UNRAVELLING THE INTERCONNECTIONS OF CELLULAR REGULATION

Johann M. Rohwer
October 2012

Unravelling the Interconnections of Cellular Regulation

Inaugural lecture delivered on 18 October 2012

Prof JM Rohwer
Department of Biochemistry
Faculty of Science

Editor: SU Language Centre
Printing: SUN MeDIA

ISBN: 978-0-7972-1396-8

About the Author

Johann Rohwer was born in Greytown on 25 May 1968 and grew up in the German settlement Hermannsburg in the KwaZulu-Natal Midlands. He received all his schooling at the Deutsche Schule Hermannsburg. After matriculating in 1985, a post-matric year (the German 'Abitur') at the Deutsche Höhere Privatschule in Windhoek followed.

From 1987 to 1989 he studied for a BSc degree at Stellenbosch University, majoring in biochemistry, chemistry and mathematics. This was followed by a BScHons and an MSc in biochemistry (all *cum laude*), the latter under the supervision of Prof Jannie Hofmeyr, studying the regulation of serine biosynthesis in *Escherichia coli*. For his doctoral studies, Johann went to the Netherlands and investigated the regulation of bacterial sugar transport under the supervision of Prof Hans Westerhoff and Dr Pieter Postma at the University of Amsterdam, graduating in April 1997.

Upon his return to South Africa in 1997, Johann took up a position as Senior Lecturer in the Department of Biochemistry at Stellenbosch University and has been with the department ever since. He was promoted to Associate Professor in 2002 and to Full Professor in 2011. Under his supervision and co-supervision fourteen MSc students and seven PhD students obtained their degrees, and he has co-authored 45 peer-reviewed articles in international journals.

Prof Rohwer was fortunate to spend two sabbaticals in overseas laboratories. During 2001, he visited Prof Philip Kuchel at the University of Sydney and learnt about applying the technique of NMR spectroscopy to study living cells in a non-invasive way. During 2008, he spent a year in Germany with his family as a research fellow of the Alexander von Humboldt Foundation, collaborating with Prof Mark Stitt at the Max Planck Institute of Molecular Plant Physiology on the modelling of central metabolism in plants.

Johann's research interests are computational and experimental systems biology, focusing on the analysis of the central metabolism of microbes and plants. He has received numerous awards, among others the Stellenbosch University Chancellor's Medal (1993), the President's Award from the South African National Research Foundation (2001), the Silver Medal of the South African Society of Biochemistry and Molecular Biology (2003), and the Vice Chancellor's Award for Excellent Research from Stellenbosch University (2010). He serves on the international STRENDA Commission and on the editorial boards of *BMC Systems Biology* and *Frontiers in Plant Systems Biology*. Johann is married to Christa and they have three children–Nicola (9), Saskia (7) and Martin (19 months).

Acknowledgements

No scientist works in isolation, and it is indeed a privilege to be standing on the shoulders of giants. My first and foremost appreciation goes to my scientific mentors, Jannie Hofmeyr and Hans Westerhoff, who introduced me to the field of systems biology long before it became the buzzword that it is today. All my colleagues at the Department of Biochemistry, I would like to thank you for your collegiality and friendship over the last fifteen years, and for making it the fun place to work that it is! I am particularly grateful for all the additional hours that you so unassumingly put in during my lengthy illness last year.

When Jacky Snoep joined our department as professor in 2000, our research group soon became known as the 'Triple-J' group (Jannie, Jacky and Johann). Indeed, it has been a privilege to work with you over the past 12 years; our common research interests mean that someone can always act as a sounding board for new ideas, and I remember many inspiring discussions. A special mention goes to Arrie Arends, our lab manager, for keeping the show on the road and the gears oiled.

It gives me great pleasure to acknowledge the contributions of all the MSc and PhD students who travelled with me on the often long and winding road that is scientific research: Otini Kroukamp, Arno Hanekom, Lafras Uys, Kristy Meyer, Brett Olivier, Garth Cronwright, Mandisi Mrwebi, Du Toit Schabort, James Dominy, Christiaan Crous, Justin Smith, Sandra Jordaan, Olona Gqwaka, Tim Akhurst, Wolfgang Schäfer, Sue Bosch, Christie Malherbe and Franco du Preez. To the current gang, I thank you for your contributions to date and wish you all the best with the final stretch: Theo van Staden, Carl Christensen, Darryl Michaels, Francois October, Athlee Maclear and Johann Eicher. To the three post-doctoral fellows who have spent time in my lab: Brett Olivier, Che Pillay and Julian Westoll, thank you for all your contributions. Che, I am particularly pleased that we can continue our collaboration now that your have your own group at UKZN. Brett, it's great to still hack on PySCeS together and watch it grow.

There is no space to acknowledge all the collaborations that are so essential to one's scientific career, but my two sabbatical hosts, Philip Kuchel and Mark Stitt, deserve special mention. Thank you for your hospitality, for giving me the freedom to explore new ideas, and for always being up to a discussion. To live and work in another country broadens the mind.

Finally, I want to thank my wife, Christa, for her unwavering love and support, especially during difficult times last year. To my kids, Nicola, Saskia and Martin, thank you for being such wonderful children and for your love. I know having an academic as a father is not always easy. . . Thank you to my parents, Christel and Rolf, Mutti und Vati, for your support and love over all these years. And thanks to my brothers Mark and Chris for being there.

Unravelling the Interconnections of Cellular Regulation

Johann M. Rohwer

1 Introduction

Ever since I studied for my BScHons degree at Stellenbosch University in 1990, when my then supervisor and now colleague Jannie Hofmeyr introduced me to the fields of metabolic control analysis and modelling of cellular pathways, the urge to formulate quantitative mathematical descriptions of biological processes has been the major driver of my scientific research activity. In a time when biologists all too often shied away from mathematics—and many still do today, although things have changed to some extent[1]—I took the unusual step of combining mathematics with biochemistry and chemistry as majors for my BSc degree. It is a decision I have never regretted. On the contrary, over the last two decades I have become convinced that mathematical formulations of biological processes can yield much additional information about how they work, and are indeed essential to their understanding.

This booklet will highlight some of these analyses from my own research career, which all fall into the field of what has been coined 'systems biology' during the last decade (although many of the seminal papers in the field are much older). Rather than following a chronological path, the narrative will be organised according to the functional hierarchy of living organisms, starting from the most fundamental unit of cellular activity, the enzyme, and building up to the regulatory networks these enzymes are involved in. Finally, possibilities of extending this approach to higher organisational levels are outlined. But first, the field of systems biology needs to be introduced briefly.

Systems biology The 20th century has brought huge advances in the fields of biochemistry and molecular biology; much of the cellular map has been elucidated, the components of life were separated, identified and their properties characterised. As such, this era of structure and function has resulted in a build-up of a huge archive of knowledge about the components of life. However, most of this information is hugely reductionist in nature. In many cases, the challenge today has shifted from component identification and characterisation towards assembling this disparate information into an integrated view at the molecular, cellular and organismal level. This has led to the emergence and rapid growth of the field of systems biology (e.g. Kitano, 2002), which aims to integrate the information into a "systems" view through a combination of interdisciplinary approaches that include mathematics and biophysics.

