Linear models for endocytic transformations from live cell imaging

John A. Belward[1] Kevin Burrage[2]
Rohan D. Teasdale[3] Nicholas A. Hamilton[4]

(Received 22 December 2010; revised 11 May 2011)

Abstract

Endocytosis is the process by which cells internalise molecules including nutrient proteins from the extracellular media. In one form, macropinocytosis, the membrane at the cell surface ruffles and folds over to give rise to an internalised vesicle. Negatively charged phospholipids within the membrane called phosphoinositides then undergo a series of transformations that are critical for the correct trafficking of the vesicle within the cell, and which are often pirated by pathogens such as *Salmonella*. Advanced fluorescent video microscopy imaging now allows the detailed observation and quantification of these events in live cells over time. Here we use these observations as a basis for building differential equation models of the transformations. An initial investigation of these interactions was modelled with reaction rates proportional to the sum of the concentrations of the individual constituents. A first order linear system for the concentrations results. The structure of the system enables analytical expressions to

http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/3801 gives this article, © Austral. Mathematical Soc. 2011. Published May 31, 2011. ISSN 1446-8735. (Print two pages per sheet of paper.) Copies of this article must not be made otherwise available on the internet; instead link directly to this URL for this article.

be obtained and the problem becomes one of determining the reaction rates which generate the observed data plots. We present results with reaction rates which capture the general behaviour of the reactions so that we now have a complete mathematical model of phosphoinositide transformations that fits the experimental observations. Some excellent fits are obtained with modulated exponential functions; however, these are not solutions of the linear system. The question arises as to how the model may be modified to obtain a system whose solution provides a more accurate fit.

Contents

1 Introduction

The endolysosomal system of the eukaryotic cell is a highly dynamic network of heterogeneous membrane bound compartments that mediates the internalization of nutrients, signaling factors, particulate matter and fluid from the extracellular environment into the cell [1]. In addition to being essential for the maintenance of cellular homeostasis, the endosome plays a

primary role in the sorting and segregation of internalized receptors which in turn modulates both the type and extent of the elicited signaling response. As an entry point into the cell, endocytic pathways are often exploited and modulated by pathogens [2].

Mathematical modelling of biological systems often suffers from the sparsity of experimental data available. To create a detailed model of how certain concentrations of interacting molecules change over time usually requires, at the very least, experimental snapshots of the concentrations at one or more time points. However, recent advances in automated fluorescent microscope imaging technology enable the experimental determination of a protein's sub-cellular localisation and its dynamic trafficking within a range of cellular contexts in live cells, and hence to a deeper understanding of the proteins and their function.

Here we show how this new detailed time based imaging can be utilised to generate a model of endocytic transformations. In particular, we focus on the phosphoinositides, a class of 3-phospholipids populating the membrane surfaces of endosomes (Figure 1). Through the action of kinases such as PIKfyve these may be transformed into one another as part of the maturation process of the endosome. Here, by combining experimental live cell imaging with modelling techniques we build a detailed rate based model of the transformations being observed. In the future we aim to use such models to predict modulations to the system such as those produced by pathogen invasion.

2 Biomolecular experimental data and an analysis of a linear model

We rely on experimental data generated to observe the molecules PI(3,4,5)P_3, PI(3)P, PIKfyve and a mutant (inactive) form of PIKfyve. Kerr et al. [3] describe the experimental details in full. The essence of the approach is to

