Journal of Experimental Botany, Vol. 55, No. 400,
Understanding Photosynthetic Performance Special Issue, pp. 1177–1186, May 2004
DOI: 10.1093/jxb/erh090 Advance Access publication 8 April, 2004

Journal of Experimental Botany

Applications of metabolic modelling to plant metabolism

M. G. Poolman*, H. E. Assmus and D. A. Fell

School of Biology and Molecular Science, Oxford Brookes University, Headington, Oxford OX3 OBP, UK

Received 12 September 2003; Accepted 8 December 2003

Abstract

In this paper some of the general concepts underpinning the computer modelling of metabolic systems are introduced. The difference between kinetic and structural modelling is emphasized, and the more important techniques from both, along with the physiological implications, are described. These approaches are then illustrated by descriptions of other work, in which they have been applied to models of the Calvin cycle, sucrose metabolism in sugar cane, and starch metabolism in potatoes.

Key words: Calvin, computer, kinetic, metabolism, modelling, photosynthesis, potato, starch, structural, sucrose.

Introduction

The use of computers to model metabolic systems has a history dating back at least as far as the early 1960s with the work of Garfinkel and Hess (1964). Since that time, the performance/price ratio of computers has continued to rise exponentially, operating systems and user interfaces have improved beyond recognition, and there is a plethora of software (free or proprietary, specialized or general purpose), suitable for undertaking metabolic modelling studies. During this time, a number of theoretical tools and concepts have emerged, notably within the fields of Metabolic Control Analysis (MCA) (Kacser and Burns, 1973; Heinrich and Rapoport, 1974; Kacser *et al.*, 1995; Hofmeyr *et al.*, 2002) and Metabolic Engineering (Stephanopoulos *et al.*, 1998) capable of providing a framework within which both the behaviour of metabolic systems can be understood, and strategies for the investigations of systems and models can be developed.

During this period, numerous modelling studies of plant, especially photosynthetic, metabolism have been undertaken (Giersch *et al.*, 1991; Woodrow, 1986; Hahn, 1987; Laisk *et al.*, 1989; Pettersson and Ryde Pettersson, 1988; Poolman *et al.*, 2001). Despite all this, the impact from modelling work tends to remain confined to the modelling community, with relatively little impact on the experimental community. Exceptions to this generalization can be found within the biotechnology industry, but, in the main, are restricted to manipulations of commercially important microbial metabolism, for example, Bongaerts *et al.* (2001) and Koffas *et al.* (2003).

This paper reviews, from a non-technical perspective, the techniques of metabolic modelling, and discusses some recent applications of such techniques to plant systems, with the aim of providing the non-specialist with enough of the concepts and vocabulary to interpret and apply to their own fields of interest results from modelling studies, and to provide at least a potential starting point for those contemplating modelling investigations of their own.

Modelling

As the term 'modelling' is used with different meanings in various contexts, its meaning must first be defined here: a metabolic model is simply a selected list of reactions and associated properties, assumed to be present in the system under investigation, along with with a description of the environment within which the system is assumed to reside. The information that can be gleaned from the model is determined by the information used to specify the reactions. This information must be expressed in a form usable by a particular software package, at which point the modelling process can begin.

Modelling software

In principle, modelling can be undertaken with nothing more sophisticated than a pen and paper, but, in practice, most of the activities comprising the process of modelling (see below) can only be undertaken with a software tool for any but the most trivial of models.

* To whom correspondence should be addressed. Fax: +44 (0)1865 483242. E-mail: mgpoolman@brookes.ac.uk

Software commonly used for modelling can be divided into two broad groups: general purpose mathematical software, and specialized modelling tools. The former (for example, matlab, http://www.mathworks.com) possibly offer the greatest flexibility, but place a considerable burden on the user (for example, in turning a set of reaction descriptions into a set of differential equations), a burden which, in the authors' opinion, outweighs the advantages in most cases.

Specialized modelling software has, as its design criteria, the ability to allow the user to think primarily at the level of biology, rather than mathematics. Reactions are entered in a fashion that will be familiar to the biologist, and no further mathematical ability is assumed. Commonly used packages of this type include 'WinScamp' (http://www.cds.caltech.edu/~hsauro/Scamp/scamp.htm) and 'Gepasi' (http://gepasi.dbs.aber.ac.uk/softw/gepasi.html). The disadvantage of the approach is the opposite of that of general purpose software: a relative lack of flexibility of action on the user's part.

As part of an effort to overcome these disadvantages simultaneously, a software tool 'ScrumPy' (http://mudshark.brookes.ac.uk/ScrumPy), is being developed that places metabolic modelling functionality within the context of a general purpose programming language, 'Python' (http://www.python.org). ScrumPy was used to investigate the models of the Calvin cycle and starch metabolism described in the section on Applications. The 'Jarnac' (http://www.cds.caltech.edu/~hsauro/Jarnac.htm) package has a similar design philosophy, but defines its own language.

The modelling process

Modelling investigations tend to fall, by accident or design, into a reasonably predictable pattern, as exemplified by Fig. 1. The starting point must always be an initial hypothesis: that the behaviour of the system under investigation can be explained as a function of some collection of reactions. In general, these will be enzyme-catalysed reactions, but this is not mandatory. Transport processes can, for modelling purposes, also be treated as reactions. From this point the model must be defined in a form usable by the software. The minimum level of useful definition is a list of reaction names, and their associated stoichiometries.