The sequencing of the first genome of an organism (the yeast *Saccharomyces cerevisiae*: Goffeau *et al.*, 1996) led to a rapid explosion of the field of bioinformatics, and today new genome sequences are commonplace. However, genome sequences are static information, and it is becoming widely appreciated that the computational approach to studying the *dynamic aspects* of cell processes is essential for understanding their function (Kitano, 2005). In addition to the computational systems biology endeavour, this requires additional experimenta-

[1] A remarkable and welcome change has been the recent introduction of an undergraduate BSc programme in biomathematics at Stellenbosch University.

tion, as the analysis has to be extended to the dynamic level, making time a central variable. In fact, it has been suggested that the integration of genetics, molecular biology and cell biology into the interdisciplinary field of systems biology will enable biology to advance to the next level (Kitano, 2002; Westerhoff & Palsson, 2004).

2 The enzyme as the unit of cellular networks

The observation by Eduard Buchner more than a century ago, that extracts of yeast cells were able to ferment glucose into ethanol (Buchner, 1897), can be viewed as having given birth to the field of biochemistry. As has by now been well established, the catalytic units responsible for this conversion are enzymes, a set of proteins in all living cells that are capable of each performing a chemical reaction with remarkable specificity and efficiency. Soon after Buchner's discovery, Leonor Michaelis and Maud Menten pioneered the field of enzyme kinetics (Michaelis & Menten, 1913), which aims to describe the dependence of the rate of an enzyme-catalysed reaction on the concentration of its reactants and products. Enzyme kinetics forms one of the pillars of systems biology, and as such is still highly relevant today. A brief overview follows.

2.1 Properties of enzymes

One of the defining features of an enzyme-catalysed reaction is that its rate does not increase indefinitely with increasing reactant concentration, in contrast to normal 'mass-action' chemical reactions (Figure 1). Michaelis and Menten explained this through the existence of an 'enzyme-substrate-complex' (the reactant of an enzyme-catalysed reaction is termed *substrate*); if the substrate molecules outnumber those of the enzyme, all enzymes will be occupied with substrate and any increase in substrate concentration will not increase the enzyme's rate. This phenomenon is called *saturation*.

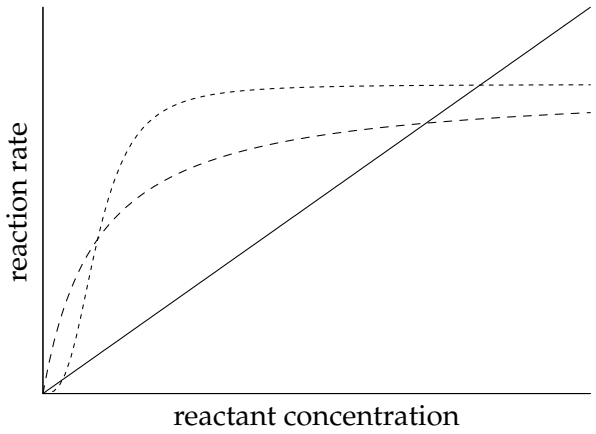

Figure 1: Reactant concentration dependence of the rate of a chemical (—), enzyme-catalysed (– –) and cooperative enzyme-catalysed (· · ·) reaction.

Note that enzyme kinetic profiles can have different shapes; some enzymes exhibit cooperative kinetics, leading to a sigmoidal, S-shaped rate-*vs*-substrate concentration curve as in Figure 1. Such enzymes can be 'switched on' (i.e. change from low to high rate) over a much narrower band of substrate concentrations than ordinary Michaelis-Menten-type enzymes.

The field of enzyme kinetics aims to describe the dependence of reaction rate on substrate concentration with a mathematical rate law (see Cornish-Bowden, 2012, for a good introductory text). The details will not be presented here; suffice it to say that every enzyme has a number of so-called kinetic parameters, which quantitatively describe its dynamic function: K_m, the Michaelis constant, is an indication of how strongly the enzyme binds to the substrate; V_max, the limiting rate, tell us 'how fast the enzyme can go'; and if an enzyme is inhibited by a compound, the inhibition constant K_i quantifies the strength of that inhibition.

Many enzyme mechanisms and their associated rate laws can be extremely complex: enzymes can have more than one substrate and product, they can be inhibited or activated by other compounds, and they can exhibit cooperative kinetics (Figure 1). As a result, the corresponding rate laws have a multitude of parameters, which can be difficult or laborious to determine experimentally. To overcome this, our