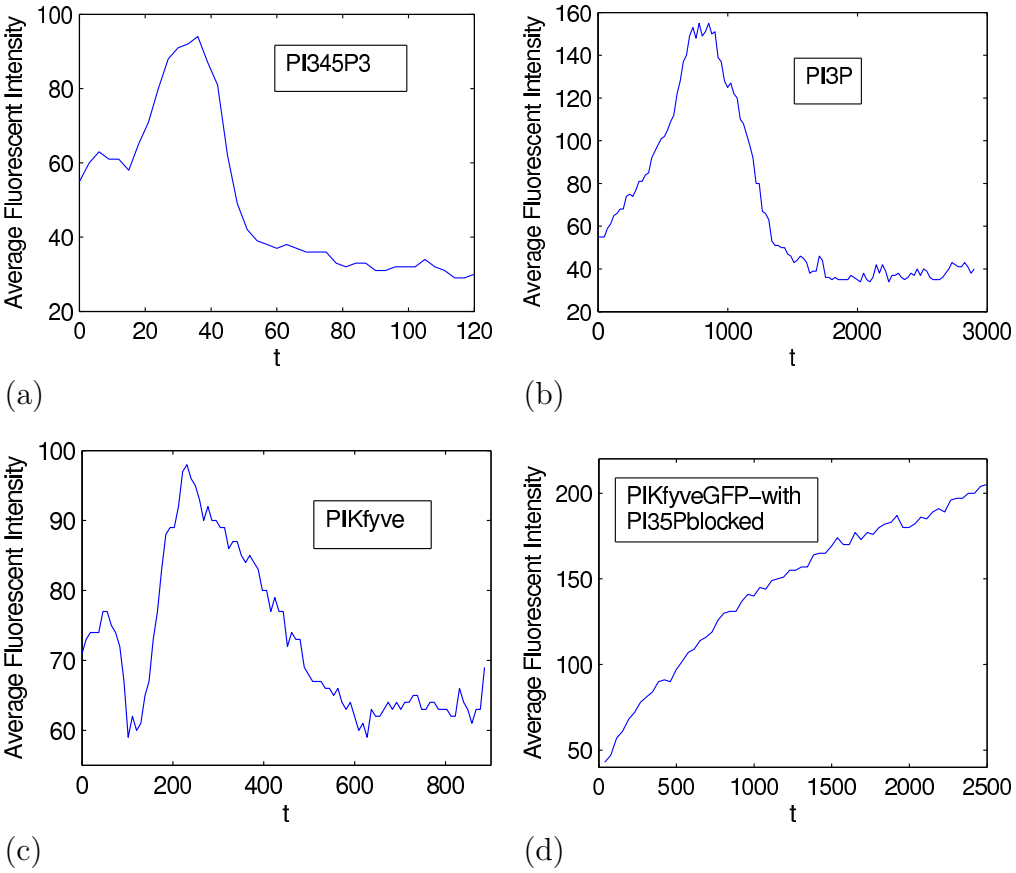


FIGURE 1: These plots are (a) PI(3,4,5)P_3, (b) PI(3)P, (c) PIKfyve and (d) a mutant PIKfyve that blocks the transformation of PI(3)P to PI(3,5)P_2; the time is measured in seconds. Adapted from results by Kerr et al. [3].

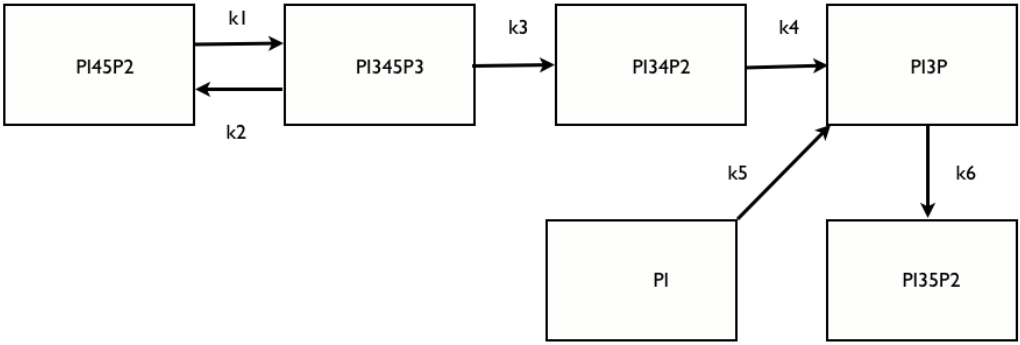


FIGURE 2: Schematic representation of the reactions.