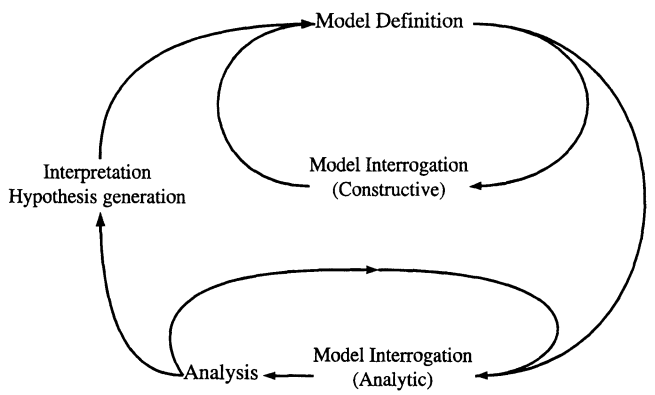


Fig. 1. Typical work flow during a modelling investigation. Scientific value is only gained during during the interpretation/hypothesis generation phase, which takes place outside of the computer.

Once the model is defined, the investigator is in a position to extract information from, i.e. to interrogate, the model. Frequently, the behaviour of an initial definition of a model is far from biological observation. This may be due to a trivial mistyping of some reaction information, or something more serious such as the omission of an essential reaction. In any event, the modeller is likely to have to spend some time altering the definition of the model, in order to achieve 'sufficiently realistic' behaviour. This is, in some ways, analogous to the initial testing and debugging of a piece of software: it can only be sensibly used once correct behaviour has been established. This phase is described as 'constructive' interrogation, as its objective is to aid the construction of the model. There is no firm definition of 'sufficiently realistic'; it depends upon the context of the investigation. A rough measurement might be the degree of surprise that such behaviour would cause were it to be observed experimentally. (Although it must be pointed out that the purpose of modelling is never merely to replicate experimental data, but to explain it.) The phase of constructive interrogation is quite distinct from the initial collation of data from which the model is to be constructed. In general, this is can require significant effort, possibly involving considerable literature searching, or even conducting limited experimental work to fill gaps in the literature.

Having achieved satisfactory behaviour in the model, it is possible to start investigating it, a stage we have called 'analytic' interrogation. In this stage, the model and the software used to investigate it can be regarded as a virtual laboratory: any component can be altered or replaced, and (within very broad limits) any aspect of its behaviour can be determined. It is at this point that modelling realizes its unique advantages. Not only can any experiment be undertaken, but it is also possible to replicate any experiment many thousands of times under varying circumstances, allowing the user to build up a characteristic description of the system's behaviour. The results of such interrogation may naturally lead to further interrogation, but will lead at some point, it is hoped, to a useful physiological interpretation, an end in its own right, or generate new hypotheses, which may either be tested via the model, completing the cycle, or experimentally.

It is worth emphasizing, that, although the analytic interrogation is the point at which the computer does the most work, in practice, this tends to take a small proportion of the total time in a modelling project, and is of little or no

value if due consideration is not given to the other stages of the process.

Kinetic modelling

Kinetic models are the most widely published metabolic models, and the most intuitively accessible to the novice user. Each reaction in the model is defined in terms of its stoichiometry and rate equation. The user must also specify values for the various kinetic parameters (V_{max}, K_m etc.), and initial values for metabolite concentrations[†]. Having done this, a number of elementary operations may be carried out on the model. (1) Values of individual metabolite concentrations or parameters may be varied. (2) The current set of reaction rates, as determined by the current set of parameter and metabolite values, can be evaluated. (3) The values of reaction rates and metabolite concentrations after the model system has evolved for some period time can be determined. (4) The reaction rates and metabolite values at which there is no further time-dependent change in these values (the steady-state) may be determined.

More sophisticated forms of interrogation may then be carried out in terms of these operations, for example, time-course simulations are achieved by simply repeating (3), and the various control coefficients of MCA can be defined in terms of (1) and (4). Although most metabolic software packages (rightly) hide such detail from the user, this may not be the case if a more general mathematical software package is used. Also some packages (ScrumPy, Jarnac) make the elementary operations available to the user, so that they may be used to form new, higher level operations not provided in the original program.

Structural modelling

Powerful though kinetic modelling is, it has certain inherent and unavoidable drawbacks. Firstly, there may be considerable uncertainty as to both the form of the kinetic equations, and the associated parameter values. Furthermore, even if all are known with certainty, then this would define a model of one specific instance of the system, and the common goal of the scientist is to be able to draw general conclusions. The large number of parameters involved makes a systematic characterization of such models impossible (e.g. sampling 10 points over 10 parameters needs 10^{10} evaluations). For similar reasons, if a kinetic model does not exhibit expected or desired behaviour, it is very difficult to determine whether or not the behaviour is a possibility of the model being investigated.

One approach to these problems is to construct models with no kinetic information, in which reactions are described purely in terms of their stoichiometry. Although at first sight it may appear that such a drastic step is bound to lead to loss of information, it in fact transpires that such an approach enables the determination of a variety of model properties that could not be found by any other means, generating information that is complementary to, rather than a subset of, that obtainable via kinetic modelling.

The stoichiometry matrix: Structural modelling in this context is so called because the focus is on the network topology (i.e. structure) of the system. It has no relation to the efforts of those producing impressive 3D models of the physical structure of proteins. Structural modelling techniques are based on the consideration of the stoichiometry matrix, **N**, of the system. This matrix is constructed such that the columns represent reactions and the rows metabolites. Each element of the matrix then represents the stoichiometric coefficient of a metabolite in a particular reaction: values of 0 indicate that the metabolite is not involved, positive values that it is produced, and negative values that it is consumed by the corresponding reaction. If a vector, $\boldsymbol{v}$, whose elements correspond to reaction rates, is defined then it transpires that the steady-state of the system can be written as:

$$\mathbf{N}\boldsymbol{v}=0 \qquad (1)$$

This simple equation is readily amenable to the well-known tools of linear algebra, which provide the foundation for the structural modelling techniques described here. A detailed mathematical explanation is beyond the scope of this paper (but see the Appendix for a brief introduction, and Heinrich and Schuster (1996) and Hofmeyr *et al.* (2002) for a more complete treatment), but the advantage of the treatment is that, as long as the stoichiometries are known, the results that follow are absolute and unambiguous.