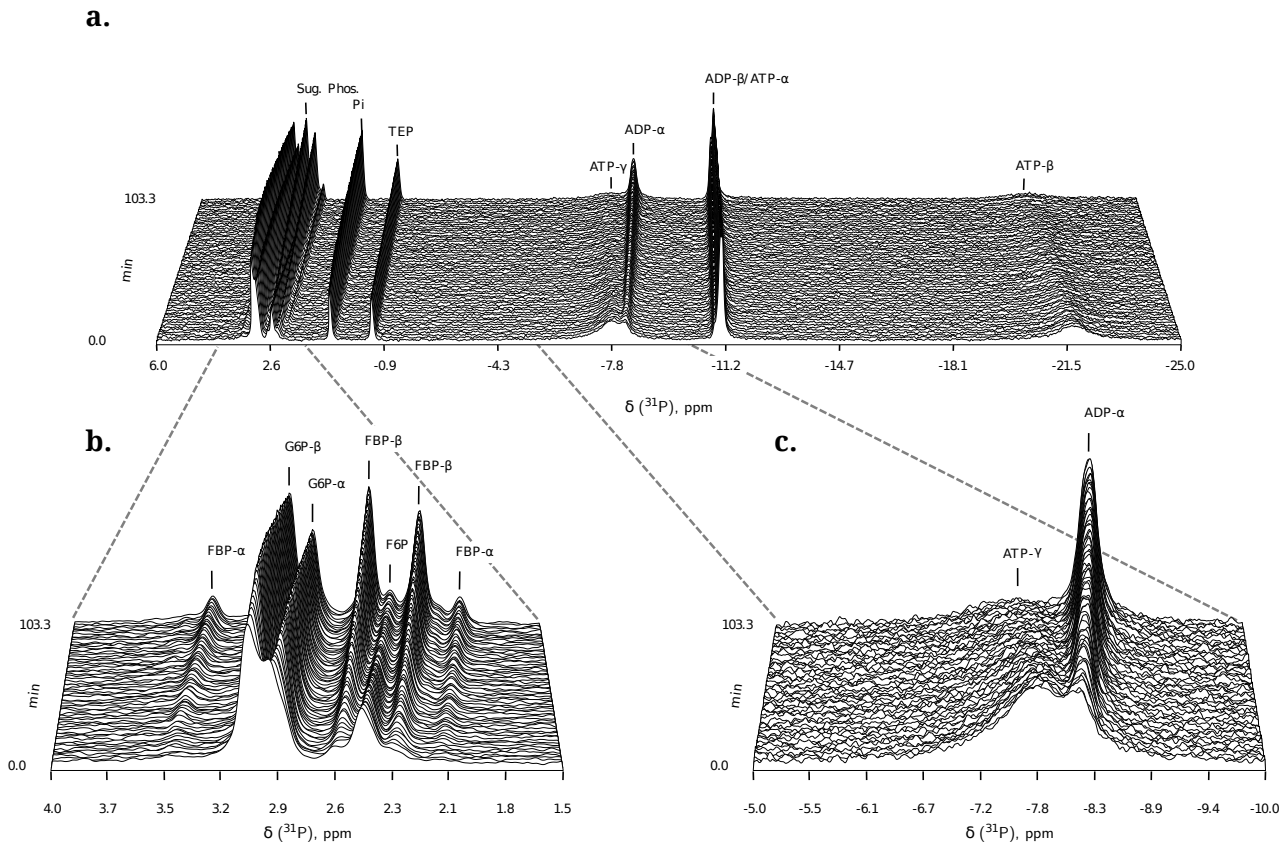

Figure 2: ^{31}P-NMR time course of a phosphofructokinase reaction. Peaks are labelled to identify the reactants. Successive NMR spectra are 130 s apart. **a.** Full NMR spectrum. **b.** Expansion of the sugar-phosphate region. **c.** Expansion of the nucleoside phosphate region.

group has developed generic equations for systems biology (Rohwer *et al.*, 2006, 2007), which have fewer kinetic parameters than the detailed mechanistic equations. These parameters are moreover defined operationally, so that they can be easily determined experimentally: the equations use half-saturation constants (i.e. the concentration of substrate or inhibitor giving half-maximal reaction rate or inhibitory effect, respectively) rather than intrinsic K_m values. Additionally, Rohwer & Hofmeyr (2010) have shown how to cleanly split and calculate the contributions of rate capacity, mass action and enzyme binding to the overall reaction rate, thus clearly delineating how enzyme catalysis modifies reaction rate and substrate/product dependence. This provides a method for quantitatively distinguishing enzyme-catalysed reactions from uncatalysed chemical ones.

2.2 Determining enzyme properties

Enzyme kinetic parameters are typically determined in an experiment where the concentration of one substrate or product is varied independently of the others, and the initial rate of the reaction measured (Cornish-Bowden, 2012) in a procedure termed an *enzyme assay*. If one of the substrates or products absorbs light at a particular wavelength, its concentration change over time (and thus the reaction rate) can be determined with a spectrophotometer. Alternatively, if none of the reactants absorb light, the reaction may be linked in one or more steps to a reaction whose reactant does absorb light. In all these cases, though, a large number of kinetic assays needs to be performed to obtain a complete picture of its function, especially in the case of multiple substrates and products and inhibition or activation by other compounds.

This limitation of classical enzyme assays

could in principle be overcome if the temporal change in multiple substrates and products could be measured simultaneously. We therefore developed a method (Eicher *et al.*, unpublished data) that captures information from the whole progression of the enzymatic reaction, rather than only the initial rate. NMR spectroscopy is used to quantify reaction substrates and products, and the reaction rate is estimated from the slopes of their time-course curves. Kinetic parameters are fitted by non-linear regression of these time courses, and can be obtained in much fewer runs than with classical initial-rate studies.

To illustrate the approach, Figure 2 shows an NMR time course of the phosphofructokinase reaction (G6P + ATP → FBP + ADP), an enzyme from the glycolysis fermentation pathway. Changes in peak area over time indicate changing substrate and product concentrations; these peaks are integrated to quantify the respective reactants, and from performing a total of six such experiments, the following rate law could be parametrised:

$$v = \frac{V_f \mathrm{f6p}^h \mathrm{atp}^h}{\left(\frac{1+\mathrm{pep}^h}{1+\alpha^{4h}\mathrm{pep}^h}\right) + \left(\frac{1+\alpha^{2h}\mathrm{pep}^h}{1+\alpha^{4h}\mathrm{pep}^h}\right)\left(\mathrm{f6p}^h + \mathrm{atp}^h\right) + \mathrm{f6p}^h\mathrm{atp}^h},$$

where f6p, atp and adp refer to the concentrations of F6P, ATP and ADP scaled by their half-saturation constants, V_f is the maximal rate, h is the Hill coefficient and α is the inhibition parameter for PEP. All these parameters could be accurately identified from the data.

3 Networks of enzymes

While enzymes have almost exclusively been studied in isolation using the methods described above, during the development of biochemistry over the last century, they almost never act in isolation inside the living cells. Enzymes are connected in networks and linked by chemical species that are broadly referred to as *metabolites*. Figure 3 shows an example network with enzymes indicated by numbered boxes and metabolites by subscripted letters.

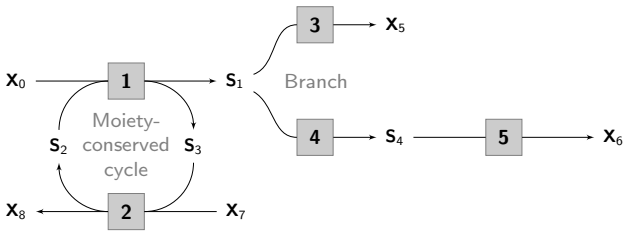

Figure 3: Example reaction scheme to illustrate a metabolic network. Terminal reactants and products (sources and sinks) are indicated by X, metabolic intermediates by S. Adapted from Hofmeyr (2001).

Although this is not a representation of a real network, it illustrates well how the network is hooked up.

Such networks are also termed *metabolic pathways*, and they are formed from the basic principle that the product of a particular reaction (say S_4, product of reaction 4) is at the same time the substrate of another reaction further 'down-stream' in the pathway (in this case reaction 5). When this network is 'running' (i.e. there is flow of matter through the pathway), S_4 will therefore simultaneously be produced by reaction 4 and consumed by reaction 5. Figure 3 also shows two other important features of metabolic pathways: (i) a branch point, where a particular metabolite (S_1 in this case) can act as substrate (or product) for two or more reactions; and (ii) a moiety-conserved cycle ($S_2 + S_3$ here), where the sum of a set of metabolite concentrations always remains constant, because when one of them is consumed, the other one is always produced in the same step, and *vice versa*.

So how are the pieces of the puzzle put together? The clue is in calculating, for each metabolite, how fast it is produced and how fast it is consumed. The net change in concentration of the metabolite over time will then be given by the difference between these two rates. So, in Figure 3, S_1 is produced by reaction 1 and consumed in reactions 3 and 4, so that its net rate of change will be given by the difference between the rate of reaction 1 (v_1 for short) and

$v_3 + v_4$. Mathematically, we can write

$$\frac{ds_1}{dt} = v_1 - (v_3 + v_4). \quad (1)$$

Equation (1) is termed an *ordinary differential equation*. Similar equations can be written for each metabolite of the network, and together this system of ordinary differential equations constitutes the *kinetic model* of the pathway (Hofmeyr, 2001; Palsson, 2006).

3.1 Computational analysis of the metabolic network

It is important to realise that each of the v_i terms in Equation (1) actually refers to an enzyme-catalysed reaction and can be described by an enzyme kinetic rate law (see Section 2.1). Together, this system of ordinary differential equations can thus be integrated to obtain the time-dependent evolution of the system from an initial state. Due to the highly non-linear nature of such rate laws, analytical solutions are usually impossible, and the result is obtained through numerical integration on the computer, for which many well-tested algorithms are available.

