A mathematical model of the tetrapyrrole biosynthesis pathway

Problem presented by: Alison Smith and Michael Moulin (Cambridge).
Report by: John Ward (Loughborough), Michael Moulin (Cambridge), Francois Feugier (Newcastle) and Saul Hazledine (John Innes Centre).
Study group contributors: Alistair Middleton, Eleanor Norris, Markus Owen, Jonathan Wattis (all Nottingham) and Ed Green (Columbia, USA).

Abstract

The tetrapyrrole biosynthesis pathway is a key part in chlorophyll production and is essential for plant survival. It involves numerous interacting compounds and, crucially, light. The understanding of the complex regulation processes involved has been the focus of extensive experimental research providing a large source of data. A particular set of data, concerned with the modelling described in this report, involves 24 hour timecourse data from seedlings exposed to constant light, following a three day period of growth from seed in darkness. This data includes the levels of key components such as chlorophyll, ATP, chlorophyllide and protochlorophyllide. Amongst the questions posed in the study-group were: i) Can the timecourse data be predicted by a model? ii) Can it predict the differences in levels of various components in found mutant strains.

To address these questions, we present in this report a model consisting of a coupled system of nonlinear ODEs that describes a simplified version of the tetrapyrrole pathway based on mass action laws. Model simulations produced results that agree qualitatively well with most, but not all, of the available timecourse data obtained from wild-type and mutant strains. Nearly all of the model's parameters are not known, so the values used in these simulations are based on estimates of the relative timescales of the reactions. An attempt at improving these estimates using data fitting techniques is also discussed.

1 Introduction

Tetrapyrroles are the most abundant pigment molecules on the planet, and chlorophyll is the only biological molecule visible from space. Chlorophyll is present only in photosynthetic organisms whereas haem (a key compound in tetrapyrrole synthesis) is universally found in all organisms, including bacteria, fungi, plants and animals. Plants also contain two other tetrapyrroles, sirohaem and phytochromobilin. These molecules are cofactors for many proteins involved in different plants functions such as light harvesting, photoreception or in nutrient assimilation and detoxification. The tetrapyrroles are synthesised by a common branched pathway which can be divided into four parts, consisting of a main branch, that serves to provide the intermediates for the other three branches, and branches involved in the synthesis of sirohaem, chlorophyll and haem.

1.1 Background on the regulation of the pathway

The compounds sirohaem, chlorophyll and haem are all essential for plant survival, but the amount of each compund the plant requires varies greatly. This results in competition for

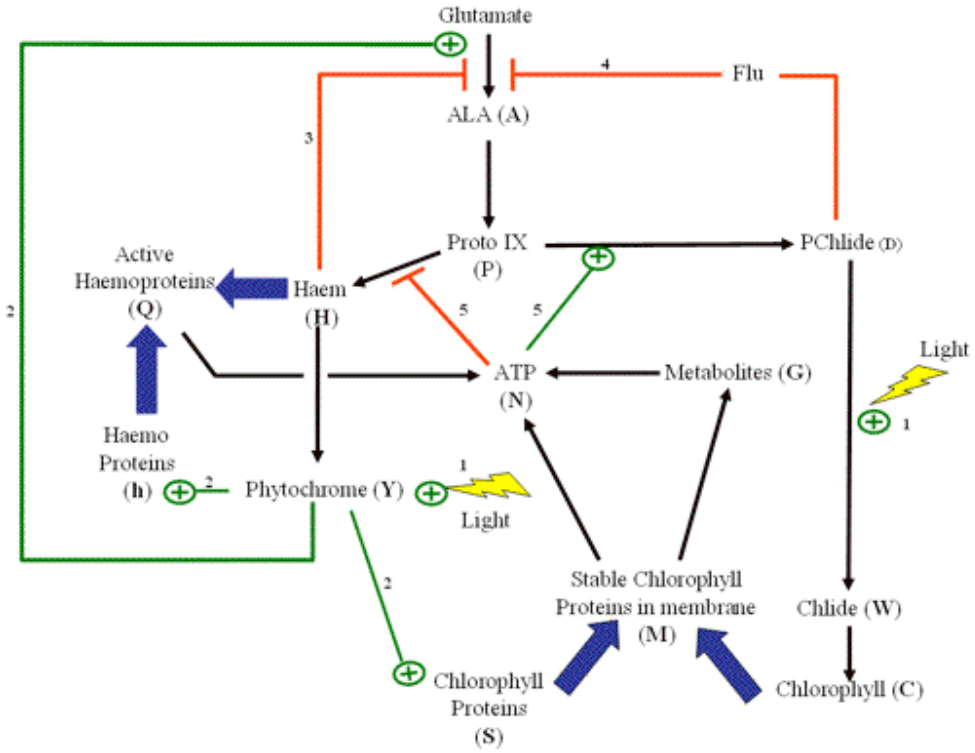


Figure 1: A Schematic diagram of the tetrapyrrole pathway and regulation processes, highlighting the role of each of the components in the system that feature in the modelling. The variable names of each component is indicated. Note we will adopt the commonly used shortened names Pchlide and Chlide for protochlorophyllide and chlorophyllide, respectively, throughout this report.

common intermediates of the pathway during their synthesis. One good example is the competition, which takes place between haem and chlorophyll. Haem (produced via synthesis, Figure 1) is essential for the plants to perceive light whereas chlorophylls is the key molecule for the conversion of photons from the sun into useable energy that can fuel plant growth. In addition, the flux of the chlorophyll branch is probably around 1000 times greater than the haem branch making the competition even more pronounced between these two branches [1].

Because of this competition between all end-products, several regulatory mechanisms have been demonstrated or proposed. The most characterised is light. Seedlings grown in complete darkness are not able to make any chlorophyll, because of the requirement of one of the enzymes, protochlorophyllide reductase (POR) for light in its enzymic reaction. Conversely, when seedlings are put in the light this induces the conversion of protochlorophyllide to chlorophyllide and then chlorophyll (Fig. 1, path labelled (1); henceforth, we will denote this using Fig. 1(1)). In addition light induces a signal cascade mediated by a photoreceptor (phytochrome) synthesised by the haem branch (Fig. 1(1)) [4]. This results in co-ordinated synthesis of the light-harvesting complex proteins, chlorophyll-binding proteins, and also the synthesis of haemoproteins involved in the photosynthesis apparatus (Fig. 1(2)) [8].