attach green fluorescent protein (GFP) to the protein of interest in live cells. Time lapse videomicroscopy is then performed on individual live cells using a Zeiss LSM 510 meta confocal scanning microscope. The GFP is excited with the 488 nm argon laser line, and a series of (3D) z-stacks are collected to create an image sequence of the protein of interest's location over time. To allow two proteins to be be simultaneously observed and distinguished another fluorescent marker, mCherry, that has a distinct spectrum from GFP can be used. In several of the experiments, a protocol was used that enabled the contents of the vesicles to be filled with fluorescently visible Dextran. This enables the region of a vesicle to be delineated, and the recruitment and/or expulsion of a fluorescently tagged protein of interest to the vesicle region to be quantified. Several live cell movies showing these dynamic transformations were published by Kerr et al. [3]. The four plots of the data are shown in Figure 1. The average fluorescent intensity, measured on an arbitrary scale, is assumed proportional to the concentration and time t is measured in seconds. There is a considerable amount of noise in the system and the changes of slope early in the plots of PI(3,4,5)P_3 and PIKfyve have been assumed due to experimental or observational effects. There is a background fluorescence of approximately 40 units, a manifestation of auto-fluorescence or non-specific localisation. Accordingly, some of the fits for these two compounds have been

made on subsets of the data.

A succession of reactions takes place at rates dependent on the concentrations of the compounds taking part in the reactions, as well as certain enzymatic kinases (which are assumed to have constant concentration). The rate of change of concentration is a linear combination of the constituents taking part in the reaction leading to a model portrayed as a system of linear constant coefficient differential equations. The assumptions implicit in such models are explained by Wilkinson [4]. This model was obtained by extracting a subset of possible phosphoinostide transformations from all those possible [5]. The differential equation for the concentrations of the components of the system is

$$\frac{d\mathbf{X}}{dt} = \begin{bmatrix} -k_1 & k_2 & 0 & 0 & 0 & 0 \\ k_1 & -k_2 - k_3 & 0 & 0 & 0 & 0 \\ 0 & k_3 & -k_4 & 0 & 0 & 0 \\ 0 & 0 & k_4 & -k_6 & k_5 & 0 \\ 0 & 0 & 0 & 0 & -k_5 & 0 \\ 0 & 0 & 0 & k_6 & 0 & 0 \end{bmatrix} \mathbf{X}, \tag{1}$$

where $\mathbf{X} = (\mathrm{PI45P2}, \mathrm{PI345P3}, \mathrm{PI34P2}, \mathrm{PI3P}, \mathrm{PI}, \mathrm{PI35P2})$. Figure 2 shows a schematic diagram of the process.

The general solution of the system given in equation (1) is

$$\mathbf{X} = \begin{bmatrix} \mathrm{PI45P2} \\ \mathrm{PI345P3} \\ \mathrm{PI34P2} \\ \mathrm{PI3P} \\ \mathrm{PI} \\ \mathrm{PI35P2} \end{bmatrix} = \mathbf{M} \begin{bmatrix} c_1 \exp(rt) \\ c_2 \exp(st) \\ c_3 \exp(-k_4 t) \\ c_4 \exp(-k_5 t) \\ c_5 \exp(-k_6 t) \\ c_6 \end{bmatrix} \tag{2}$$

where matrix M is

$$\begin{bmatrix} sk_2(s+k_6)(s+k_4) & rk_2(r+k_6)(r+k_4) & 0 & 0 & 0 & 0 \\ s(s+k_1)(s+k_6)(s+k_4) & r(r+k_1)(r+k_6)(r+k_4) & 0 & 0 & 0 & 0 \\ sk_3(s+k_6)(s+k_1) & rk_3(r+k_6)(r+k_1) & k_4-k_6 & 0 & 0 & 0 \\ sk_3k_4(s+k_1) & rk_3k_4(r+k_1) & k4 & k_5 & -1 & 0 \\ 0 & 0 & 0 & k_6-k_5 & 0 & 0 \\ k_3k_4k_6(s+k_1) & k_3k_4k_6(r+k_1) & k_6 & k_6 & 1 & 1 \end{bmatrix} \tag{3}$$