Moiety conservation: In most realistic model systems, there are groups of metabolites sharing a common functional group whose total concentration is fixed. An archetypal example is phosphate conservation in models of carbohydrate metabolism. All software packages for metabolic modelling can report the conservation relationships in a system. The exact physiological significance of such relationships is likely to depend both upon the system studied, and the objectives of the study. However, in all cases, the effect of conservation relationships is to place strict limits as to the concentrations which metabolites can attain.

Enzyme subsets: A rather less obvious, and more recently developed, concept is that of enzyme subsets, described by

[†] Although metabolite values in a kinetic model, either at steady-state, or after simulation for some defined time period, are variables of the model, *initial* metabolite values are parameters: the modeller can change them, the model, obviously, cannot.

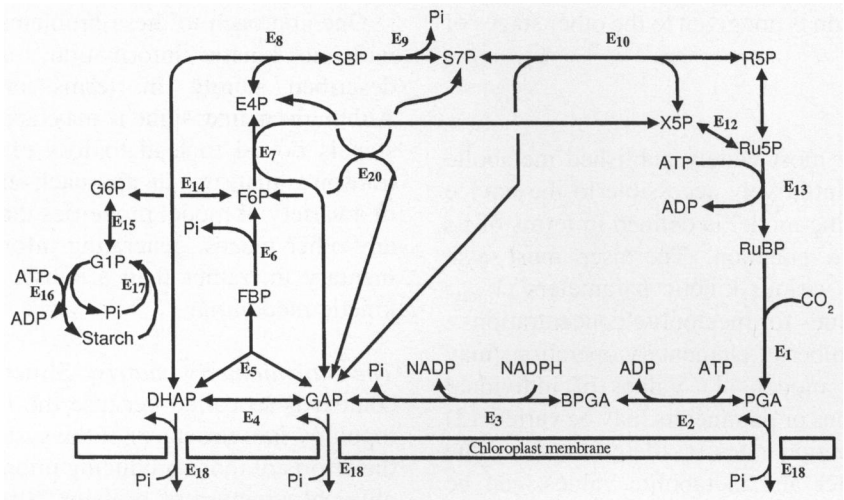


Fig. 2. Structure of the Calvin cycle model. Light reactions not shown here but were included in the model, catalysing the reaction ADP+Pi→ATP. CO_2, starch, NADP/H, and external free phosphate have fixed concentrations. BPGA, 1,3-*bis*-phosphoglycerate; DHAP, dihydroxyacetone 3 phosphate; E4P, erythrose-4-phosphate; F6P, fructose-1,6-*bis*phosphate; G1P, glucose 1-phosphate; G6P, glucose 6-phosphate; GAP, glyceraldehyde-3-phosphate; R5P, ribose-5-phosphate; Ru5P, ribulose-5-phosphate; RuBP, ribulose-1,5-*bis*phosphate; 2-PGA, 2-phosphoglycerate; 3-PGA/PGA, 3-phosphoglycerate; S7P, sedoheptulose-7-phosphate; SBP, sedoheptulose-1,7-*bis*phosphate; X5P, xylulose 5-phosphate.

Pfeiffer *et al.* (1999) and Klamt *et al.* (2002). These are groups of reactions that, at any steady-state, carry flux in a fixed ratio. Determining general biological significance of enzyme subsets is an area of current research, however, it is clear that they reflect a higher level of metabolic organization than that described by individual reactions.

Elementary modes: Perhaps the most useful theoretical tool to emerge from structural modelling is that of elementary modes (Schuster *et al.*, 1999; Schilling *et al.*, 1999; Papin *et al.*, 2003). This may be described as a minimal set of reactions that, at steady-state, catalyse some net reaction between input(s) and output(s) with a characteristic stoichiometry. The complete metabolic capability of a given system is thus encapsulated by its elementary modes. An elementary mode can be thought of as a minimal biochemical pathway, capable of operating independently from the rest of the system, determined by an objective and well-defined algorithm. The Appendix provides a more detailed introduction to the unerlying theory of elementary modes analysis.

Applications

The Calvin cycle

The Calvin cycle is probably the plant system most studied by metabolic modelling. Reasons of space constrain the authors to a brief discussion of their own work, but interested readers are directed to work described by Giersch *et al.* (1991); Giersch (1994), Laisk *et al.* (1989); Laisk and Walker (1989), and Pettersson and Ryde Pettersson (1988, 1990). Fell *et al.* (1999), Poolman and Fell (2000) and Poolman *et al.* (2001, 2003) have previously described kinetic and structural modelling studies of a model of the Calvin cycle closely based on that described by Pettersson and Ryde-Pettersson (1988), and shown in Fig. 2.

It has been shown experimentally (Heldt *et al.*, 1977; Flügge *et al.*, 1980; Portis, 1982) that in isolated chloroplasts, carbon flux through the triose phosphate-phosphate translocator (TPT, reaction E_{18} in Fig. 2) responds positively to increases in external phosphate, while net flux to starch (determined in the model by the flux carried by reactions E_{14} and E_{15}) responds negatively. Qualitatively, this behaviour is readily replicated in the model shown in Fig. 3A. Now, it is commonly asserted that the regulation of starch synthesis is predominantly controlled by ADPG-pyrophosphorylase (AGPase, lumped with starch synthase as reaction E_{16} in the model), which is itself activated by phosphoglycerate (PGA) and inhibited by stromal phosphate.

These kinetic properties are reflected in the model and it was found (data not shown) that the behaviour of stromal PGA and free phosphate is entirely consistent with this idea: as cytosolic phosphate rises, (steady-state) stromal phosphate rises and stromal PGA decreases. However, within the environment of a model this becomes a testable hypothesis: 'The response of starch synthesis flux to changes in external free phosphate is due to the sensitivity of AGPase to PGA and (stromal) free phosphate'.