A number of tetrapyrrole intermediates are phototoxic, accumulation causes necrotic lesions, further complicating the regulation of this pathway. To circumvent the production of

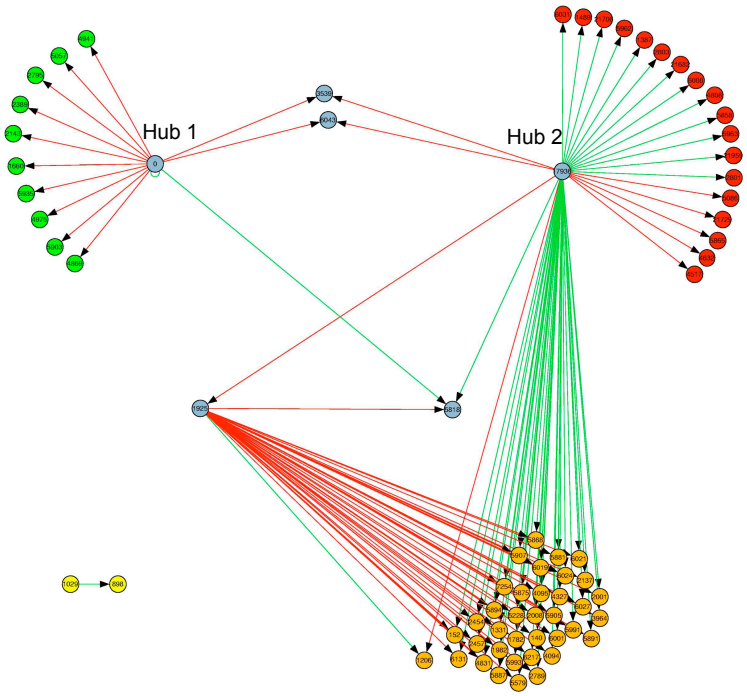


Figure 2: Gene network obtained by state-space modelling of the tetrapyrrole pathway [12], showing two major hubs that influence expression of many other genes. Hub 1 = chlorophyll; Hub 2 = enzyme of the trunk branch.

these toxic compounds the haem and cholorophyll branches of the pathway have been proposed to acts a regulators by inhibiting the first step in the tetrapyrrole pathway (Fig. 1(3)&(4)).

Chlorophyll and haem differ in the type of metal ion they contain, chlorophyll contains magnesium (Mg) and haem contains iron (Fe). The biochemical reaction for metal ion insertion is likely to be similar for both compounds, however, the chelatase enzymes that catalyse these reactions are quite different. Mg-chelatase requires ATP for activity (Fig. 1(5)) whilst it has been shown that Fe-chelatase is inhibited by ATP. So in the light when ATP levels are higher, the Mg branch of the pathway would be favoured; conversely, in the dark Mg chelation would be reduced.

We think these features make it an ideal system with which to carry out wet/dry studies of a subcellular network. However, although we have collected a lot of genomic data, we tend to interpret it based on our existing understanding of the pathway and its regulation. We have previously carried out a mathematical investigation of this pathway using state space modelling. The results shown in Figure 2 demonstrate that an enzyme of the trunk branch of the pathway might play a key role in the regulation Hub 2, unfortunately testing this hypothesis requires extensive analysis and generation of specific mutants. We decided instead to take part in the Maths Study group meeting in order to know if another mathematical approach could lead to an objective non-biased model that would allow us to confirm, or otherwise, the regulatory models, and to generate a further series of predictions that we could test experimentally.

1.2 Questions posed and aims of the modelling

The questions posed are

1. Can a model be built to describe the feedback regulation with the pathway?
    - Can it predict the time course data for chlorophyll increase?
    - Can it predict the levels of accumulation of intermediates in mutants?
2. Is the state-space model valid?
    - Does an alternative analysis of the data implicate a key role of the "trunk branch" enzyme (Hub 2)?

The modelling undertaken during the four days of the meeting and described below attempts to tackle the first question, in which a deterministic model is developed to simulate the experimental work. We did not attempt the second question, indeed this enzyme is not explicitly considered in the model, though its effects can be investigated by the changing of a single parameter (A_0).

2 Mathematical model

The mathematical model is derived from the tetrapyrrole biosynthesis pathway diagram shown in Figure 1. The pathway does not show all the molecules involved in this process, there being several intermediate compounds between, for example, ALA and protoporphyrin IX (namely Uroporphyrinogen III, Coproporphyrinogen III and protoporphyrinogen IX [7]). The 13 molecules selected can be viewed as being representative in the process, either as the head of a branching process, stopping points in the absence of light or molecules for which direct measurement has been made experimentally. Such measurements will help in model validation and parameter estimation (see Section 5).

The kinetics in the model below are the simplest possible based on the Mass Action Law. All of the inhibition processes in this system, namely Flu and haem on ALA output and ATP on haem output, act on intermediate production processes rather than directly on the molecules themselves, hence the inhibitors effect is to reduce the reaction rates of production for the molecules concerned. The effect of light is modelled as a dimensionless switch function $f(L)$ as follows

$$f(L) = \begin{cases} 0 & \text{No light,} \\ 1 & \text{Light on.} \end{cases}$$

In truth the action of light will depend on the light's intensity, leading to more complicated forms of $f(L)$; however, the experimental work concerned with this study had the lights being on or off. By the usual modelling assumptions and those just described the following system of ordinary differential equations (ODEs) can be derived,

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{A_0(1+\lambda_{AY}Yf(L))}{(1+\lambda_{AH}H)(1+\lambda_{AD}D)} - \mu_A A, \tag{1}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \beta_{PA}A - \frac{\beta_{HP}P}{1+\lambda_{HN}N} - k_{PN}PN - \mu_P P, \tag{2}$$

Molecule	Variable
ALA	A
Protoporphyrin IX	P
Haem	H
Haemoproteins	h
Active haemoproteins	Q
Phytochrome	Y
Chlorophyll Proteins	S
Pchlide (protochlorophyllide)	D
Chlide (chlorophyllide)	W
Chlorophyll	C
Membrane bound chlorophyll	M
Metabolites	G
ATP	N

Table 1: Variables used in the model.