and r and s are the roots of the quadratic equation

$$\alpha^2 + (k_1+k_2+k_3)\alpha + k_1k_3 = 0\,. \tag{4}$$

Since the discriminant, D, of equation (4) is

$$D = (k_1+k_2+k_3)^2 - 4k_1k_3 = (k_1+k_2-k_3)^2 + 4k_2k_3\,, \tag{5}$$

the right-hand expression guarantees that both r and s are real and the left-hand expression, being less than $k_1+k_2+k_3$ in magnitude ensures that the two roots are negative. On the assumption that the reaction rates k_i are positive, all the exponents in the solution are negative, including r and s. This means that all compounds decay to zero, except PI(3,5)P_2. The exponents in the expression for the solution may be all different. If the eigenvalues of the matrix in equation (1) have repeated roots, then terms of the form $t\exp(-k_it)$ or even $t^2\exp(-k_it)$ are possible. In fact r and s cannot be equal since the discriminant of the quadratic in α is always positive.

A second system is obtained when the evolution of PI(3,5)P_2 from PI(3)P is blocked. The box representing PI(3,5)P_2 is absent as is the rate constant k_6. The system matrix is order five and comprises the first five rows and columns of that in equation (1) with the rate k_6 set to zero. The general solution is similar to that of (1) except that the PI(3,5)P_2 component is absent and, as in (2), k_6 is zero. Experimentally this case corresponds to a mutant form of the PIKfyve kinase which does not enable PI(3)P to be phosphorylated to PI(3,5)P_2.

3 Fitting the linear model

The general solution of the model problem is a linear combination of decaying exponentials. In principle a set of least squares problems can determine the rates and magnitudes of the compounds. However, the experimental data is far from complete and the problem is to determine as much information as possible from the experimental data.

The approximants are obtained by a combination of interpolation, least squares fitting and numerical experiment. We set up a minimisation problem using a mathematical programming routine to locate the parameter values which optimise the quality of the approximation. The latter is measured by some function of the residuals, the differences between the value of the observation and the approximant at some chosen set of points. Least squares objective functions are an obvious choice leading to a nonlinear least squares problem in the case of the functions appearing in equation (2). This is sometimes successful, but in several instances the objective has multiple optima, making it difficult for a black box minimisation routine.

It is possible to make the problem easier for a black box routine by determining the parameters in a sequence of stages. In the examples above most progress was made by piecing together these stages, sometimes solving one or two variable problems by algebraic methods.

One starting point for a multistage approximation algorithm is to argue that for large times, only one term will remain, that corresponding to the smallest magnitude exponential. So we fit the tails and express the solution as the sum of this known part plus further exponential terms. This technique was applied to PI(3,4,5)P_3; the solution is, according to equation (2), the sum of two exponentials.

When PI(3,5)P_2 is blocked the data is given as the fluorescence of PIKfyve which binds to PI(3)P and whose presence is responsible for the production of PI(3,5)P_2 from PI(3)P. In this way the PIKfyve fluorescent intensity is seen

as a proxy for measuring PI(3)P. The time scales for PI(3)P and PI(3,4,5)P_3 suggest that these reactions are uncoupled. Thus the time dependent part may be captured as a single exponential term. Figure 3(a) shows a fit to a mutant form of the PIKfyve kinase which blocks the transformation of PI(3)P to PI(3,5)P_2. The fitted curve is

$$\text{PIKfyve} = 247.6 - 204.6\exp(-0.00064t), \tag{6}$$

which fits the data with remarkable accuracy. Whether it is determined by a least squares fit or by a simple analytic manipulation the residuals are small and alternate in sign reminiscent of the optimality criterion of the Remez algorithm.

