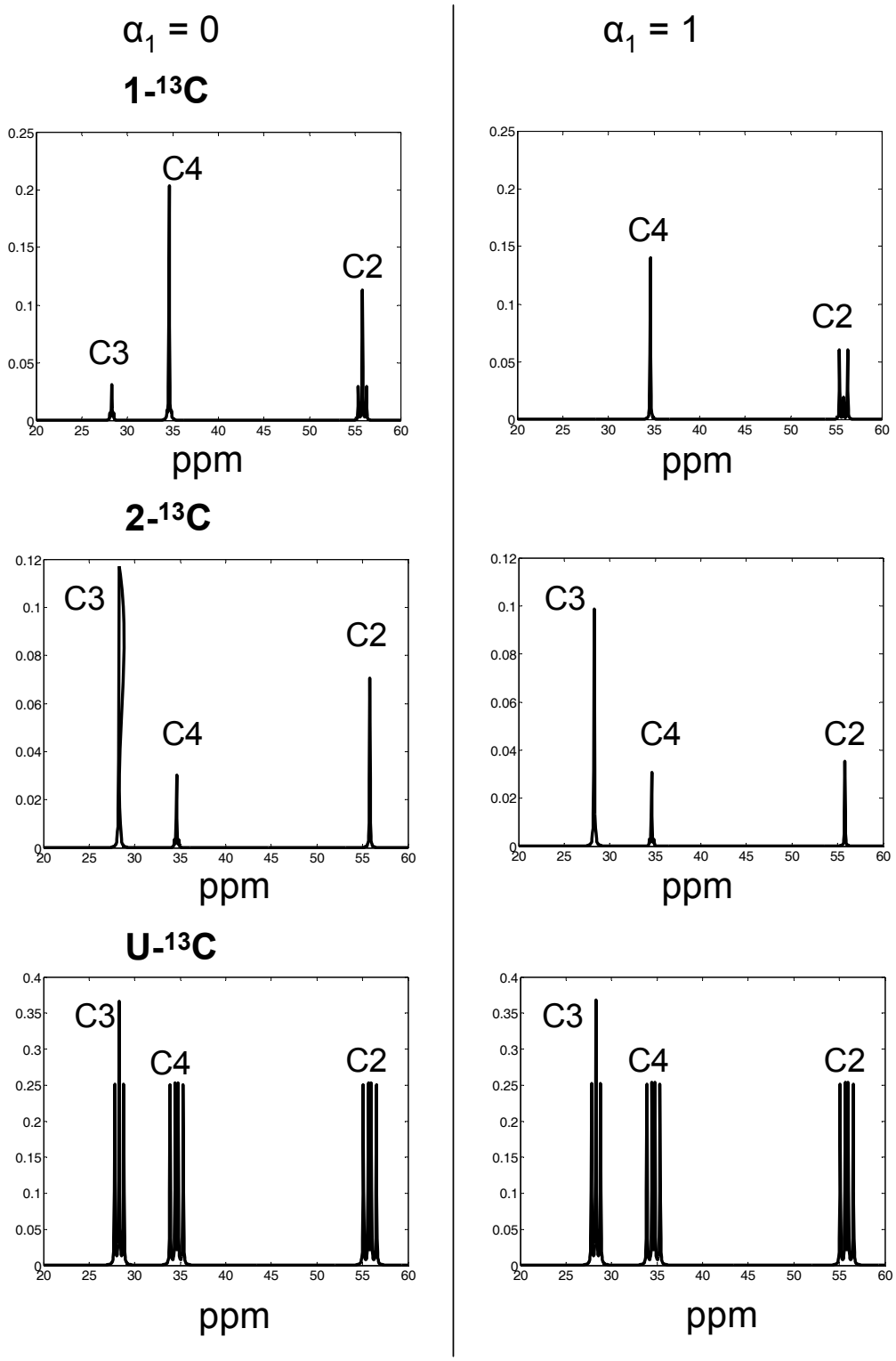


Figure 5

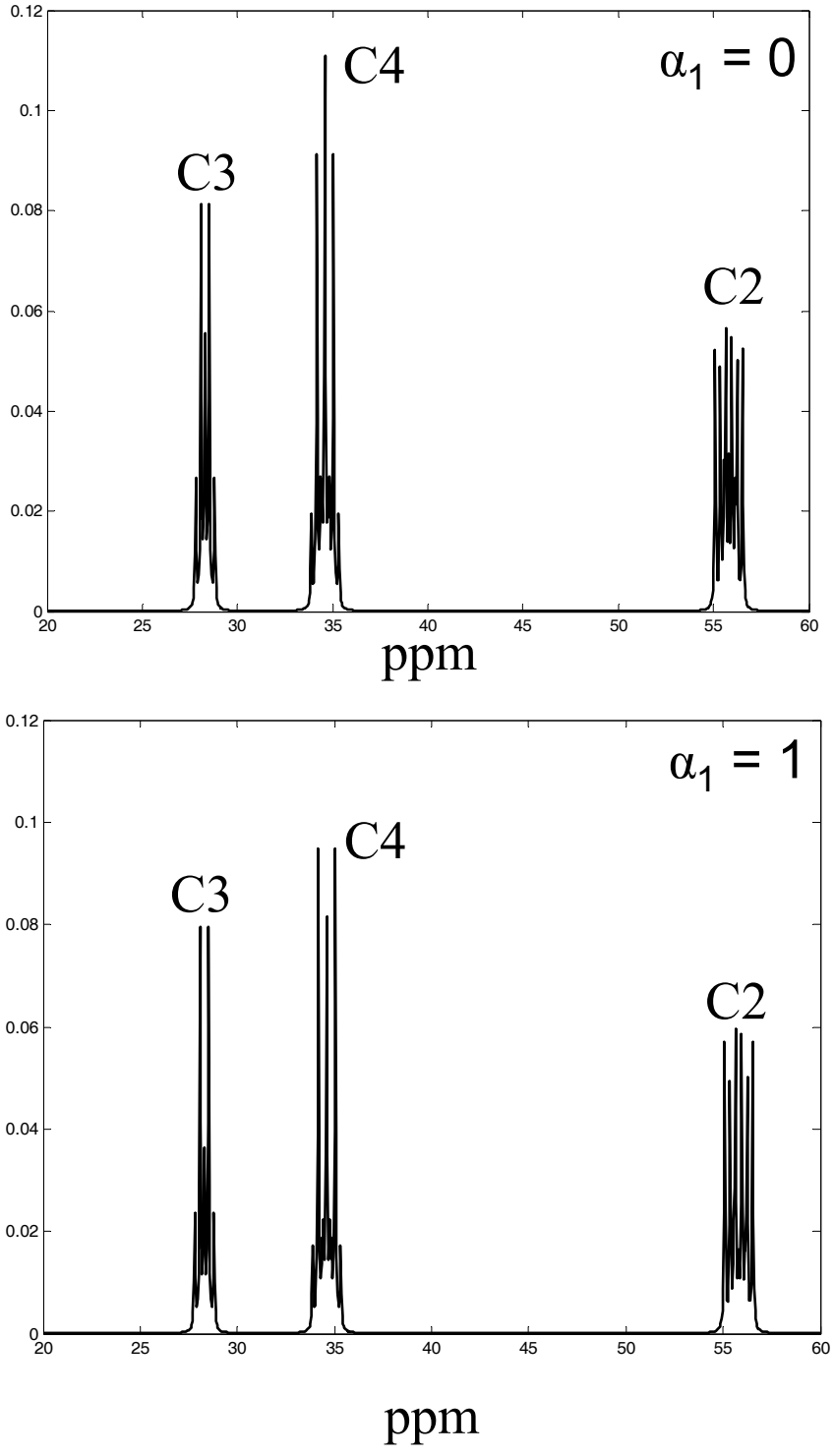
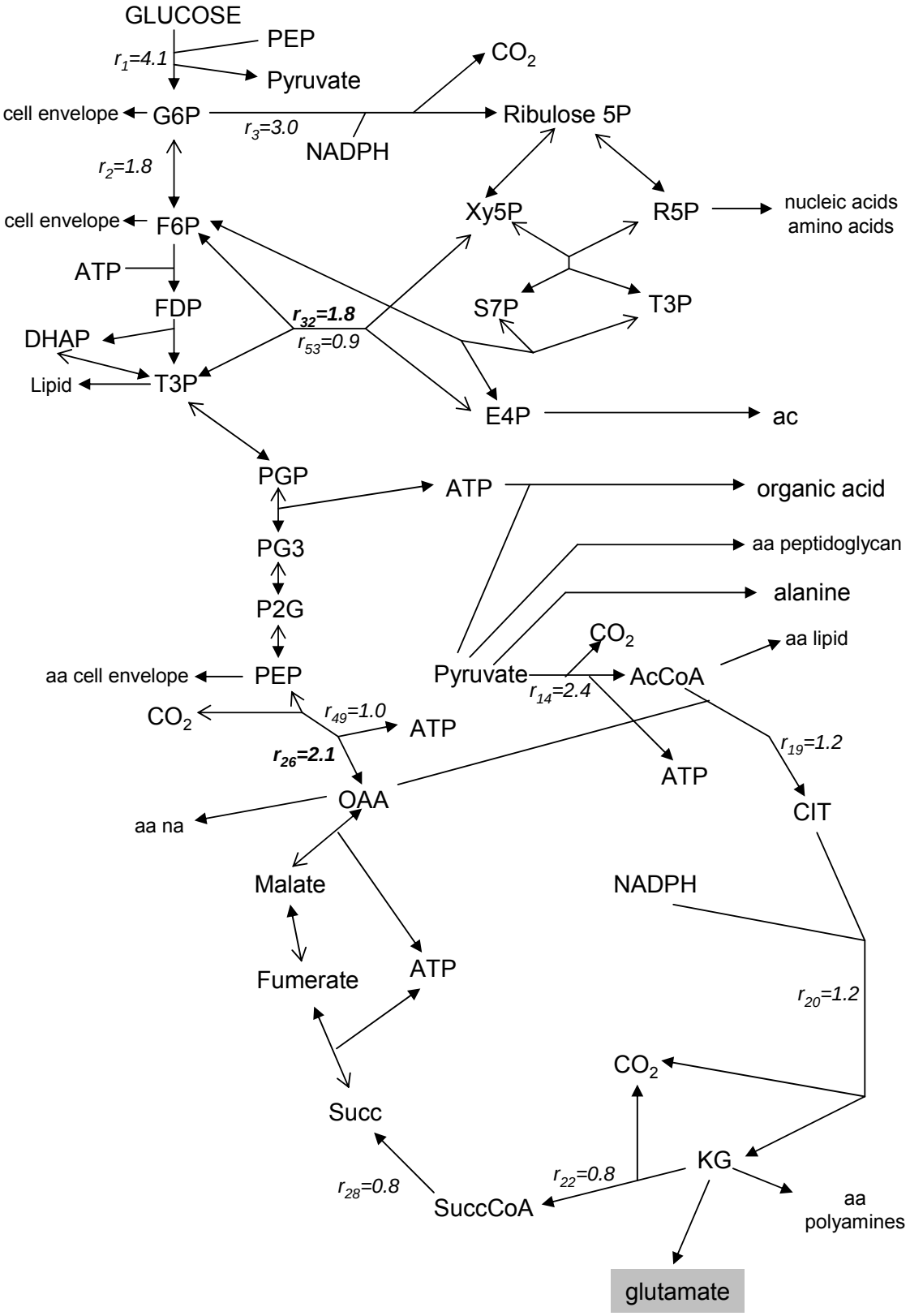


Figure 6


GLUCOSE
PEP
r_1=4.1
Pyruvate
CO_2
cell envelope
G6P
r_3=3.0
NADPH
Ribulose 5P
r_2=1.8
Xy5P
R5P
nucleic acids
amino acids
cell envelope
F6P
ATP
FDP
S7P
T3P
r_{32}=1.8
r_{53}=0.9
DHAP
Lipid
T3P
E4P
ac
PGP
ATP
organic acid
PG3
aa peptidoglycan
alanine
P2G
CO_2
aa lipid
aa cell envelope
PEP
Pyruvate
AcCoA
r_{14}=2.4
CO_2
r_{49}=1.0
ATP
r_{26}=2.1
ATP
r_{19}=1.2
OAA
aa na
CIT
Malate
NADPH
Fumerate
ATP
r_{20}=1.2
CO_2
Succ
KG
r_{22}=0.8
r_{28}=0.8
SuccCoA
aa
polyamines
glutamate