Metabolic systems, being open systems (i.e. they exchange matter and energy with their environment), frequently evolve to a so-called *steady state*. This state is characterised by metabolite concentrations that remain constant with time, but at a *finite flux* of matter through the system. This means that each metabolite is produced at exactly the same rate as at which it is consumed. Think of the analogy of a basin with a leaky plug. If I now adjust the flow of the tap to exactly match the rate at which water is leaking from the plug, then the level of water in the basin will remain constant, i.e. at steady state. Computationally, the steady state of a network can be calculated by setting the ordinary differential equations (see Equation (1)) equal to zero.

Even higher levels of analysis exist for such networks. One of these is the framework of metabolic control analysis, which was developed independently almost forty years ago by Kacser & Burns (1973) and Heinrich & Rapoport (1974). Metabolic control analysis is a kind of sensitivity analysis that aims to identify which enzymes in the network have the greatest control of the flux (i.e. the flow of matter through the pathway) or on the steady-state concentration of a metabolite. The degree of control is quantified by so-called control coefficients, which can take on a particular numerical value (see the sugarcane example in Section 3.2). While not fully appreciated during the initial decade after its publication, metabolic control analysis has made immense contributions to the study of cellular networks (Kacser & Burns (1973) has been cited 1249 times, Heinrich & Rapoport (1974) 900 times!). To name just one example, it has dispelled the dogma of a 'rate-limiting step', which has been (and still is being) held by many biochemists, and which states that a single enzyme step always determines the flux through a pathway and acts as 'pacemaker'. In contrast, metabolic control analysis states that flux control can be, and has been shown to be, shared between a number of steps, with all the control coefficients for a particular flux adding up to one (see Fell, 1996, for review).

I have recently reviewed in detail the kinetic modelling of metabolic pathways with a particular emphasis on plant metabolism, and the reader is referred to Rohwer (2012) for further information.

Software Generically, any system of coupled ordinary differential equations can of course be solved with standard numerical tools such as Mathematica or Matlab, or their open-source alternatives Octave and SciPy. However, simulation of these networks always involves a number of steps that have to be repeated: defining the network structure and kinetics, setting the parameters and initial conditions, performing an analysis (e.g. time-course simulation, steady state or metabolic control analysis), and visualising or storing the results. Consequently, a number of dedicated software tools have been developed, which simplify and automate many of these repetitive tasks. A standard model description language, the Systems Biology Markup Language or SBML (Hucka *et al.*, 2003), allows model interchange between tools

that can understand it.

In the early 2000s, our research group regularly made use of some of these existing software tools. During the development of new models, we were struck by the limitation that, while freely available, these tools were not open source and could thus not be modified or extended when they were lacking an important feature that we needed. This prompted the development of our own open-source tool, PySCeS, the Python Simulator for Cellular Systems (Olivier *et al.*, 2005). This program is still actively being developed and being used widely, both in our own research group and by others. (Box 1 highlights the importance of the open-source framework for my work.)

3.2 Examples of metabolic models

Bacterial PTS To conclude this section, two examples of kinetic models of metabolic pathways from my own work will be discussed briefly. The first stems from my doctoral research at the University of Amsterdam and concerns the bacterial phosphoenolpyruvate-dependent phosphotransferase system (PTS) in the bacterium *Escherichia coli*. The PTS is a complex of four enzymes, responsible for the uptake of glucose by the bacterium from its environment. Inside the bacterium, the glucose is then further fermented by the glycolysis pathway. In addition to sugar transport, the PTS is involved in the regulation of a number of other cellular processes, including the update of alternative carbon and energy sources. My focus was on the PTS itself, though, and prompted by the experimental observation (van der Vlag *et al.*, 1995) that the sum of the flux-control coefficients of the four PTS enzymes on glucose uptake did not add up to one as expected, but was in fact less, while theoretical analyses (van Dam *et al.*, 1993) had predicted that this sum could be as high as two due to the nature of the interconnections between the PTS enzymes. The PTS is a group-transfer pathway and a phosphate group is transferred from phosphoenolpyruvate sequentially along the enzymes to the glucose molecule as it is taken up by the cell. To understand this process better, I developed a kinetic model of the pathway based on measured kinetic data for the constituent enzymes (Rohwer *et al.*, 1998, 2000). The main conclusions from this model are three-fold:

1. The model predicted the existence of complexes between the PTS enzymes (these complexes could not be detected by direct experimental means).
2. The complexes drastically affect the behaviour of the PTS; when they are more

Box 1: Open source, open access

The decision to open-source our simulation tool PySCeS (Olivier *et al.*, 2005) has paid off numerous times. PySCeS is based on the Python programming language and its SciPy set of scientific libraries (Jones *et al.*, 2001–), themselves both open-source projects. Compared to commercial closed-source solutions, open-source alternatives have the following advantages:

- They are free, as in 'free beer'. They cost no money. This lowers the barrier to access and is particularly pertinent for developing countries like South Africa.
- They are free, as in 'free speech'. This empowers users and allows them to modify the source code. We have incorporated a number of mathematical algorithms into PySCeS, which were previously not available in SciPy.
- They are based on a spirit of reciprocation and sharing. Open-source software developers are generally keen to test code and contribute improvements; this fosters collaboration.

The decision by ever-increasing numbers of scholarly journals to follow an open-access publication model should be viewed in a similar favourable light.[2]

[2]The contributions of Stellenbosch University to open access through its SUNScholar repository and by hosting the Berlin10 Open Access Conference in December 2012 are welcomed.

prevalent, the sum of the PTS enzyme flux-control coefficients decreases (Rohwer *et al.*, 2000). Moreover, they are more prevalent under *in vivo* conditions than in the test tube, due to macromolecular crowding in the intracellular environment (Rohwer *et al.*, 1998).

3. The model explains why we sometimes observed a quadratic dependence of the PTS flux on the total enzyme concentration (Rohwer *et al.*, 1998), in contrast to a linear dependence that is common for metabolic pathways. The origin of the discrepancy lies in the way the PTS enzymes react with each other: being a group-transfer pathway, the phosphate transfer reactions are bimolecular, and if no significant complex formation occurs, this can lead to a greater-than-linear flux response.

Sugarcane The second example deals with the modelling of sucrose (cane sugar) accumulation in sugarcane (Figure 4). Sugarcane is remarkable for its ability to accumulate sucrose to concentrations higher than those in most other plants, which makes it the major agricultural crop for producing table sugar. However, this accumulation ability differs between varieties of sugarcane, and understanding the factors that control the extent to which sucrose is accumulated may aid in developing strategies for optimising agricultural yields. Sucrose metabolism in the sugarcane stalk is characterised by concurrent sucrose breakdown and re-synthesis (Komor, 1994; Zhu *et al.*, 1997), a process which has been termed 'futile cycling'. This process is considered to be energetically wasteful (due to the expenditure of cellular energy during the re-synthesis reactions) and it was reasoned that decreased futile cycling should lead to an increase in sucrose accumulation. To investigate this problem further, we constructed a detailed kinetic model of the central sucrose metabolism, its futile cycling, and its accumulation in the stalk (Rohwer & Botha, 2001) .