A fit to the data for PI(3,4,5)P_3 appears in Figure 3(b):

$$f(t) = ae^{\alpha t} + be^{\beta t} + c \tag{7}$$

with

$$\begin{aligned} a &= -1.003 \times 10^4, \quad \alpha = -0.568 \times 10^{-1}, \\ b &= 1.006 \times 10^4, \quad \beta = -0.559 \times 10^{-1}, \quad c = 28.00\,. \end{aligned} \tag{8}$$

The long time value is attributable to background noise fluorescence; however, a function of the form of (7) could not generate the two stationary points which appear in the experimental data of Figure 2(a), thus the fit is made from $t = 15\,\text{s}$ on the basis that the data prior to this time is unreliable. The initial value of the fluorescence was captured by constraining the sum of the coefficients. Least squares fitting does not provide a visually satisfactory solution and so adjustments were made to the parameters to capture the important features of the measured data plots.

For PI(3)P there is a significant difference in the time intervals over which the reactions take place. The reaction involving PI(3,4,5)P_3 ceases after 60 seconds whereas that for PI(3)P has hardly commenced. This means that the exponential constants r and s will have much greater magnitude than

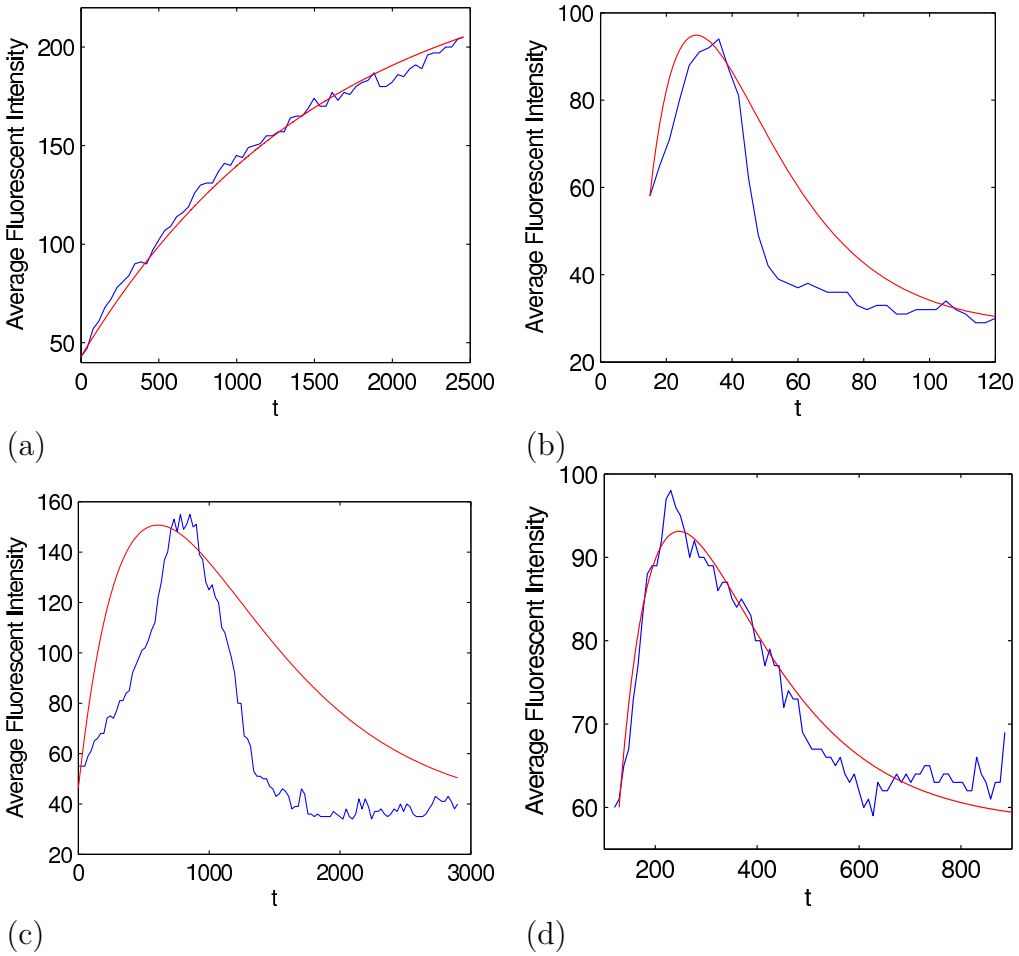
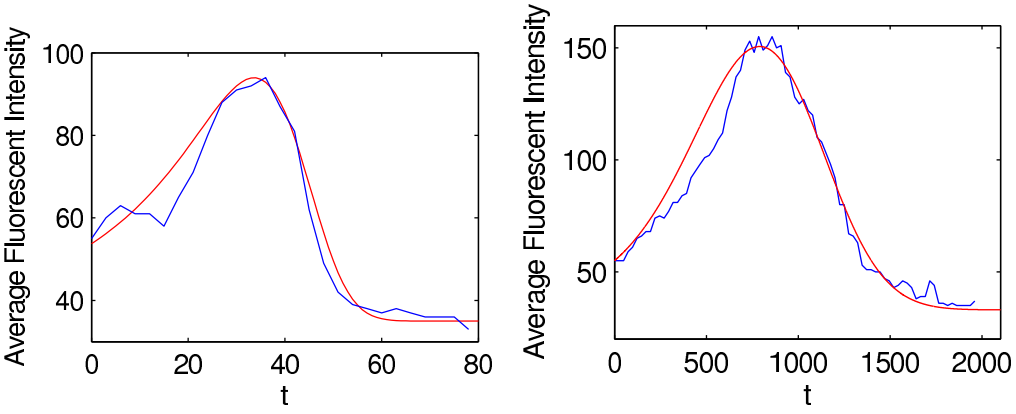