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\beta_{HP}P}{1+\lambda_{HN}N} - \mu_H H - \beta_{HY}H - k_{hH}hH, \tag{3}$$

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = k_{hH}hH - \mu_Q Q, \tag{4}$$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \beta_{HY}H - \beta_{YS}Yf(L) - \mu_Y Y, \tag{5}$$

$$\frac{\mathrm{d}h}{\mathrm{d}t} = H_0(1+\lambda_{hY}Yf(L)) - \mu_h h - k_{hH}hH, \tag{6}$$

$$\frac{\mathrm{d}D}{\mathrm{d}t} = k_{PN}PN - \beta_{DW}f(L)D - \mu_D D, \tag{7}$$

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \beta_{DW}f(L)D - \beta_W W, \tag{8}$$

$$\frac{\mathrm{d}C}{\mathrm{d}t} = \beta_W W - k_{SC}SC - \mu_C C, \tag{9}$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \beta_{YS}Yf(L) - \mu_S S - k_{SC}SC, \tag{10}$$

$$\frac{\mathrm{d}M}{\mathrm{d}t} = k_{SC}CS - \mu_M M, \tag{11}$$

$$\frac{\mathrm{d}G}{\mathrm{d}t} = k_{GM}NMf(L) - \beta_G(1+\delta_{GL}f(L))G + G_0(t), \tag{12}$$

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \delta_{NG}\beta_G(1+\delta_{GL}f(L))G + k_{NQ}MQf(L) - \mu_N N - k_{NM}NMf(L), \tag{13}$$

The variables are listed in Table 1 and correspond to the bracketed letters in Figure 1. The only feature of the model that has not been discussed so far is the function $G_0(t)$ in equation (12). This represents metabolite sources other than that obtained from photosynthesis, including nutrients still in the seed and that obtained from soil or growth media.

To close the system of ODEs we need a set of initial conditions. With regards to the experimental work concerned, we assume the light is switched on at $t = 0$ (in simulations $t = 10$) and all the variables start at the steady-state values for the case $f(L) = 0$. We note for a plant that has never been exposed to light we have $S = 0, W = 0, C = 0$ and $M = 0$, corresponding to a plant without chlorophyll.

A description of all of the parameters are listed in Table 2.

2.1 Quasi-steady approximations

It is routine at this point to determine where possible relative timescales of each of the processes in the system; these can often be exploited to simplify a system of ODEs, by expressing some of the variables in terms of the others in the form of algebraic equations. If all the reaction rates were known, then this process can be made systematic. However, lacking such data we took a more heuristic approach. Further work in the parameter estimation process may go some way to assist in going the systematic route.

The overall timescale of interest is about 24 hours and the "slow" events seem to occur on a timescale of hours. Reactions and gene expression will occur much faster (on the order of minutes). Experience of plants surviving and staying green overnight suggest to us that changes in stable chlorophyll (M) and metabolites (G) is rather slow, and experimental time-course data suggests that Chlide (W) and active-haemoproteins (Q) also take a few hours to change or decay. For the remaining variables we debated which were the fast (seconds), medium (minutes) and slow (hours) timescales and proposed the following quasi-steady assumptions.

$$Y \sim \frac{\beta_{HY} H}{\beta_{YS} f(L) + \mu_Y}, \quad D \sim \frac{k_{PN} PN}{\beta_{DW} f(L) + \mu_D}, \tag{14}$$

$$h \sim \frac{H_0(1 + \lambda_{hY} Y f(L))}{k_{hH} H + \mu_h} \sim \frac{1}{k_{hH} H} \left(\frac{\beta_{HP} P}{1 + \lambda_{HN} N} - \mu_H H - \beta_{HY} H \right), \tag{15}$$

$$S \sim \frac{\beta_{YS} Y f(L)}{k_{SC} C + \mu_S} \sim \frac{\beta_W W - \mu_C C}{k_{SC} C}, \tag{16}$$

$$A \sim \frac{A_0}{\mu_A} \frac{(1 + \lambda_{AY} Y f(L))}{(1 + \lambda_{AH} H)(1 + \lambda_{AD} D)} \sim \frac{1}{\beta_{PA}} \left(\frac{\beta_{HP} P}{1 + \lambda_{HN} N} + k_{PN} PN + \mu_P P \right), \tag{17}$$

$$N \sim \frac{\delta_{NG} \beta_G (1 + \delta_{GL} f(L)) G + k_{NQ} MQ f(L)}{k_{NM} M f(L) + \mu_N}. \tag{18}$$

It can be shown by examination of these expressions that it is possible to express all the quasi-steady variables in terms of the slow variables M, G, W and Q, though care would be needed as multiple solutions are possible. The complexity of these terms prevents easy simplification of the full system by this route and, consequently, no further progress was made in this analysis.

2.2 Steady-state analysis

Since the model does not take into account plant growth, the steady-state analysis summarised here will only be of relevance for the initial conditions, at which the plant is exposed to light for the first time. For such a plant, the Pchlide-Chlide and phytochrome-chlorophyll protein pathways will always have been blocked, leading to $S = W = C = M = 0$ (i.e. no chlorophyll production). If the remaining slow processes (involving G and Q, see Section 2.1) operate sufficiently fast in terms of the plant growth timescale, then the steady-state values corresponding to $f(L) = 0$ should provide appropriate initial conditions for the simulations to follow.

In the absence of light the non-zero steady-states, denoted with a *, eventually satisfy,

$$A_A^* = \frac{\mu_D A_0}{\mu_A (1 + \lambda_{AH} H^*)(\mu_D + \lambda_{AD} k_{PN} P^* N^*)}, \tag{19}$$

Parameter	Description of constants	Simulation	Estimate
A_0	Background ALA production rate	10	0.110
H_0	Background haemoproteins production rate	1	0.583
G_0	Background metabolite production rate	0.1	0.262
k_{PN}	Protoporphyrin IX → Pchlide conversion rate	100	1
k_{hH}	Haem-haemoprotein reaction rate	100	1.11×10^{-8}
k_{SC}	Chlorophyll-chlorophyll-protein reaction rate	100	0.0853
k_{GM}	Metabolite synthesis rate from photosynthesis	100	0.488
k_{NQ}	ATP output rate from photosynthesis	2	0.0232
k_{NM}	ATP loss rate from metabolite synthesis	1	0.470
λ_{AY}	Phytochrome enhanced ALA synthesis const.	1	0.972
λ_{AH}	Haem inhibition of ALA constant	5	0.650
λ_{AD}	Pchlide inhibition of ALA constant	1	0.535
λ_{HN}	ATP inhibition of haem constant	100	7.29×10^{-9}
λ_{hY}	Phytochrome enh. haemoprotein prod. const.	1	0.240
β_{PA}	ALA → protoporphyrin IX conversion rate	100	5.21×10^{-9}
β_{HP}	Protoporphyrin IX → haem conversion rate	100	1
β_{HY}	Haem → phytochrome conversion rate	10	3.34×10^{-9}
β_{YS}	Phytochrome → chlorophyll protein conv. rate	10	0.0790
β_{DW}	Pchlide → Chlide conversion rate	100	1
β_{DW} (PORA)	Mutant's Pchlide → chlide conversion rate	-	30
β_W	Chlide → chlorophyll conversion rate rate	1	0.00226
β_G	Metabolite → ATP conversion rate	1	0.500
δ_{NG}	Metabolite → ATP conversion factor	2	1.76
δ_{GL}	Enhanced metabolism rate due to light	2	0.328
μ_A	ALA decay rate	100	0.327
μ_P	Protoporphyrin IX decay rate	10	1
μ_H	haem decay rate	1	6.57×10^{-9}
μ_Q	Active haemoproteins decay rate	1	0.0547
μ_Y	Phytochrome decay rate	100	3.06×10^{-7}
μ_h	haemoprotein decay rate	100	0.728
μ_D	Pchlide decay rate	5	1
μ_C	chlorophyll decay rate	0.1	7.53×10^{-7}
μ_S	chlorophyll protein decay rate	1	0.00212
μ_N	Overall ATP consumption rate	100	0.857
μ_M	Membrane bound chlorophyll decay rate	1	0.0793