Based on metabolic control analysis, the control coefficient of each reaction on the futile cycling of sucrose (defined as the ratio between

Figure 4: Cut sugarcane. By Rufino Uribe (caña de azúcar) [CC-BY-SA-2.0], via Wikimedia Commons.

sucrose breakdown and sucrose accumulation) was calculated (Figure 5). The five reactions with the numerically largest control coefficients were the uptake of fructose and glucose into the cells (-0.86 and -0.90 respectively), the transport of sucrose into the storage compartment (-0.51), phosphorylation of glucose (1.09) and breakdown of sucrose (0.71). The model calculations showed that a decrease in futile cycling should translate into increased sucrose accumulation (Rohwer & Botha, 2001). On the basis of these results and the negative values of the control coefficients, we predicted that an *increase* in the transport proteins for glucose and fructose across the cell membrane, as well as of the protein responsible for transporting sucrose into the storage compartment, should favour sucrose accumulation. Conversely, *decreasing* the enzymes that break down sucrose and phosphorylate glucose should reduce futile cycling (their control coefficients are positive, Figure 5). The model thus identified these steps as the most promising biotechnological targets for reducing futile cycling of sucrose and increasing sucrose accumulation.

By way of experimental validation, Rossouw *et al.* (2007) demonstrated an increase in sucrose accumulation in sugarcane suspension cells by experimentally decreasing neutral invertase activity (the enzyme that breaks down sucrose), albeit at the expense of reduced respiration and growth. Thus, while the increase in sucrose accumulation was correctly predicted by the model, its scope did not extend far enough to predict the effect of changes

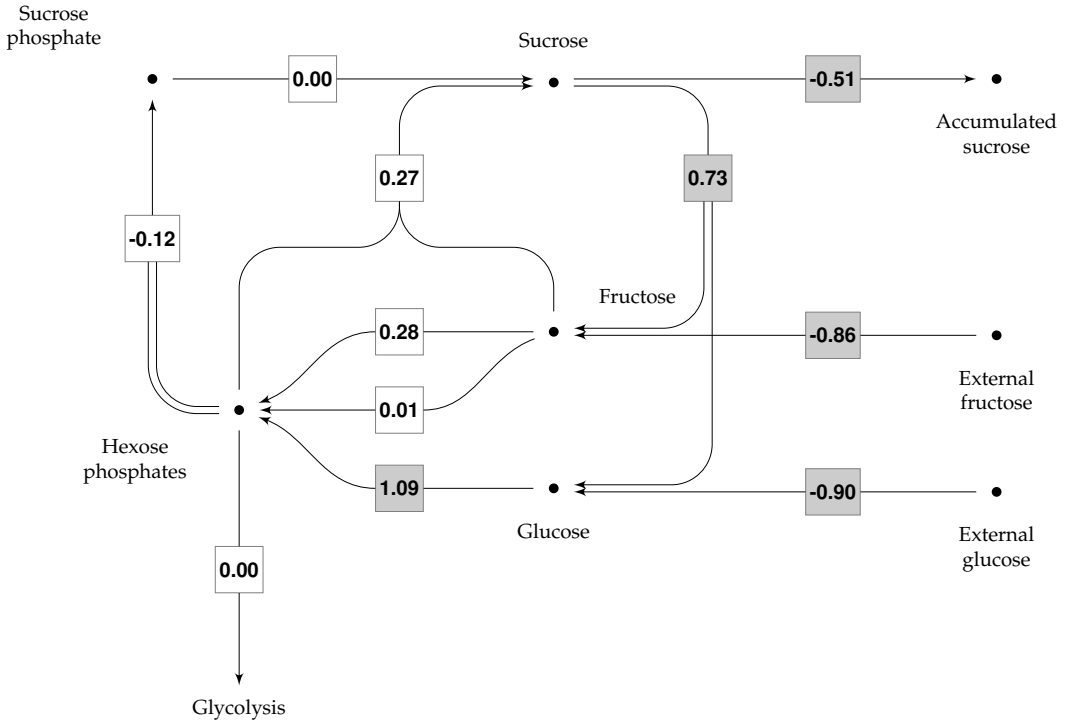

Figure 5: Sucrose futile cycling control coefficients as calculated by the sugarcane kinetic model (Rohwer & Botha, 2001). Metabolites are indicated by a labelled black dot, enzymes by a box containing the numerical value of the control coefficient. The steps with the numerically largest control coefficients are highlighted in grey.

in neutral invertase activity on respiration and growth. More recently, these results were repeated in whole transgenic sugarcane plants, with Rossouw *et al.* (2010) demonstrating increased sucrose and decreased hexose levels, as well as reduced futile cycling in the stalk tissues of plants with reduced neutral invertase activity.

The sugarcane stalk is divided into nodes and internodes (Figure 4), with the younger internodes at the top. Sucrose is accumulated preferentially in the older mature internodes at the bottom of the stalk. The original model described above was built to simulate medium-mature tissue. To gain a better understanding of the metabolic changes that occur during the maturation of the stalk tissue, this model was extended to 'model a stalk in segments' by substituting experimentally determined data on how the levels of all the pathway enzymes change in the various internodes (Uys *et al.*, 2007). This provided a more comprehensive view of the maturation process but still had significant shortcomings, as discussed below.

4 Towards the organism scale: networks of networks

The model of Uys *et al.* (2007) described above distinguished between internodes purely by substituting enzyme activity data for each internode. This is a simplification, as in reality the internodes are linked by the phloem (i.e. the plant connective tissue that transports carbohydrates—mainly sucrose—from the leaves, where they are synthesised during photosynthesis, via the stalk to the roots). The internodes are thus not independent, but linked by mass flow. Moreover, the phloem is a separate compartment from the cytosol (the main intracellular compartment) and the vacuole (the main storage compartment in plant cells). Solute transport between these compart-

ments therefore also needs to be accounted for.

To overcome these shortcomings, we developed a method for incorporating phloem flow into the sugarcane model. The movement of solutes through the phloem has been extensively studied by Thompson and co-workers, who have also developed a mathematical model of the process (reviewed in Thompson, 2006). The basis of this formulation is the establishment of an osmotic potential due to an unequal distribution of solute in the phloem tubes. This causes an osmotic pressure gradient (termed *turgor*) that can push fluids up or down.

The challenge was now to combine this mathematical formulation of phloem flow with our existing kinetic model of metabolic reactions. This was achieved by

- considering the phloem, apoplast (extracellular space), symplast (linked cytosols of adjacent cells) and vacuole as separate compartments;
- considering the stalk as a one-dimensional object, discretising its length into a fixed number of finite volumes and defining the above compartments on each of these volumes;
- adding the metabolic reactions from our previous kinetic model; and
- including steps that transport metabolites between compartments.