It is a trivial modelling exercise to modify the kinetic function for AGPase to remove these sensitivities and repeat the determination of the model's response to external phosphate. The results of doing this are shown in Fig. 3B. It is clear that there is no qualitative difference in the two curves, and that the main effect is to change the

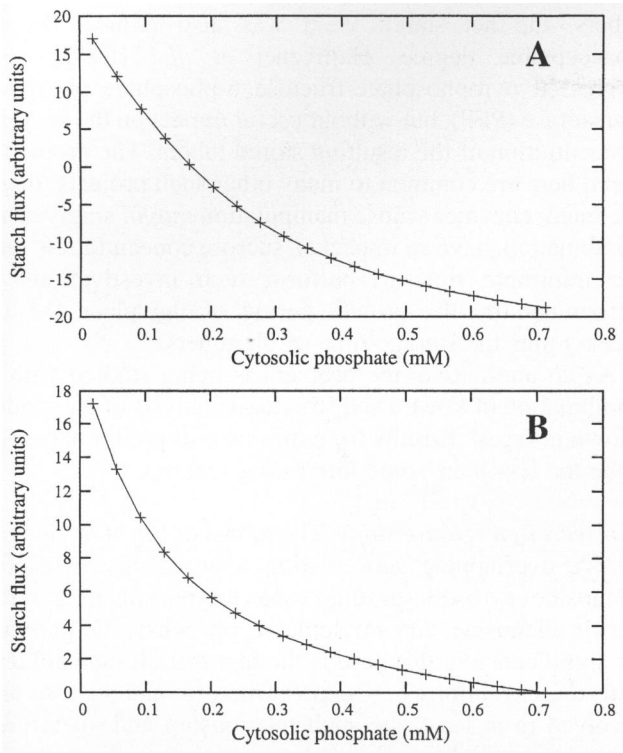


Fig. 3. The effect of the phosphate inhibition and the PGA activation of AGPase/starch synthase (E^{16}) on the response of the model shown in Fig. 2 to variation in cytosolic phosphate: before (A), and after (B) the removal of the sensitivity of E_{16} to these metabolites.

concentration of cytosolic phosphate at which the flux carried by phosphoglucomutase (PGM, E_{15}) becomes negative. Thus the hypothesis is disproved and it is possible to say that the particular kinetic characteristics of AGPase are not essential under all circumstances to explain the observed behaviour of stromal starch metabolism. Furthermore, the manipulation of other parameters anywhere in the model results in similar observations: the shape of the curve remains the same, and the only differences are quantitative (not shown). This may lead to an alternative hypothesis: 'The response of starch synthesis flux to changes in external free phosphate is due to some feature of the structure of the whole system'.

It has been shown (Poolman, 1999) that the qualitative behaviour of the detailed model can be replicated in a model that is reduced to four components: assimilation, regeneration, storage, and export. It is proposed (although this is unverified) that the features of the system that determine its behaviour are: (1) Auto-catalysis in the cycle formed by assimilation and regeneration. (2) Saturable kinetics of one or more components. (3) Concentrations of internal metabolites constrained by moiety conservation. (4) Two (or more) outputs (in this instance TP and starch)

Beyond the requirement that reactions must be saturable (which must always be the case for enzyme-catalysed reactions), the qualitative behaviour is entirely independent of the kinetics of individual reactions, it is a property of the system as a whole.

Most other aspects of the behaviour of the model are broadly consistent with experimental investigations in which enzyme activities have been modulated by genetic manipulation. However, such experiments are rather harder to interpret in the context of a kinetic model, especially with regard to the manipulation (typically by antisense repression) of a single enzyme. This is firstly because such manipulations affect the whole development of the plant, possibly leading to changes in the expression of other enzymes, secondly because such investigations report measurements on the whole plant (or at least whole tissue), whereas the model described here, represents an isolated chloroplast in an environment in which the only metabolic demand is for TP. For example, Riesmeier *et al.* (1993) and Hausler *et al.* (2000*a*, *b*) all report starch synthesis responding negatively to TPT activity, consistent with the model results. However, they also report an increased turnover of starch and neutral sugar metabolism, complicating any attempt to discuss the differences and similarities between the two sets of results. See also Poolman (1999) for a comparison of results from other genetic manipulation experiments and the model described here.

Another feature of Fig. 3 is the fact that at high levels of external phosphate the net starch synthesis flux becomes negative, corresponding to a situation in which the total exported carbon flux via TPT exceeds the assimilation flux via rubisco (E_1), with the excess carbon requirement being fulfilled by starch degradation. This raises the question as to whether or not this indicates a potential route for the chloroplast to produce TP species in the dark, i.e. can this model support a TPT flux, using starch as an initial substrate, if the light reaction flux (and hence the Rubisco flux) is zero? This too appears to be a trivial kinetic modelling problem. The investigator simply has to set the parameter representing the light reaction activity to zero, and recalculate the steady-state. In fact when this was done, all fluxes fell to zero. This leaves the much harder problem of determining whether or not this is a kinetic effect (i.e. due to an inauspicious choice of kinetic equations or parameter values) or a structural one. This is a typical example of the inherent problems associated with kinetic models. The number of parameters in the system renders impossible any kind of systematic exploration to discover values that do support a flux.

The problem is, however, readily solved by the use of elementary modes analysis. It has been shown (Poolman *et al.*, 2003) that all elementary modes of the system shown in Fig. 2, including the starch-degrading modes, have both Rubisco and the light reactions as components and there is, therefore, no set of kinetic equations or parameters that can sustain flux in the absence of light. Thus, despite the fact that there is obvious connectivity between starch and TP,

and the fact that the model can exhibit net starch degradation, this degradation is dependent on the activity of Rubisco.