Table 2: Table of the model's parameters including description and values used in the simulations of Section 4 (guessed from estimates of timescales) and curve-fitted estimates found in Section 5. Parameter units: $*_0$ = [conc.]/[time], k_{**} = 1/[conc.][time], λ_{**} = 1/[conc.], $\beta_{**}, \beta_*, \lambda_{**}$ = 1/[time] and δ_{**} = dimensionless.

$$A_P^* = \frac{1}{\beta_{PA}}\left(\frac{\beta_{HP}}{1+\lambda_{HN}N^*} + k_{PN}N^* + \mu_P\right)P^*, \quad (20)$$

$$P^* = \frac{(1+\lambda_{HN}N^*)}{\beta_{HP}}\left(\mu_H + \beta_{HY} + \frac{k_{hH}H_0}{\mu_h + k_{hH}H^*}\right)H^*, \quad (21)$$

$$Q^* = k_{hH}h^*H^*/\mu_Q, \quad (22)$$

$$Y^* = \beta_{HY}H^*/\mu_Y, \quad (23)$$

$$D^* = k_{PN}P^*N^*/\mu_D, \quad (24)$$

$$h^* = H_0/(\mu_h + k_{hH}H^*), \quad (25)$$

$$G^* = G_0^*/\beta_G, \quad (26)$$

$$N^* = \delta_{NG}G_0^*/\mu_N, \quad (27)$$

which are 9 expressions derived from the respective equations (1)-(7), (12) and (13), where (24) is used to obtain (19). We note we have used A_A^* and A_P^* to emphasise that the steady-state expressions for ALA result from the "A" and "P" equations (1) and (2), respectively. The parameter G_0^* is a "representative" metabolite feed rate at the time the light is switched on.

Substitution of (21) into the equation resulting from that formed by writing (19)=(20) results with a quintic polynomial in H^* (details omitted in this discussion due to the rather ugly nature of the coefficients). In principle, this could mean that there may be 5 physical solutions of H^* (i.e. real and positive) representing 5 different "valid" steady-states. However, examination of the signs of the coefficients of successive powers in the quintic reveal that there is only one change of sign, which, by Descartes' Rule of Signs (see [9]), means that there is one and only one root H^* that is positive and real; i.e. only one of the 5 roots is biologically relevant. Using this root, all the other steady-state terms can be found uniquely using (19)-(25). In practice, it is simpler to numerically solve the system (1)-(13) for $f(L) = 0$ long enough so each of the variables have reached their steady-state values, and use these as initial conditions in the simulations.

3 Experimental results

The evolution of various components of the pathway as measured in experiments are shown in Figure 3, namely, the levels of Pchlide and Chlide (a), chlorophyll (b) and ATP levels (amongst others, (c)). We observed that Pchlide levels drop dramatically on exposure to light, apparently increasing slowly after 10 hours or so. Meanwhile Chlide levels has increased significantly over the first 7-8 hours, slowing down and perhaps saturating by 24 hours. This pattern is strongly reflected in the levels of Chlorophyll. ATP levels in both the wild-type and mutant rise sharply during the first hour and then fall away over a longer timescale as observed in Figure 3(c). Capturing the qualitative forms of these curves using the model proposed above is the aim of the first part of the next section.

4 Simulation results

We run the simulations of the model using parameters that give reasonably comparable results with experimental observations in a qualitative sense. The parameter values are listed in the 3rd column of Table 2. Initial choices of parameter values were selected based on estimates

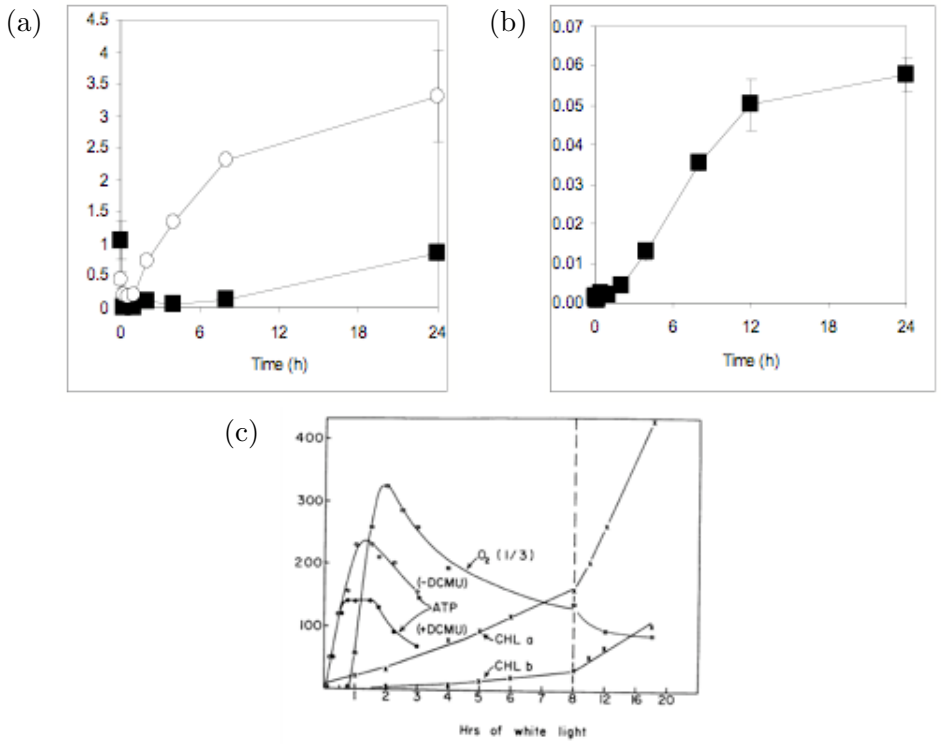


Figure 3: Time evolution of (a) Pchlide (black squares) and Chlide (white circles), (b) Chlorophyll and (c) ATP levels (amongst others, see graph labelling) measured for a plant exposed to light following 3 days in darkness after germination. Figure (c) is taken from Oelze-Karow and Butler [10].

of timescales, critical concentrations etc., in which the values of 1, 10, 100 represented, for example, slow, medium and fast timescales, respectively. These initial paremeter estimates were then tweaked to get solutions that agreed reasonably with experimental results.