The distribution of sucrose throughout the stalk was then modelled with the following mathematical formalism (Uys *et al.*, in preparation; Uys, 2009):

$$\frac{\partial s}{\partial t} + \frac{\partial}{\partial z}(\vec{u}\, s) + \frac{\partial}{\partial z}\left(D_s \frac{\partial s}{\partial z}\right) = v_{rx} + v_{tr} \qquad (2)$$

which is illustrated here for only one metabolite S (with concentration s). Similar equations were constructed for all metabolites in all compartments. Time is indicated by t, z is the distance along the length of the stalk, $\vec{u}$ is a vector of velocity of fluid flow for the particular compartment, D_s is the diffusion coefficient of S, v_{rx} is the sum of all metabolic reactions in which S participates, and v_{tr} is the sum of all inter-compartmental transport processes involving S. Equation (2) is called a *partial differential equation*, and simply stated, from left to right

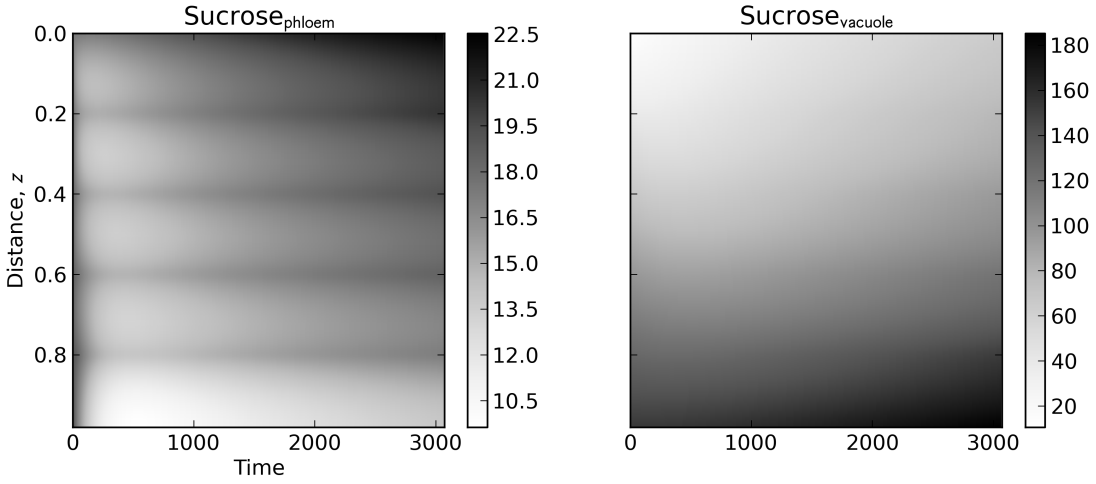

Figure 6: Time-evolution of sucrose concentrations in the phloem and vacuole (storage compartment) in an advection-diffusion-reaction model of a sugarcane stalk (Uys, 2009). The x-axis shows time. The y-axis shows the length of the stalk from 0 (young tissue) to 1 (older tissue). Leaves are attached at the internodes at positions 0, 0.2, 0.4, 0.6 and 0.8 on the stalk. Sucrose concentrations are indicated by a grey-scale value according to the corresponding bar on the right-hand side of each figure.

the terms mean that the concentrations change with time, that metabolites may undergo transport by a fluid—in this case the phloem sap—due to the fluid's bulk motion in a particular direction (the technical term is *advection*), that species may diffuse, and lastly that they engage in metabolic reactions or are transported between compartments. The set of equations is termed an *advection-diffusion-reaction* system.

The advection-diffusion-reaction framework was implemented in a model comprising five node-internode pairs. The system was simulated with another Python-based open-source software, FiPy (Guyer *et al.*, 2009). The pathway used made a number of simplifying assumptions, as the model aimed to capture the essential features of the system but was not based in detail on experimentally determined data. Such 'core' or 'toy' models can often yield biologically relevant results, in spite of the simplified structure and default parameters used. In our case (Figure 6), the concentration of sucrose in the phloem was highest at the nodes due to phloem 'loading' (import of sugar from the attached leaves), but then spread out in both directions to create 'sawtooth' profiles. Importantly, sucrose could also be transported back up the stalk from the leaf attachment point, due to the advective flow in the phloem. Also, the 'filling-up' behaviour of the stalk could be observed; sucrose in the storage compartment (vacuole) was accumulated at first preferentially in the more mature internodes, but gradually the younger internodes also progressively accumulated sucrose (Uys, 2009).

Parameter sensitivities: FAST For any mathematical model, a crucial question is how sensitive the model outputs are to changes in the values of its parameters. For kinetic models based on ordinary differential equations at steady state, this can be achieved with metabolic control analysis (Section 3.1). However, the advection-diffusion-reaction framework cannot be analysed with metabolic control analysis as it is not at steady state and moreover requires the use of partial differential equations. The model can be analysed, however, with a method widely applied in the engineering sciences, called the Fourier Amplitude Sensitivity Test (FAST, Cukier *et al.*, 1978). This is a global sensitivity analysis method that calculates the concomitant effect of changes in all model parameters on its output by allowing these parameters to oscillate at uniquely assigned frequencies, resulting in oscillating model output. Since the parameter frequency is known, its contribution to the model variation can be isolated using some properties of a Fourier transform, which converts the time domain back to the frequency domain.

By way of example, we used the Fourier Amplitude Sensitivity Test to isolate the steps with the largest contribution to the difference between the rate of appearance and rate of disappearance of sucrose in the cytosolic compartment (i.e. 'futile cycling' of sucrose). Only four steps had any significant effect on this quantity: the uptake of sucrose into the cytosol from the phloem, the synthesis of sucrose by the enzyme sucrose synthase, and the breakdown of sucrose by two enzymes, i.e. neutral invertase in the cytosol and acid soluble invertase in the vacuole (Uys, 2009). Note that this is only one example, and the Fourier Amplitude Sensitivity Test may of course be used to calculate the parameter sensitivity of any model output that the experimenter may be interested in.

5 Unravelling the interconnections

The number of published kinetic models of cellular pathways increases weekly, as any inspection of online model databases such as JWS Online (Olivier & Snoep, 2004) or BioModels (le Novère *et al.*, 2006) will reveal. Moreover, these models are increasing in size and complexity. Most recently, Karr *et al.* (2012) published a whole-cell computational model of the human pathogen *Mycoplasma genitalium*, accounting for the function of every annotated gene and aiming to describe its life cycle from the level of interactions of individual molecules. One problem with increasing model size is that the level of complexity of the models approaches that of the organisms they

aim to describe. While such models provide powerful tools that are often more accessible to query and interrogation than experimental systems, without proper frameworks of analysis they remain little more than collections of data, albeit big and comprehensive.

In the framework of supply-demand analysis, Hofmeyr & Cornish-Bowden (2000) have posited that the metaphor of an economy of supply and demand may prove useful to understand the regulation of cellular pathways. One of its tenets is that flux control of a pathway may lie beyond the boundaries of what has traditionally been considered the pathway, i.e. in the demand for its end-product. One of the problems with supply-demand analysis, however, is that the complexity of large models may preclude us from finding a 'natural' subdivision of the system into supply and demand blocks, and that it is therefore difficult to apply the analysis in an unbiased way (Hofmeyr & Rohwer, 2011).