FIGURE 3: From left to right and top to bottom these plots are (a) mutant PIKfyve (with PI(3,5)P_2 blocked) , (b) PI(3,4,5)P_3, (c) PI(3)P, and (d) PIKfyve.



(a) (b)

FIGURE 4: Plots of (a) PI(3,4,5)P_3, and (b) PI(3)P fitted using the modulated exponentials of section 4.

k_4 and k_6. It implies that the reaction for PI(3)P is largely uncoupled from that of PI(3,4,5)P_3, and suggests that the behaviour of PI(3)P can be represented by a linear combination of two exponentials. The heuristic argument about rates of decay concludes that the lesser magnitude exponential coefficient will account for the large time behaviour. We commence by fitting $b\exp(-\beta t)$ to the data over the values for large time, having subtracted out the long time equilibrium values. A second exponential term is then added and the function, with b and β perturbed perhaps, fitted. The technique was used to obtain the fits shown in 3(c) . The equation of the fit is equation (7) with

$$\begin{aligned} a &= -2.2756\times 10^5, \quad \alpha = -1.5922\times 10^{-3}, \\ b &= 2.2757\times 10^5, \quad \beta = -1.5904\times 10^{-3}, \quad c = 36.50\,. \end{aligned} \tag{9}$$

The enzyme PIKfyve is monitored since its presence facilitates the production of PI(3,5)P_2 and thereby gives information on that compound. Taking the data from $t = 129\,s$ to $t = 886\,s$ the plot in Figure 3(d) is obtained. The

constants are

$$\begin{aligned} a &= 1.644 \times 10^{2}, \quad \alpha = -6.134 \times 10^{-3}, \\ b &= -1.624 \times 10^{2}, \quad \beta = -1.095 \times 10^{-2}, \quad c = 58.00\,. \end{aligned} \tag{10}$$

The investigations of PI(3,4,5)P_3 and PI(3)P leads to several conclusions. First, in both examples the two exponential rate constants are almost equal. This is consistent with the requirement that each term contributes significantly to the fitted function. In order to capture the observed slopes of the data the coefficients of the exponential terms must have opposite signs. Next, there can be considerable variation in the constants of proportionality although, to achieve a balance in the terms across the entire time scale, the magnitudes of their values are always similar. Since the relationship between the fluorescence and mass, say, is not known, the coefficients are not expected to provide information concerning these aspects of the model. We conclude that the plots only provide information about the temporal behaviour of the data.