The model equations were simulated using Mathematica. These simulations consisted of two phases: a first phase of ten time units in dark condition where $F = 0$, then a light phase using $F = 1$. By $t = 10$ all the variables have typically settled to a steady-state and are used as initial conditions at the point the light is switched on. This should be a good approximation to the plant's condition following its first 3 days following germination in darkness.

Shown in Figure 4 is the evolution of each of the 13 variables using the standard parameter set, which represents the "wild-type" plant. For the purposes of matching timescales of the model solutions with the experimental data shown in Figure 3, two simulated time units represents about 24 hours. The simulations predict a very rapid decay of the phytotoxic chemicals like Pchlide, being in agreement with the experimental data described in Section 3. Furthermore, the predicted increase of stable chlorophyll and Chlide on longer timescales is again in agreement with observation. The increase levels of ATP is perhaps too sharp, but the general rapid increase and gradual decay is in broad agreement to the picture shown in Figure 3(c).

To further validate the model, simulations of mutant strains were performed and the results compared with experimental studies. Figure 5 shows the model solutions of a *hy1* mutant,

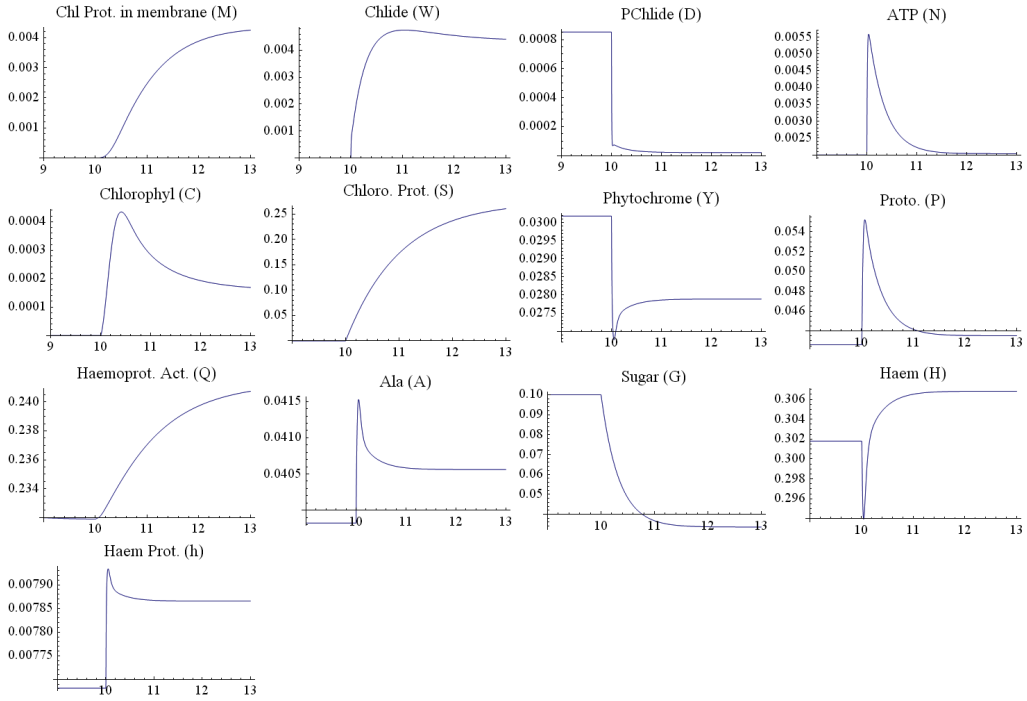


Figure 4: Time evolution of each variables for the wild type strain. A single time unit for this data is approximately 12 hours.

which has reduced expression of haem oxygenase, an enzyme responsible for the conversion of haem to phytochrome. In the hy1 mutation leads to a reduced rate of phytochrome synthesis to about half that of the wild type (parameter β_{HY}). The experimental studies only reported comparisons of Chlorophyll levels between the wild-type and mutant. In agreement with experiments, the model predicts only a marginal reduction in chlorophyll levels. The simulations predict that the reduction in phytochrome output in the mutant leads to enhanced accumulation of haem and consequently activated haemoprotein production. The reduced phytochrome levels, however, lead to the reduction in chlorophyll protein production and hence lower stable chlorophyll levels. In addition, an increase of the haem level affects the production of tetrapyrrole intermediate (Figure 5, Proto (P)) by inhibiting the first step of the tetrapyrrole biosynthesis

Interestingly, metabolite and ATP levels appear to be unaffected in Figure 5, as well as in Figures 6 and 7. Examination of the model's parameter and these simulation results show that the background production rate of the metabolite ($G_0 = 0.1$time units^{-1}) far exceeds that of Chlorophyll based production ($k_{GM}NMf(L) = O(0.0001)$time units^{-1}). This is probably unrealistic and will need to be investigated in the future.

The *gun5* mutant expresses altered Mg-chelatase activity, which is involved in the conversion of protoporphyrin IX to Mg-protoporphyrin IX, intermediate of the Pchlide synthesis. The gun5 mutation leads to a reduced rate of Pchlide synthesis to about half that of the wild type. The model's prediction of this mutant's response to light is depicted in Figure 6. The reduction in stable chlorophyll (M) in the mutant is quite significant. These results are in complete accordance with Mochizuki *et al.*[5] that gun5 mutants accumulate less Chlorophyll than the wild type. All components downstream of the protoporphyrin IX-Pchlide pathway