To overcome this limitation, Rohwer & Hofmeyr (2008) have generalised supply-demand analysis so that it can be applied to models of arbitrary size and complexity in a systematic, computer-driven way. In essence, the approach works by 'clamping' or fixing each of the model's variable metabolites in turn, thus changing it into a parameter of the system. Its value is then altered in a range above and below the reference. Concomitantly, the dependences of all the rates that produce the metabolite (the *supply reactions*) and all the rates that consume it (the *demand reactions*) on the concentration of the metabolite are then graphed in what is called a combined rate characteristic. Because this is done for every metabolite of the system, the approach is unbiased.

Generalised supply-demand analysis now compares, for each of the supply and demand fluxes, how sensitively the individual enzyme that interacts directly with the clamped metabolite responds to changes in its level. This is compared to how the whole supply (or, respectively, demand) flux responds to the metabolite. This comparison yields important information about how the pathway is regulated, and can identify:

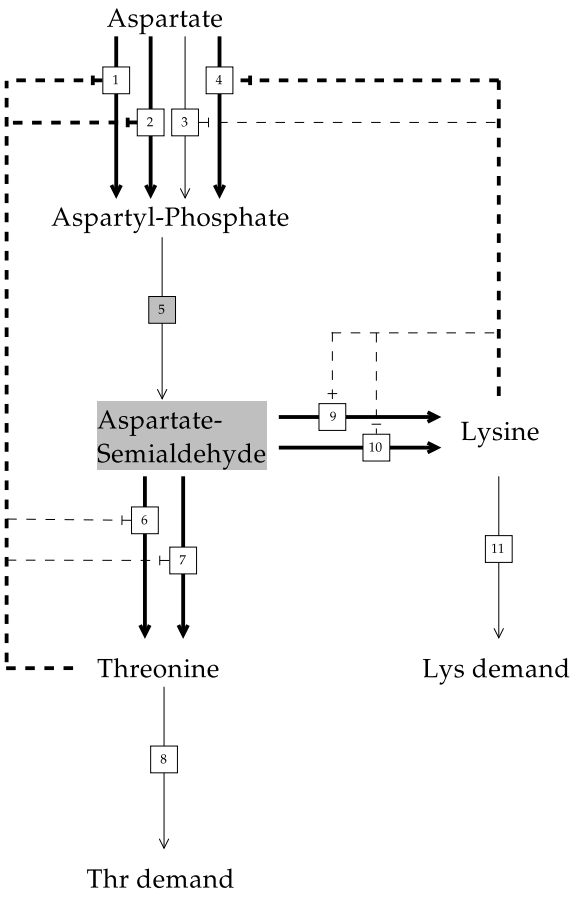

Figure 7: Aspartate metabolism in *Arabidopsis thaliana*. Enzymes are in numbered boxes. The pathway products threonine and lysine exert negative feedbacks on enzymes in their synthesis pathways (dashed lines). Bold lines indicate major routes of interaction (see main text). Adapted from Curien *et al.* (2009).

- potential sites of regulation;
- regulatory metabolites;
- the quantitative relative contribution of different routes of interaction from a metabolite to a supply or demand block; and
- sites of functional differentiation where one of the supply or demand blocks predominantly controls the flux, and the other determines to what extent the linking metabolite is buffered against changes in its concentration.

We have developed a module for the PySCeS software that can automatically perform generalised supply-demand analysis on a model by doing the computations and graphing the results.

By way of example, consider the pathway in Figure 7, which shows the metabolism of aspartate (an amino acid) to threonine and lysine (two other amino acids) in the thale cress (*Arabidopsis thaliana*), which is the model plant for botanists and plant physiologists. The pathway has a number of interesting features. First, a number of reactions are catalysed by more than one enzyme (for the first reaction, there are actually four enzymes!). These multiple enzymes are called *isozymes* and have different properties. Second, numerous feed-back loops exist from the end-products lysine and threonine to enzymes in their synthesis pathways, both to the very first enzymes and to the first enzymes in their dedicated branch.

To better understand the function of these isozymes and feed-back loops, Curien *et al.* (2009) constructed a detailed model of the pathway in Figure 7 (their model has a number of additional reactions, which have been omitted from Figure 7 for clarity). Two of the main conclusions of their work are that, first, some of the feed-back loops serve to maintain independence between fluxes in competing pathways (such as the branches to threonine and lysine), and second, that isozymes are not redundant, because they contribute unequally to the regulation of the flux through the pathway.

We performed a generalised supply-demand analysis on this model to obtain additional information on the function of the pathway and its regulation. Only one result is presented here. Consider the analysis around aspartate-semialdehyde, and specifically its interaction with the supply enzyme 5 (its biochemical name is aspartate-semialdehyde dehydrogenase). Both are shaded grey in Figure 7. The analysis has shown that the major route of communication of aspartate-semialdehyde with its supply enzyme is not the direct interaction; rather all the major routes run via the demand enzymes and feedback loops (indicated in bold), and for lysine, only one of the feedback loops (to enzyme 4) plays a significant role. This was an unexpected result, as the regulatory routes follow a 'detour'. It also shows the use of generalised supply-demand analysis in quantifying the relative contribution of routes of interaction (third bullet point above). Note that the importance of every route can be quantified with a number, thus ranking the various routes of interaction (not shown in Figure 7).

As pointed out by Hofmeyr & Cornish-Bowden (2000), the mere existence of a feedback loop does not mean that it is always active, and the regulation can sometimes follow another route. Generalised supply-demand analysis thus identify active regulatory routes for a *particular set of conditions*, which depend on the particular state of the model and cannot be inferred from model structure alone. The strength of the approach lies in that it provides a tool for the systematic functional analysis of large models, yielding an entry point for further refined analyses, where the modeller can zoom in to those parts that show interesting regulatory behaviour.

6 Conclusion

In this essay I have outlined some of the experimental and computational tools that can be used to unravel the complex networks of cellular regulation. The field of systems biology has drawn attention to the fact that these networks cannot be understood on the basis of structure alone, but that dynamic aspects and the interactions between the components play a crucial role.

The enzyme as the central unit of catalysis was introduced in Section 2, and techniques to study its properties were outlined. The interaction of enzymes in metabolic networks, the analysis of such networks with computational models, and the additional levels of complexity that come with it were discussed in Section 3. Two examples from my own research were discussed in greater detail. A framework to extend the network analysis to the organism scale in plants was introduced in Section 4, by including the physico-chemical properties of solute transport along the plant stalk. Fi-

nally, Section 5 introduced generalised supply-demand analysis as a computational method for unravelling regulatory interconnections in large models of cellular pathways.

The analysis has thus followed a bottom-up approach, from small to big, to arrive at an overall picture of how such a system functions dynamically. Importantly, though, every step in this process is necessary, and we cannot simply forget about the properties of individual enzymes, for example, if we are to understand the regulatory function of the whole network.

