



OPEN

# A model of the PI cycle reveals the regulating roles of lipid-binding proteins and pitfalls of using mosaic biological data

Francoise Mazet<sup>1</sup>✉, Marcus J. Tindall<sup>2</sup>, Jonathan M. Gibbins<sup>1</sup> & Michael J. Fry<sup>1</sup>

The phosphatidylinositol (PI) cycle is central to eukaryotic cell signaling. Its complexity, due to the number of reactions and lipid and inositol phosphate intermediates involved makes it difficult to analyze experimentally. Computational modelling approaches are seen as a way forward to elucidate complex biological regulatory mechanisms when this cannot be achieved solely through experimental approaches. Whilst mathematical modelling is well established in informing biological systems, many models are often informed by data sourced from multiple unrelated cell types (mosaic data) or from purified enzyme data. In this work, we develop a model of the PI cycle informed by experimental and omics data taken from a single cell type, namely platelets. We were able to make a number of predictions regarding the regulation of PI cycle enzymes, the importance of the number of receptors required for successful GPCR signaling and the importance of lipid- and protein-binding proteins in regulating second messenger outputs. We then consider how pathway behavior differs, when fully informed by data for HeLa cells and show that model predictions remain consistent. However, when informed by mosaic experimental data model predictions greatly vary illustrating the risks of using mosaic datasets from unrelated cell types.

The phosphatidylinositol (PI) cycle is a key component of the signaling machinery downstream of receptor protein-tyrosine kinases (RTK) and G protein-coupled receptors (GPCR). The cycle can be found in all eukaryotic cells, is the source of multiple second messengers through the actions of phospholipase C (PLC) and phosphoinositide 3-kinase (PI3K) and is assumed to function the same way in different cell types. The universal nature of the pathway means it is of wide interest, but its multiple components are technically difficult to measure, making it a good candidate to explore using mathematical modelling approaches. There has been a number of prior attempts to model aspects of the PI cycle looking at portions of the signaling cascade using ordinary differential equations (ODEs), informed to a large degree by data from different cell types<sup>1-3</sup>. We have been unable, however, to combine them into a single model of the complete pathway and recapitulate the different published biological outputs. This issue applies to a number of cell signaling models, with their development being hampered by a lack of cell type specific biological data to inform kinetic rate constants and protein concentrations. In particular, we note current cell signaling pathway models often lack cell-specific time course data to inform model parameter values. They also usually incorporate purified enzyme kinetic data which may bear little resemblance to the cellular kinetics<sup>4</sup>, or experimental values obtained from multiple, unrelated, cell types, with reactant concentrations often estimated. Signaling can be cell context dependent producing specific responses to stimuli and “mosaic” models. Although useful in investigating and informing the general biological processes involved, models informed in these ways may not necessarily recapitulate cell specific dynamic behaviors.

We postulate that cell-type specific datasets generated by omics approaches coupled to time-course analysis of the gene expression or protein modifications would provide a more consistent approach to informing mathematical models. This would allow us to focus on determining in-situ reaction kinetic rate constants which once obtained should be possible to use for other cell types as long as specific quantitative proteomic data are available. Here we describe a PI cycle model making use of a quantitative proteomic dataset<sup>3</sup> and experimental

<sup>1</sup>ICMR, School of Biological Sciences, The University of Reading, Whiteknights, Reading RG6 6AS, UK. <sup>2</sup>Department of Mathematics and Statistics, The University of Reading, Whiteknights, Reading RG6 6AX, UK. ✉email: f.m.mazet@reading.ac.uk

phospholipid (PL) and inositol phosphate (IP) time-course data produced in platelets, under similar conditions (Fig. 1A)<sup>6–11</sup>. We use the model to demonstrate the importance of lipid-binding proteins in regulating homeostasis and to inform the regulation of several key proteins in platelets. We next investigate how the model can be used to simulate the PI cycle in other cell types. In doing so we leave our rates unchanged but inform the model with specific cell-type proteomic data (Fig. 1B). Finally, we demonstrate the pitfalls of using mosaic data to inform such a model and how it can lead to erroneous conclusions. We discuss how this limits the use of combining cell signaling models and directing the design of future experiments.

## Results

We first sought to combine previously published mathematical models of the PI pathway in order to construct a model of the pathway in platelets. Whilst the structure of the pathways and their respective mathematical formulations were generally similar, how the respective reaction rate constants and concentrations were informed varied greatly. For instance, values were not consistently reported both in terms of their magnitude and units, it was not always clear how all values had been informed and values to inform cell type specific models had been obtained from data related to different cell types. We also sought to consider a previous model of PI signalling in platelets in response to ADP signalling<sup>1</sup>, which we could extend to model the response of platelets to a variety of GPCR ligands. We found that whilst the model had been useful in informing platelet biology, it was informed using a range of different cell type parameter data and experimental results from both ADP and Thrombin but only considering ADP receptor to trigger the cell activation.