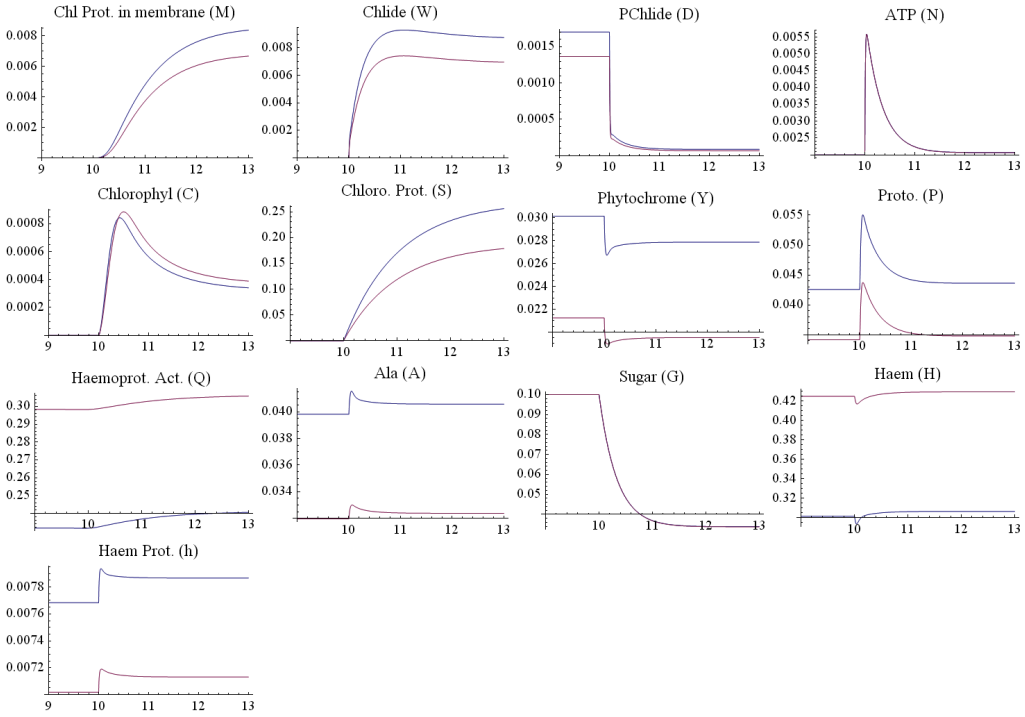


Figure 5: Time evolution of each variables for the *hy1* mutant (red, $\beta_{HY} = 5$) and the wild type (blue, $\beta_{HY} = 10$).

have been reduced to about half the levels of the wild-type. Surprisingly, the model predicts that the levels of all other components are almost identical between the mutant and wild-type strains.

Figure 7 shows the response from the PORA over-expressing line, which over expresses an enzyme that enhances the rate of Pchlide-Chlide conversion (increased β_{DW}). The results suggest that response is almost identical to that of the wild-type. This does not compare well with experimental results as levels of stable-chlorophyll (M) are recorded to be higher in the over-expressing line than the wild-type. The most likely explanation for this discrepancy is the choice of parameter values, which were not compiled with a great deal of sophistication. Another possibility is that there is another control step not accounted for by the model, however, much more scrutiny of the parameters is required before any such conclusions can be drawn.

5 Parameter estimation

5.1 Introduction

After constructing a mathematical model of a system, it is useful to see how closely the output of the model matches available experimental data. In order fit the model to the data, a set of parameters must be chosen that control the model to optimally reproduce the real world data. In the proposed model for the tetrapyrrole biosynthesis pathway, these are the parameters listed in Table 2.

One way of finding an optimal set of parameters is to check all combinations, measuring

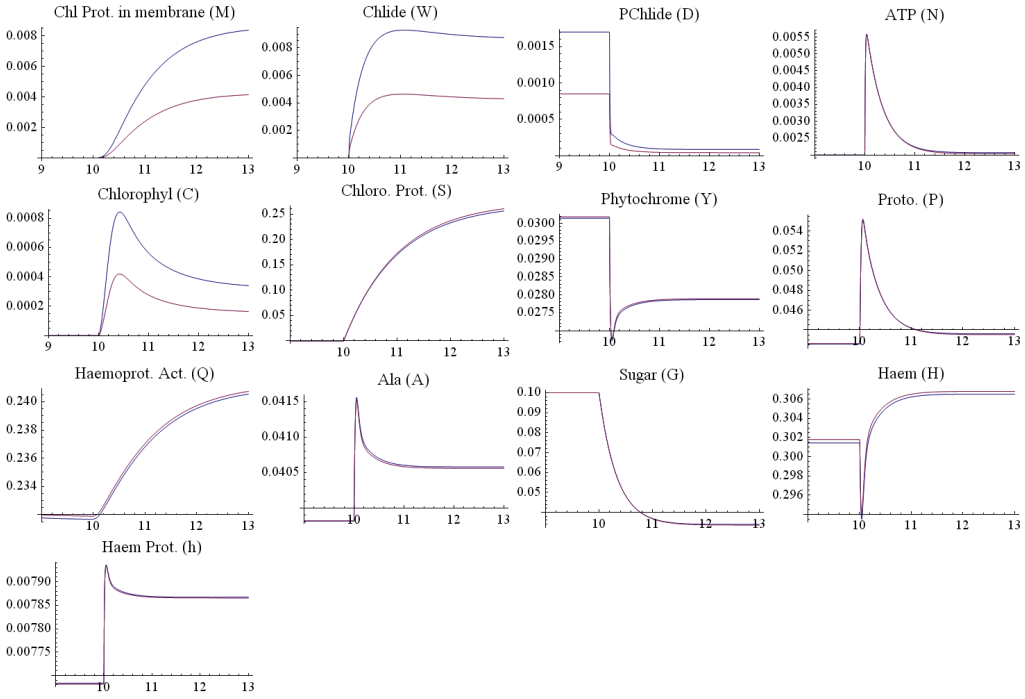


Figure 6: Time evolution of each variable for the *gun5* mutant (red, $k_{PN} = 50$) and the wild type (blue, $k_{PN} = 100$).

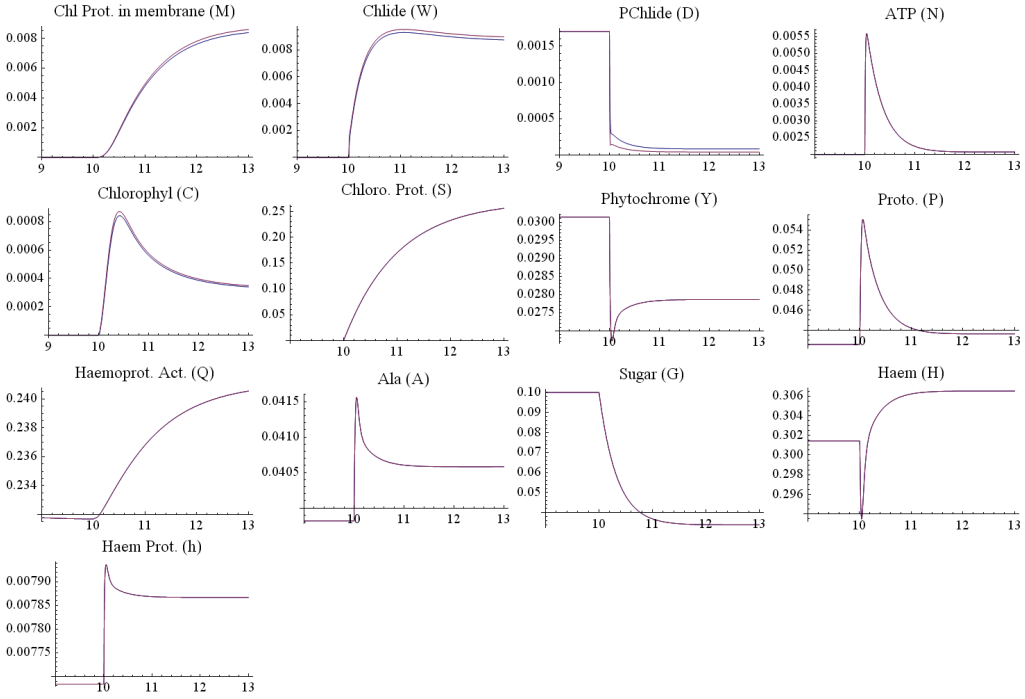


Figure 7: Time evolution of each variable for the PORA over-expressing line (red, $\beta_{DW} = 200$) and the wild type (blue, $\beta_{DW} = 100$).