The model used here is a simplified version of a more complex model and further simplification does not appear possible. Thus, since the rate constants for PI(3,4,5)P_3 and PI(3)P are of different orders of magnitude this can only be reconciled with the presence of PI(3,4)P_2 as an intermediary in the linear model.

When determining fits of equation (7) by first fitting the long time tail for b and β, the values of the exponential constants showed little variation. Tests were made on PI(3)P using the data up to $t = 80\,\mathrm{s}$—beyond this point the reaction has probably terminated. The data when PI(3,5)P_2 is blocked may be fitted with a single exponential. The eigenvalue corresponding to the blocked configuration is precisely k_4 thus we have its value immediately.

Since the constants r and s are the roots of equation (4) we can draw some important conclusions concerning k_1, k_2 and k_3. Since all k_i are nonnegative, D is nonnegative. Thus D can only be zero if, and only if, k_2 is zero and $k_1 = k_3$.

Denote

$$s = k_1 + k_3 \quad \text{and} \quad d = k_1 - k_3\,, \tag{11}$$

and then

$$D = (k_2 + s)^2 - (s^2 - d^2) = 2sk_2 + {k_2}^2 + d^2. \tag{12}$$

Since each term in this expression is nonnegative, D is small if and only if each term is small.

On the grounds of continuity, we use the heuristic argument that if the roots are nearly equal (which is an experimental observation) then so too are k_1 and k_3. An estimate of k_2 is obtained from equation (12) by setting $k_1 = k_3$ and discarding k_2^2:

$$k_2 = |D|/2s = |\alpha_1 - \alpha_2|/[2(\alpha_1 + \alpha_2)] \approx 0.003697\,. \tag{13}$$

4 Data fits with modulated exponential functions

While displaying many of the broad features and time scales of the data, the above fits are not particularly close to the experimental data. Hence we also investigated other function forms to fit to the experimental data. Burrage observed that Hill functions, similar to $h(x)$ in equation (14), are often utilised in biological modelling of inhibition and activation. Even though the functions introduced below have not been determined from the current model, the quality of the fits merits presentation and warrants further investigation.

The fitted plots of PI(3,4,5)P_3 and PI(3)P in Figure 4 have the general form

$$y = a\exp\bigl[kxh(x)\bigr] + g \quad \text{where} \quad h(x) = \frac{1 - cx^p}{1 + dx^p}\,, \tag{14}$$

with k and p both nonnegative. Note that the constant g is not part of the model, it merely describes the background fluorescence. When g is removed

TABLE 1: Parameters for the modulated exponential functions of equation (14).

Component	a	c	d	g	k	p
PI(3,4,5)P_3	18.75	3.63×10^{-9}	3.63×10^{-10}	35	0.041	5
PI(3)P	22	5.123×10^{-7}	5.123×10^{-8}	33	0.00322	2

TABLE 2: Summary of the results; the units are in seconds and 1/seconds.