Sucrose metabolism

Rohwer and Botha (2001) have recently used similar tools, in conjunction with MCA to study a rather more applied problem: that of genetically manipulating sugar cane to maximize sucrose production. These authors used a model of sucrose synthesis in sugar cane culm, comprising 11 reactions interconverting hexose and hexose phosphate species, sucrose, and sucrose phosphate. Investigation of this revealed a total of six elementary modes capable of producing sucrose, but also five futile cycles hydrolysing either ATP or UTP.

Consideration of these enabled a calculation of a control coefficient over the total flux in all futile cycles, those enzymes with large negative or positive values were presumed to be potential targets for genetic manipulation. It must be remembered that the value calculated for a particular control coefficient is specific to the parameter set at which it was calculated. Thus, although control coefficients may indicate good potential candidates for manipulation, it is still necessary to model such manipulations. These authors went on to investigate the effects of modulating the enzymes with large control coefficients over futile cycling with the aim of reducing it, thereby increasing net sucrose production.

Such an investigation can be carried out many times more quickly, and orders of magnitude more cheaply, than attempting to determine the same information by experimental genetic manipulation. It is, of course, not guaranteed that the results of genetic manipulation will exactly match the model predictions, however, use of a model in this fashion provides a rational basis for identifying those enzymes to be targeted in such an enterprise. The information from which such a model is constructed represents the modellers' (in general, not just these particular authors) knowledge of the components of the system, and the resulting model behaviour is solely determined by this knowledge.

Starch metabolism

Another example of applied modelling work is an ongoing project within this group investigating starch metabolism in potato tubers. A problem of particular interest is that of so-called 'cold sweetening': when stored for extended periods, there is notable starch degradation, accompanied by a concomitant increase in sucrose concentration, making the resulting tubers less useful to the food-processing industry (Burton, 1989).

The problem has previously been approached via genetic modification. Burrell *et al.* (1994) over-expressed phosphofructokinase (PFK), but found that, although there was a decrease in sucrose concentration in stored tubers, the net starch yield was also reduced to an unacceptable degree. Hajirezaei *et al.* (1994) over-expressed pyrophosphate-fructose 6-phosphate phosphotransferase (PFP), but without useful impact on the sucrose concentration of the resulting stored tubers. The problems faced here are common to many other such projects: there are many enzymes whose manipulation *might*, singly or in combination, have an impact on sucrose concentration, and the minimum time to perform such investigations is determined by the growth period of the plant (13–17 weeks) plus the storage time of the tubers.

As an alternative, the problem is being studied with a combination of kinetic and structural analysis of the model shown in Fig. 4. Results from this are still preliminary, but none the less have some interesting features.

Conservation relationships: The model of Fig. 4 has a total of five overlapping conservation relationships: cytosolic adenosine, cytosolic uridine, cytosolic phosphate, amyloplastic adenosine, and amyloplastic phosphate. Of particular significance in this case is the fact that all metabolites, with the exceptions of glucose, fructose, and sucrose are involved in at least one such relationship and, therefore, the only metabolites whose concentration can increase in an unconstrained fashion are those exceptions.

Enzyme subsets: Determination of the enzyme subsets revealed four subsets comprising 14 reactions from a total of 30 (the remainder operating singly). One of these subsets consists of the pair of transport reactions E_{27} and E_{29}. The implication of this is, as might be seen by inspection, that ATP transport into the plastid, and hence plastidic ATP-dependent metabolism, cannot be sustained in the absence of a mechanism transporting free phosphate in the opposite direction.

Neuhaus and MaaB (1996) have described experimental evidence for a protein-mediated phosphate uniport mechanism in isolated cauliflower-bud amyloplasts, which if shown to be universally present, would certainly solve the problem. A corollary of the fact that, in this study's model, the phospate uniporter appears in a subset with the ATP-ADP exchanger, is that not only is ATP-dependent metabolism in the amyloplast unable to proceed in the absence of the phosphate uniporter, but the steady-state export of phosphate from the amyloplast stroma cannot proceed in the absence of the ATP-ADP exchanger.

Elementary modes: The model as shown has a total of 200 elementary modes. Of these, 77 import sucrose and synthesize starch, but all produce pyruvate as well. This is unsurprising as it is only via glycolysis that the ATP requirement for starch synthesis can be satisfied. All the starch-synthesizing elementary modes contain not only ATP/ADP translocator (E_{27}) but also the phosphate transporter (E_{29}) as could be predicted from the preceding