As summarised in Rohwer (2012), I believe the future challenges to lie in the integration of different levels of the cellular hierarchy (such as gene transcription, protein synthesis and signalling) with the metabolic level. Also, new methods have to be developed to extract information from the huge datasets generated by current 'omics' techniques, that can measure RNA, protein, enzyme activity or metabolite levels on a system-wide scale. Following through to the supra-cellular level, approaches are needed for integrating metabolic and cellular regulation with organismal regulation, and providing links to (plant and animal) physiology as well as ecology.

In conclusion, returning to the importance of mathematical analysis, the tenet of my work has been and will continue to be that a thorough understanding of the molecular physiology of cellular systems requires a multidisciplinary approach comprising quantitative experimental analysis and numerical simulation, all within a rigorous theoretical framework. This can result in the formulation of predictive models that will enhance our ability to understand organisms and manipulate them in a directed, targeted way (e.g. in biotechnology or medicine), and holds a great future for biology and physiology.

References

Buchner E. 1897. Alkoholische Gährung ohne Hefezellen. *Berichte der Deutschen Chemischen Gesellschaft* **30**, 117–124.

Cornish-Bowden A. 2012. *Fundamentals of Enzyme Kinetics*. Weinheim, Germany: Wiley-Blackwell, 4th edn.

Cukier RI, Levine HB, Shuler KE. 1978. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics* **26**, 1–42. Paper copy.

Curien G, Bastien O, Robert-Genthon M, Cornish-Bowden A, Cárdenas ML, *et al.* 2009. Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters. *Molecular Systems Biology* **5**, 271.

Fell DA. 1996. *Understanding the Control of Metabolism*. London: Portland Press.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, *et al.* 1996. Life with 6000 genes. *Science* **274**, 546.

Guyer JE, Wheeler D, Warren JA. 2009. FiPy: Partial differential equations with Python. *Computing in Science and Engineering* **11**, 6–15.

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.

Hofmeyr JHS. 2001. Metabolic control analysis in a nutshell. In: **Yi TM, Hucka M, Morohashi M, Kitano H**, eds., *Proceedings of the 2nd International Conference on Systems Biology*. Madison, WI: Omnipress, 291–300.

Hofmeyr JHS, Cornish-Bowden A. 2000. Regulating the cellular economy of supply and demand. *FEBS Letters* **476**, 47–51.

Hofmeyr JHS, Rohwer JM. 2011. Supply-demand analysis: a framework for exploring the regulatory design of metabolism. *Methods in Enzymology* **500**, 533–554.

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, *et al.* 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531.

Jones E, Oliphant T, Peterson P, *et al.* 2001–. SciPy: Open source scientific tools for Python.

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

Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, *et al.* 2012. A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389–401.

Kitano H. 2002. Computational systems biology. *Nature* **420**, 206–210.

Kitano H. 2005. International alliances for quantitative modeling in systems biology. *Molecular Systems Biology* **1**, 2005.0007.

Komor E. 1994. Regulation by futile cycles: the transport of carbon and nitrogen in plants. In: **Schulze ED**, ed., *Flux Control in Biological Systems*. San Diego: Academic Press, 153–201.

le Novère N, Bornstein B, Broicher A, Courtot M, Donizelli M, *et al.* 2006. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research* **34**, D689–D691.

Michaelis L, Menten M. 1913. Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift* **49**, 333–369.

Olivier BG, Rohwer JM, Hofmeyr JHS. 2005. Modelling cellular systems with PySCeS. *Bioinformatics* **21**, 560–561.

Olivier BG, Snoep JL. 2004. Web-based kinetic modelling using JWS Online. *Bioinformatics* **20**, 2143–2144.

Palsson BO. 2006. *Systems Biology—Properties of Reconstructed Networks*. New York: Cambridge University Press.

Rohwer JM. 2012. Kinetic modelling of plant metabolic pathways. *Journal of Experimental Botany* **63**, 2275–2292.

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

Rohwer JM, Hanekom AJ, Crous C, Snoep JL, Hofmeyr JHS. 2006. Evaluation of a simplified generic bi-substrate rate equation for computational systems biology. *IEE Proceedings—Systems Biology* **153**, 338–341.

Rohwer JM, Hanekom AJ, Hofmeyr JHS. 2007. A universal rate equation for systems biology. In: **Hicks MG, Kettner C**, eds., *Experimental Standard Conditions of Enzyme Characterizations. Proceedings of the 2nd International Beilstein Workshop*. Frankfurt: Beilstein-Institut zur Förderung der Chemischen Wissenschaften, 175–187.

Rohwer JM, Hofmeyr JHS. 2008. Identifying and characterising regulatory metabolites with generalised supply-demand analysis. *Journal of Theoretical Biology* **252**, 546–554.

Rohwer JM, Hofmeyr JHS. 2010. Kinetic and thermodynamic aspects of enzyme control and regulation. *Journal of Physical Chemistry B* **114**, 16280–16289.

Rohwer JM, Meadow ND, Roseman S, Westerhoff HV, Postma PW. 2000. Understanding glucose transport by the bacterial phosphoenolpyruvate:glycose phosphotransferase system on the basis of kinetic measurements *in vitro*. *Journal of Biological Chemistry* **275**, 34909–34921.

Rohwer JM, Postma PW, Kholodenko BN, Westerhoff HV. 1998. Implications of macromolecular crowding for signal transduction and metabolite channeling. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 10547–10552.

Rossouw D, Bosch S, Kossmann J, Botha F, Groenewald JH. 2007. Downregulation of neutral invertase activity in sugarcane cell suspension cultures leads to a reduction in respiration and growth and an increase in sucrose accumulation. *Functional Plant Biology* **34**, 490–498.

Rossouw D, Kossmann J, Botha F, Groenewald J. 2010. Reduced neutral invertase activity in the culm tissues of transgenic sugarcane plants results in a decrease in respiration and sucrose cycling and an increase in the sucrose to hexose ratio. *Functional Plant Biology* **37**, 22–31.

Thompson MV. 2006. Phloem: the long and the short of it. *Trends in Plant Science* **11**, 26–32.

Uys L. 2009. Coupling kinetic models and advection-diffusion equations to model vascular transport in plants, applied to sucrose accumulation in sugarcane. Ph.D. thesis, Stellenbosch University.

Uys L, Botha FC, Hofmeyr JHS, Rohwer JM. 2007. Kinetic model of sucrose accumulation in maturing sugarcane culm tissue. *Phytochemistry* **68**, 2375–2392.

van Dam K, van der Vlag J, Kholodenko BN, Westerhoff HV. 1993. The sum of the control coefficients of all enzymes on the flux through a group-transfer pathway can be as high as two. *European Journal of Biochemistry* **212**, 791–799.

van der Vlag J, van't Hof R, van Dam K, Postma PW. 1995. Control of glucose metabolism by the enzymes of the glucose phosphotransferase system in *Salmonella typhimurium*. *European Journal of Biochemistry* **230**, 170–182.

Westerhoff HV, Palsson BO. 2004. The evolution of molecular biology into systems biology. *Nature Biotechnology* **22**, 1249–1252.

Zhu YJ, Komor E, Moore PH. 1997. Sucrose accumulation in the sugarcane stem is regulated by the difference between the activities of soluble acid invertase and sucrose phosphate synthase. *Plant Physiology* **115**, 609–616.