Component	Active Time interval	rate exponents	implied rate constants
PI(3,4,5)P_3	$[15, 80]$	$\alpha = -0.170$, $\beta = -0.167$	$\alpha = k_1$, $\beta = k_3$
PI(3)P	$[0, 2000]$	$\alpha = -0.0015922$, $\beta = -0.0015904$	$\alpha = k_4$, $\beta = k_6$
PIKfyve	$[129, 645]$	$\alpha = -0.0061$, $\beta = -0.011$	
PIKfyveblocked	$[0, 2000]$	$\alpha = -0.00064$	$\alpha = k_4$

the function y is a single exponential modulated by $h(x)$. For x small, $h(x)$ is unity; whereas for x large and positive, $h(x)$ is close to $-c/d$. This is a very versatile function as the constants a and g determine y at zero and infinity. The remaining parameters can be used to capture intermediate behaviour. Because the argument of the exponential function changes sign, equation (14) captures the curvature of both PI(3,4,5)P_3 and PI(3)P at small and large times. The parameters whose values appear in Table 1 were determined on a discrete set $0, 1, \ldots, N$ and scaled to the time intervals corresponding to each experimental plot.

5 Conclusions

In the absence of direct measurement of the whole process we nevertheless construct a complete model of the reactions. The values of α and β in the expression (7) and the time intervals on which they were determined appear in Table 2. The constants of proportionality are not given since their values may not be of significance in the current analysis.

The rate constants from the model which correspond to α and β are also listed. The constant k_2 was estimated by a speculative analysis given in section 3, equation (13). The rate k_5 has not been estimated; however, it is probably of minor importance. Two values of k_4 have been estimated by two independent processes. They are of similar orders of magnitude, differing by a factor of 2.5. Experimental variation could account for this difference and further data from repetition of the experiments may help. A second experiment that would be helpful in understanding this difference would be the simultaneous imaging and quantification of PI(3,4,5)P2 and PI(3)P to give more detail of the k_3 and k_4 rates in a single experiment. We did have this data for one experiment, but the imaging proved difficult to quantify reliably due to a high density of overlapping endosomes in the region of interest. We hope to obtain further experimental data in the future.

Acknowledgements Thanks to Dr. Markus Kerr for the generous use of his data and invaluable discussions and support of the ARC Centre of Excellence in Bioinformatics, The University of Queensland.

References

[1] Y. Mosesson, G. B. Mills, Y. Yarden. Derailed endocytosis: an emerging feature of cancer. *Nature Reviews Cancer* 2008 Nov; 8(11):835–50. doi:10.1038/nrc2521 C157

[2] L. R. Prost, S. Sanowar, S. I. Miller. Salmonella sensing of anti-microbial mechanisms to promote survival within macrophages. *Immunological Reviews* 2007; 219:55–65. doi:10.1111/j.1600-065X.2007.00557.x C158

[3] M. C. Kerr, T. Wang, N. A. Castro, N. Hamilton, L. Town, D. L. Brown, F. A. Meunier, N. F. Brown, J. Stow and R. D. Teasdale.Inhibition of the PtdIns(5) kinase PIKfyve disrupts intracellular replication of Salmonella. *EMBO* 29:1331–1347, 2010. doi:10.1038/emboj.2010.28 C158, C159, C160

[4] D. J. Wilkinson *Stochastic Modelling for Systems Biology* Chapman and Hall. 2006. C161

[5] P. J. Cullen, G. E. Cozier, G. Banting, H. Mellor. Modular phosphoinositide-binding domains—their role in signalling and membrane trafficking. *Current Biology* Volume 11, Issue 21, 30 October 2001, Pages R882–R893. doi:10.1016/S0960-9822(01)00523-1 C161

Author addresses

1. **John A. Belward**, ARC Centre of Excellence in Bioinformatics, The University of Queensland, Australia and the Queensland Consortium for Scientific Computation and Mathematics.
mailto:john.belward@qcscm.com
2. **Kevin Burrage**, ARC Centre of Excellence in Bioinformatics, The University of Queensland, Australia and Comlab, University of Oxford, U.K.
3. **Rohan D. Teasdale**, ARC Centre of Excellence in Bioinformatics, The University of Queensland, Australia.
4. **Nicholas A. Hamilton**, ARC Centre of Excellence in Bioinformatics, The University of Queensland, Australia.