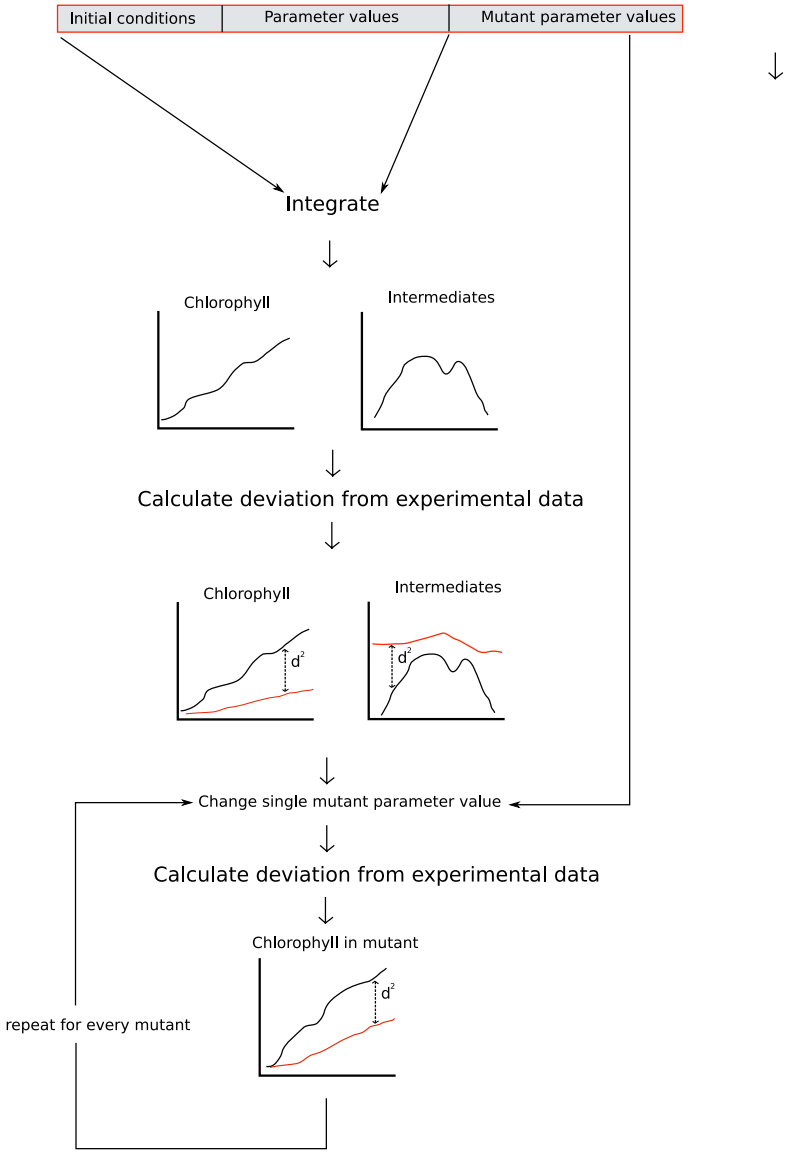


Figure 8: Parameter values generated from the SRES algorithm are used to integrate the model. The results are then compared to experimental data. Each mutant is expected to differ by only a single parameter value. The applicable parameter is changed, the model integrated and a comparison done with mutant experimental data.

how well each combination fits the experimental data. However, this method rapidly becomes unfeasible as the number of parameters increases. As an example, if there are 10 parameters and only 8 different parameter values are considered, the model has to be fitted to the data 8^{10} times. In order to reduce this computational workload, more discerning methods exist for proposing and combining parameter values.

5.2 Method

We chose a parameter estimation method known as SRES, the Stochastic Ranking Evolution Strategy [11]. SRES has been shown to perform well with models of biochemical pathways [6]

and is freely available as a library for C programmers [3].

The algorithm works by testing around 300 sets of parameters in a single step known as a generation. The best sets of parameters are then selected to breed new parameters, which are subjected to small changes in a step known as mutation, for the next generation. In order to assess if one set of parameters is better than another, the algorithm requires a fitness function to be provided by the user. When the fitness function is given a set of parameters, it returns a number indicating how closely the model fits the experimental data when used with the parameter values. A lower number indicates a better fit and a higher fitness.

An overview of the fitness function we used is given in Figure 8. The experimental data is compared, with data generated from the model, using the Normalised Root Mean Square Deviation (NRMSD) method. The use of NRMSD allows different measurement scales to be compared by the fitness function. Given n experimental data points $\mathbf{y}$ and n data points generated from the model $\mathbf{z}$, the NRMSD is calculated as:

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - z_i)^2}{n}}, \tag{28}$$

$$\text{NRMSD} = \frac{\text{RMSD}}{y_{max} - y_{min}}, \tag{29}$$

where y_{max} and y_{min} are the maximum and minimum values of the data.

When calculating a value for the fitness, the wild type chlorophyll experimental results are given a weighting of 2.0. Intermediates and mutants given a weighting of 1.0 to emphasise the importance of the chlorophyll production. Currently, there are 36 parameters for the wild-type, one parameter for the mutant and 10 initial conditions to estimate from 40 data points. However, only the PORA over-expressing line is currently being considered and experimental data from additional mutants is available.

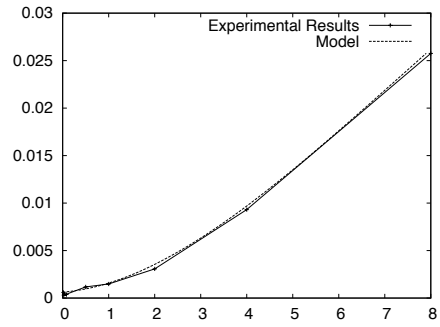
5.3 Results

Figure 9 shows the fit between experimental and model data after an overnight run of the SRES algorithm over 144000 generations. All parameter values were constrained between 0.0 and 1.0 except β_{DW} for the PORA over-expressing line, which could take any value between 5.0 and 30.0.