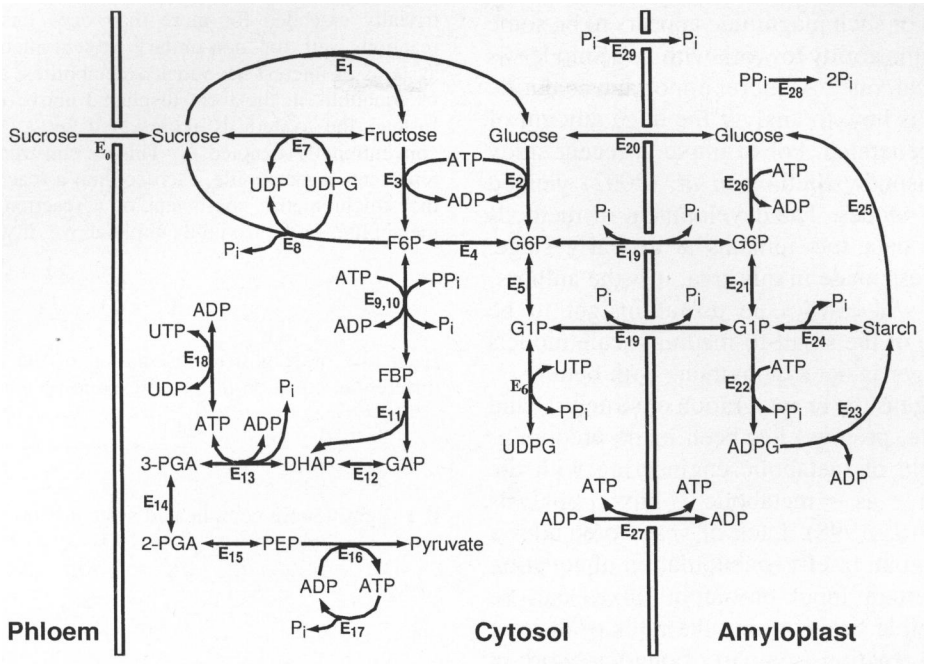


Fig. 4. Model of carbohydrate metabolism in potato tuber cells. The model as shown here corresponds to a cell of a growing tuber. (Extracellular sucrose, cytosolic pyruvate, and the starch in the plastid are the external metabolites.) Detachment of the tuber from the potato plant can be described simply by removing reaction E_0, the uptake of sucrose from the phloem. For the glucose phosphate transport E_{19} two possible substrates are considered. Whether there is only G1P or only G6P transport makes no qualitative difference to the structural analysis since PGM (phosphoglucomutase) is present in both compartments E_5 cytosolic and E_{21}). Hypothetical Pi transport across the plastid membrane (E_{29}) is included, although not experimentally confirmed, the relevance of this is addressed in the sections on Enzyme subsets and on Elementary modes. DHAP, dihydroxyacetone 3 phosphate; F6P, fructose-1,6-*bis*phosphate; G1P, glucose 1-phosphate; G6P, glucose 6-phosphate; GAP, glyceraldehyde-3-phosphate.

discussion. This observation does not prove the existence of the phosphate transporter, but rather confirms the fact, noted by Fischer and Weber (2002), that as the ATP/ADP translocator is unbalanced with respect to phosphate, it cannot carry a steady-state flux, unless some other mechanism exists to alleviate the imbalance.

There are 33 sucrose-consuming modes that have only pyruvate (with concomitant ATP synthesis) as a product. These are only realizable in the presence of one or more ATP hydrolysing reactions, represented by E^{17} in this model. When all elementary modes consuming sucrose are discounted, representing the state of the post-harvest tuber, a total of 90 remain. These all have the same net reaction, namely consumption of starch and production of pyruvate. This highlights the extreme degree of redundancy in the system, and the difficulty of pin-pointing any one enzyme as a 'key' or 'regulatory' step, in the absence of an analysis of the system as a whole.

Conclusions

Having introduced some of the basic ideas of metabolic modelling, their application to various plant systems has been discussed and demonstrated. Detailed discussion of the physiological significance of such work has deliberately been avoided, partly because this is discussed in detail in literature previously cited, but mainly because the wish was to highlight the techniques of modelling, and their application to the study of metabolic systems. It is worth repeating that the information used to construct a model is ultimately based on experimental observation, and that the behaviour of the resulting model is determined by those observations.

If the model behaviour is at variance with expectation, there are two possible causes: either a mistake has been made in the definition of the model, or the expectations are incorrect. It is under these conditions that insight is generated (there would be nothing gained if a model slavishly replicated observations already made). Either of these cases may be put to good use, if, for example, it is possible with elementary modes analysis to show that a given set of reactions cannot catalyse a net reaction shown to be present experimentally, then the presence of at least one extra reaction in the system has been demonstrated. Identifying missing reaction(s) would be rather less easy, but could be approachable by further modelling studies.

Both the theory and technology that underpin metabolic modelling continue to develop. Current challenges include the ability to make the connection between modelling and bioinformatics, and to build models containing all expressed enzymes. This is starting to become feasible for structural modelling, although the possibility of dealing

with kinetic models of such magnitude appears to be some way off. Achieving the ability to work with large models is primarily a technical one. A more important scientific problem to address is how to analyse the huge amount of data subsequently generated. For example, a recent study of *E. coli* metabolism by Stelling *et al.* (2002) yielded $\approx 5 \times 10^5$ elementary modes. The development of methods to investigate such data sets remains at an early stage. Regardless of progress made in this area, it is the authors' opinion that there will always be useful insight to be gained from the use of the small-to-medium-scale models described here. Bigger is not synonymous with better.

A further challenge is the re-integration of structural and kinetic data. To date, progress has been made predominantly within the field of metabolic engineering with the technique known as metabolic flux analysis (Stephanopoulos *et al.*, 1998). Lack of space precludes a detailed description, but, briefly, manipulation of equation 1 means that if certain input or output fluxes can be measured, it is possible to determine the rates of at least some of the internal reactions. As part of ongoing work in this laboratory, it has been found that it is possible, using similar considerations, to assign flux values to elementary modes (MG Poolman, HE Assmus, DA Fell, unpublished data).

However, perhaps the most important challenge is as much social as it is scientific: to re-integrate the work of modellers with that of experimentalists. The full potential, not only of modelling but also the spectacular experimental techniques of the post-genomic era, are unlikely to be realized if the two groups are unable to find a common language.

Appendix

The stoichiometry matrix

Metabolic systems consist of a number of metabolites and reactions that consume or produce them. Consider the simple case in which metabolite A is produced by reaction e_1 and consumed by reaction e_2:

$$\xrightarrow{e_1} A \xrightarrow{e_4}$$

The rate of change of A is:

$$\frac{dA}{dt} = v_1 + v_2$$

where v_1 and v_2 are the fluxes carried by reactions e_1 and e_2. By convention, the production of a metabolite is assigned a positive flux value, and consumption a negative value. If A is constant with respect to time then:

$$\frac{dA}{dt} = v_1 + v_2 = 0$$

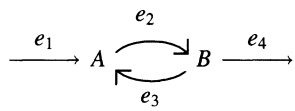
This equation represents the relationship between the two steady-state fluxes, but is independent of their actual value, and is therefore independent of their kinetic characteristics. These equations can be trivially extended for more than one consuming and producing reaction, and for non-unitary stoichiometric coefficients. Most systems of interest have many metabolites, and handling large sets of equations in the form displayed above is rather inconvenient. Instead the system is represented by a 'stoichiometry matrix' conventionally denoted **N**. This is constructed so that each row represents a metabolite, each column a reaction, and each element the stoichiometric coefficient of a reaction to the corresponding metabolite. In the trivial example above, it would be written:

$$\begin{matrix} & e_1 & e_2 \end{matrix}$$
$$N = A\begin{bmatrix} 1 & -1 \end{bmatrix}$$