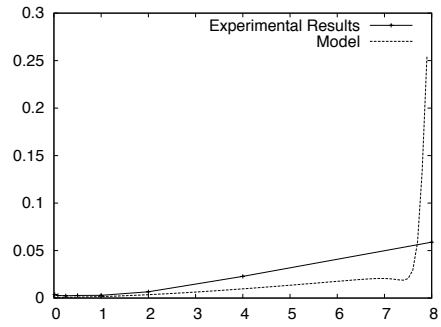
Table 2 shows the parameter values after fitting to experimental data. Table 3 gives the estimated initial conditions. Currently, no units have been considered. Both the parameters and initial conditions are given without specifying their units.

6 Discussion

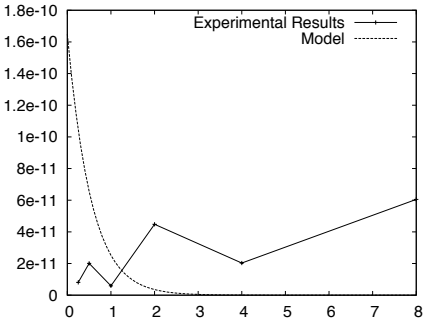
In this report a mathematical model was developed and studied to describe the tetrapyrrole regulation pathway, in particular response to light exposure after 3 days following germination in darkness. Simulations of the model seems to produce results that are consistent with experimental observation, including those of the mutant strains. It should be stressed that the results presented in Section 4 and the parameter values shown in Table 2 should not be viewed as definitive. However, the results thus far, particularly regarding the wild-typeand



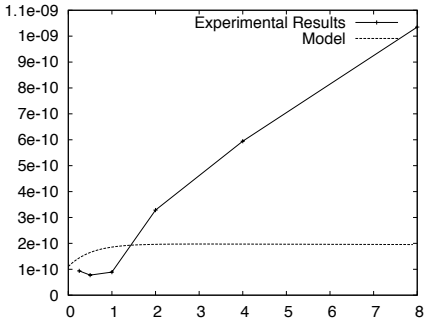
(a) Chlorophyll in Membrane WT



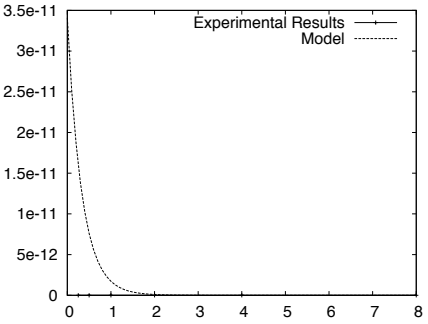
(b) Chlorophyll in Membrane PORA o/e



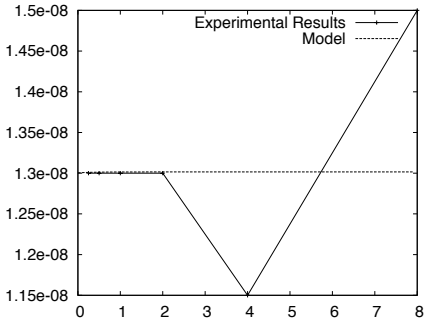
(c) PChilde



(d) Childe



(e) Protoporphyrin IX



(f) Haem

Figure 9: Model output compared to experimental results for Chlorophyll and intermediates.

hy1 and *gun5* mutant strains, are encouraging and the current model should provide the basis for further study.

With the limited duration of the studygroup we were principally concerned with getting the model to qualitatively predict observed results. However, a serious attempt was made at systematically determining suitable parameter values using curve fitting techniques with wild-type and mutant time-course data, this being described in Section 5. This is a highly non-trivial task involving the estimation of about 35 parameters plus initial conditions with about as many data points. As can be seen from the values in Table 2 there is considerable disparity between the simulated and curve fitted values, in particularly the relative sizes of parameters to each other; this may be due to the constraining of values between 0 and 1 in the curve fitting procedure. However, this work is very much in progress and there are a

Variable name	Identifier	Initial Condition
ALA	A	0.806765
Protoporphyrin IX	P	4.6e-11
Haem	H	1.3e-08
Phytochromobilin	Y	0.505698
Pchlide	D	1.97e-10
Chlide	W	9.28e-11
Chlorophyll	C	0.384744
Chlorophyll proteins	S	0.00861472
ATP	N	0.993669
Chlorophyll in membrane	M	0.000596312
Activated haemoprotein	h	0.262178
Sugars	G	0.806906
Stable haem + haemoprotein	Q	0.797839

Table 3: Estimated initial conditions after running the SRES algorithm.

number of avenues that can be tried to assist the searching algorithm; this will hopefully be undertaken in the near future.

The modelling focussed on the tetrapyrrole regulation processes for a particular experiment. There is certainly considerable potential for further work on the current model, which through improved parametrisation and tweaking of the terms will hopefully be able to provide quantitatively accurate reproductions of existing data and provide predictions and insights that will motivate further investigation. In broader applications, other regulation processes are known to be important, in particular, it seems, the circadian signal transduction process. Such factors are not accounted for in the current model and offers considerable potential for future work.

References

[1] J.E. Cornah, M.J. Terry and A.G. Smith (2003). *Trends Plant Sci.*, **8**, 224-230.

[2] J.E. Cornah, J.M. Roper, D.P. Singh and A.G. Smith (2002). *Biochem J.*, **362**, 423-432.

[3] X. Ji and Y. Xu (2006). *Bioinformatics*. **22**, 124-126.

[4] A.C. McCormac and M.J. Terry (2002). *Plant J.*, **32**, 549-559.

[5] N. Mochizuki, J.A. Brusslan, R. Larkin, A. Nagatani and J. Chory (2001). *PNAS*, **98**, 2053-2058.

[6] C.G. Moles, P. Mendes and J.R. Banga (2003). *Genome Res.*, **13**, 2467-2474.

[7] M. Moulin and A.G. Smith (2005). *Biochem. Soc. Trans.*, **33**, 737-742.

[8] A. Murakami and Y. Fujita (1991). *Plant Cell Physiol.*, **32**, 223-230.

[9] J.D. Murray, Mathematical Biology 2nd Ed. (1993), Springer-Verlag.

[10] H. Oelze-Karow and W.L. Butler (1971). *Plant Physiol.*, **48**, 621-625.

[11] T.P. Runarsson and X. Yao (2000). *IEEE Trans. Evolution. Comput.*, **4**, 284-294.

[12] B. Schmid (2006). "Modeling of microarray data for the investigation of the tetrapyrrole pathway in *A. thaliana*. Diploma thesis, University of Applied Sciences Weihenstephan.