Each row of **N** is a representation of the right hand side of the differential equation of the coresponding metabolite, i.e.:

$$\frac{dA}{dt} = v_1 + v_3 - v_2$$

If a slightly more complicated system is now considered,



The stoichiometry matrix is:

$$\begin{matrix} & e_1 & e_2 & e_3 & e_4 \end{matrix}$$
$$N = \begin{matrix} A \\ B \end{matrix}\begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix}$$

and

$$\frac{dB}{dt} = v_2 + v_3 - v_4$$

This set of differential equations can be written in matrix notation as:

$$\begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \end{bmatrix}$$

and at steady-state

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

if the vector of zeroes is denoted as 0 and the vector of reaction fluxes v_1 ... v_4 as $\boldsymbol{v}$, the whole equation can be written rather more succinctly as:

$$\mathbf{N}\boldsymbol{v}=0 \qquad (1)$$

This has the merit of holding true for any system capable of reaching steady-state: it is this equation that provides the foundation upon which stoichiometric modelling, including elementary modes analysis, is built.

Elementary modes

From a mathematical perspective, the determination of the elementary modes of a system revolves around the search for possible

vectors, v, that satisfy equation (1). Firstly, it should be noted that there are (assuming that the system can attain steady-state) an infinite number of solutions in the sense that if some vector v is a solution, then so is kv where k is an arbitrary real constant. For this reason, candidate solutions of v are considered in their normalized form (typically by dividing all elements by the value of the largest). Given this proviso there will still be, in general, a number of possible instances of v satisfying equation (1). For example, in the two-metabolite system above, one solution is

$$v = [2\ 1\ -1\ 2]^T$$

The biological interpretation of this is that reactions e_1 and e_4 carry equal flux, which is exactly twice that carried by e_2 and e_3, these two carry flux of the same magnitude but opposite sign, because they were defined in opposite directions. More useful, in this context, are solutions in which one or more elements of v is equal to zero. The vector

$$v = [1\ 1\ 0\ 1]^T \quad (2)$$

is also a solution in which reactions e_1, e_2, and e_4 carry equal flux, and e_3 carries none. This is more interesting as it shows that e_1, e_2, and e_4 can attain a steady-state in the absence of e_3 and can thus be regarded as an independent subsystem. There are, however, no further reactions which can be removed (i.e. have the corresponding elements of v set to zero) that fulfil equation (1). Thus equation (2) represents a *minimal* subsystem. Another possible solution is

$$v = [0\ 1\ 1\ 0]^T \quad (3)$$

which is also a minimal subsystem. Taken together this pair of vectors span the null-space of **N** (also called the kernel, $k(\mathbf{N})$). That is, any v that is a solution to equation 1 can be written as a linear combination of these two vectors. In this instance, two is the minimum number of vectors that will satisfy this requirement, however they are not unique: the vector

$$v = [1\ 0\ -1\ 1]^T$$

is a third possible, minimal and independent solution. Any two of these three vectors can be used to form $k(\mathbf{N})$. A final factor to be taken into consideration is that of reversibility. If any reactions are considered irreversible, then only vectors with non-negative elements in positions corresponding to irreversible reactions can be considered as solutions.

This now allows a mathematical definition of the elementary modes of the system as: 'The complete set of vectors from which $k(\mathbf{N})$ can be constructed and which do not violate irreversibility criteria'; which has as its biological equivalent: 'The complete set of minimal pathways in the system, capable of carrying a steady-state flux, and which do not require irreversible steps to carry negative flux'.

References

Bongaerts J, Krämer M, Müller U, Raeven L, Wubbolts M. 2001. Metabolic engineering for the microbial production of aromatic amino acids and derived compounds. *Metabolic Engineering* **3,** 289–300.

Burrell M, Mooney P, Blundy M, Carter D, Wilson F, Green J, Blundy K, ap Rees T. 1994. Genetic manipulation of 6–phosphofructokinase in potato tubers. *Planta* **194,** 95–101.

Burton W. 1989. *The potato*, 3rd edn. Harlow, England: Longman Scientific and Technical.

Fell DA, Thomas S, Poolman MG. 1999. Modelling metabolic pathways and analysing control. In: Bryant J, Burrell M, Kruger N, eds. *Plant carbohydrate biochemistry*. Oxford, UK: Bios Scientific Publishers.

Fischer K, Weber A. 2002. Transport of carbon in non-green plastids. *Trends in Plant Science* **7,** 345–351.

Flügge U-I, Freisl M, Heldt H-W. 1980. Balance between metabolite accumulation and transport in relation to photosynthesis by isolated spinach chloroplasts. *Plant Physiology* **65,** 574–577.

Garfinkel D, Hess B. 1964. Metabolic control mechanisms. VII. A detailed computer model of the glycolytic pathway in ascites cells. *Journal of Biological Chemistry* **239,** 971–983.

Giersch C. 1994. Photsynthetic oscillations: observations and models. *Comments on Theoretical Biology* **3,** 339–364

Giersch C, Sivak MN, Walker DA. 1991. A mathematical skeleton model of photosynthetic oscillations. *Proceedings of the Royal Society London, Series B* **245,** 77–83.

Hahn BD. 1987. A mathematical model of photorespiration and photosynthesis. *Annals of Botany* **60,** 157–169.

Hajirezaei M, Sonnewald U, Viola R, Carlisle S, Dennis D, Stitt M. 1994. Transgenic potato plants with strongly decreased expression of pyrophosphate-fructose-6-phosphate phosphotransferase show no visible phenotype and only minor changes in metabolic fluxes in their tubers. *Planta* **192,** 16–30.

Hausler R, Schlieben N, Flugge U. 2000*a*. Control of carbon partitioning and photosynthesis by the triose phosphate/phosphate translocator in transgenic tobacco plants (*Nicotiana tabacum*). II. Assessment of control coefficients of the triose phosphate/phosphate translocator. *Planta* **210,** 383–390.

Hausler R, Schlieben N, Nicolay P, Fischer K, Fischer K, Flugge U. 2000*b*. Control of carbon partitioning and photosynthesis by the triose phosphate/phosphate translocator in transgenic tobacco plants (*Nicotiana tabacum* L.). *Planta* **210,** 371–372.

Heinrich R, Rapoport TA. 1974. A linear steady-state treatment of enzymatic chains; general properties, control and effector strength. *European Journal of Biochemistry* **42,** 89–95.

Heinrich R, Schuster S. 1996. *The regulation of cellular systems*, Chapter 3. London, England: Chapman and Hall.

Heldt HW, Chon JC, Maronde D, Herold A, Stankovic Z, Walker D, Kraminer A, Kirk M, Heber U. 1977. Role of orthophosphate and other factors in the regulation of starch formation in leaves and isolated chloroplasts. *Plant Physiology* **59,** 1146–1155.

Hofmeyr J-HS, Snoep JL, Westerhoff HV. 2002. Kinetics, control and regulation of metabolic systems. http://www.ki.se/icsb2002/Snoep+Westerhoff.pdf. Tutorial Notes for the Third International Conference on System Biology.

Kacser H, Burns J. 1973. The control of flux. *Symposium of the Society for Experimental Biology* **27,** 65–104.

Kacser H, Burns JA, Fell DA. 1995. The control of flux. *Biochemical Society Transactions* **23,** 341–366.

Klamt S, Schuster S, Gilles ED. 2002. Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism of in purple nonsulphur metabolism. *Biotechnology and Bioengineering* **77,** 734–750.

Laisk A, Eichelmann H, Eatheall A, Walker DA. 1989. A mathematical model of carbon metabolism in photosynthesis: difficulties in explaining oscillations by fructose 2,6-*bis*phosphate regulation. *Proceedings of the Royal Society London, Series B* **237,** 389–415.

Laisk A, Walker DA. 1989. A mathematical model of electron transport. thermodynamic neccessity for PSII regulation—'light stomata'. *Proceedings of the Royal Society London, Series B* **237,** 417–444.

Koffas MAG, Jung GY, Stephanopoulos GN. 2003. Engineering metabolism and product formation in *Corynebacterium*

glutamicum by coordinated gene expression. *Metabolic Engineering* **5,** 32–41.

Neuhaus H, Maaß U. 1996. Unidirectional transport of orthophosphate across the envelope of isolated cauliflower-bud amyloplasts. *Planta* **198,** 542–548.

Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO. 2003. Metabolic pathways in the post-genomic era. *Trends in Biochemical Sciences* **28,** 250–258.

Pettersson G, Ryde-Pettersson U. 1988. A mathematical model of the Calvin photosynthesis cycle. *European Journal of Biochemistry* **175,** 661–672.

Pettersson G, Ryde-Pettersson U. 1990. Model studies of the regulation of the Calvin photosynthesis cycle by cytosolic metabolites. *Biochimica et Biomedica Acta* **40,** 723–732.

Pfeiffer T, Sanchez-Valdenebro I, Nuno JC, Montero F, Schuster S. 1999. Metatool: for studying metabolic networks. *Bioinformatics* **15,** 251–257.

Poolman MG. 1999. *Computer modelling applied to the Calvin cycle*. PhD thesis, Oxford Brookes University.

Poolman MG, Fell DA. 2000. Modelling photosynthesis and its control. *Journal of Experimental Botany* **51,** 319–328.

Poolman MG, Fell DA, Raines CA. 2003. Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. *European Journal of Biochemistry* **270,** 430–439.

Poolman MG, Ölcer H, Lloyd JC, Raines CA, Fell DA. 2001. Computer modelling and experimental evidence for two steady states in the photosynthetic calvin cycle. *European Journal of Biochemistry* **268,** 2810–2816.

Portis Jr AR. 1982. Effects of the relative extrachloroplastic concentrations of inorganic phosphate, 3-phosphoglycerate, and dihydroxyacetone phosphate on the rate of starch synthesis in isolated spinach chloroplasts. *Plant Physiology* **70,** 393–396.

Riesmeier JW, Flügge U-I, Schulz B, Heineke D, Heldt H-W, Willmitzer L, Frommer WB. 1993. Antisense represion of the chloroplast triose phosphate translocator affects carbon partitioning in transgenic potato plants. *Proceedings of the National Academy of Sciences, USA* **90,** 6160–6164.

Rohwer JM, Botha FC. 2001. Analysis of sucrose accumulation in the sugar cane culm on the basis of *in vitro* kinetic data. *Biochemistry Journal* **358,** 437–445.

Schilling CH, Schuster S, Palsson BO, Heinrich R. 1999. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnology Progress* **15,** 296–303.

Schuster S, Dandekar T, Fell DA. 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology* **17,** 53–60.

Stelling J, Klamt S, Bettenbrock B, Schuster S, Gilles E. 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420,** 190–193. Supplemetary material available: Stel_02_suppl.pdf.

Stephanopoulos GN, Aristidou AA, Nielsen J. 1998. *Metabolic engineeering: principles and methodologies*. London, UK: Academic Press.

Woodrow I. 1986. Control of the rate of photosynthetic carbon dioxide fixation. *Biochimica et Biophysica Acta* **851,** 181–192.