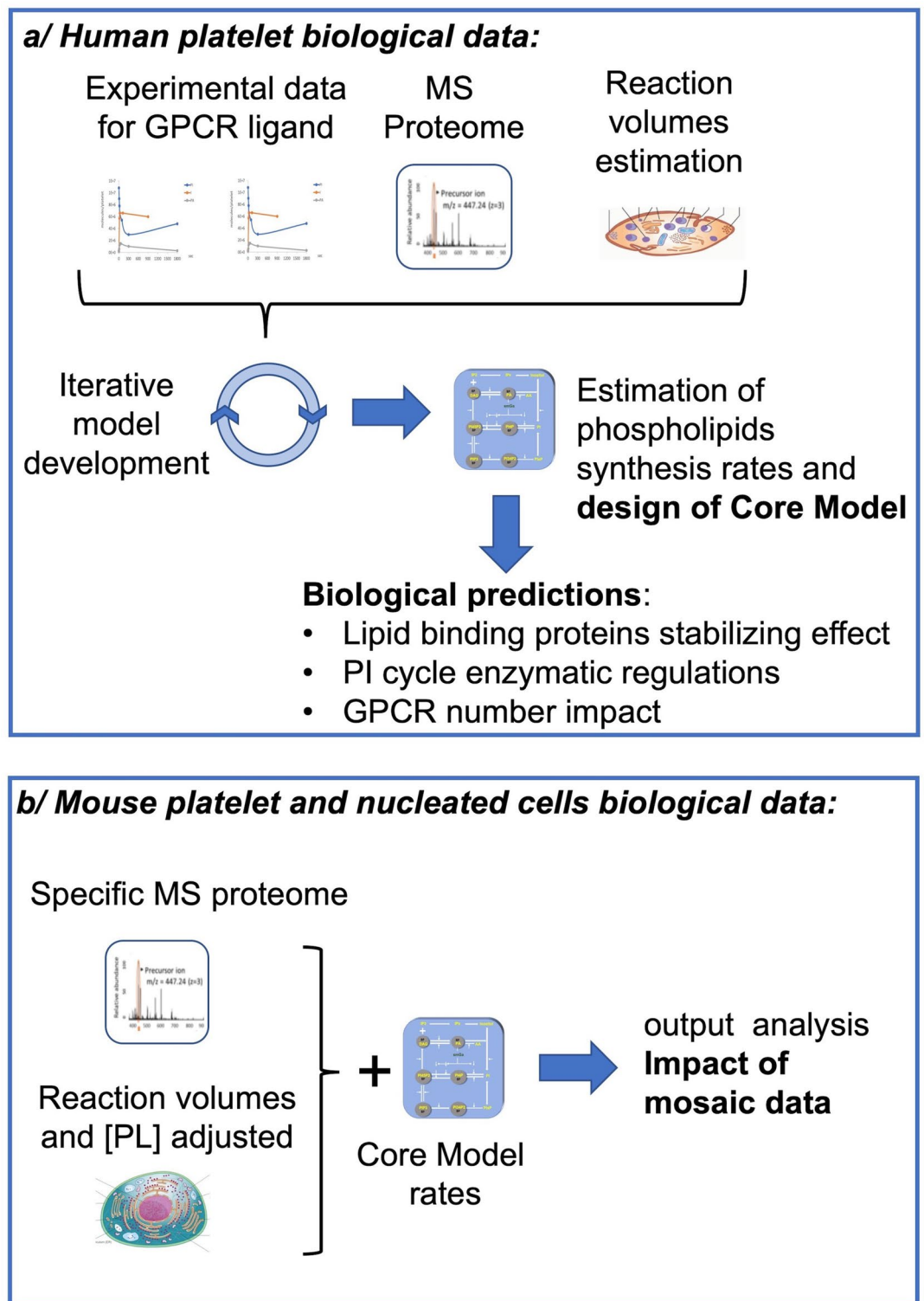
In light of these points and knowing that considerable biological data for informing platelet biology is now available, we thus decided to formulate a complete model of the PI cycle in platelets, exclusively using specific parameters for PI cycle pathway proteins, phospholipid substrates and resulting second messengers (Fig. 2). In respect of platelet specific data, literature mining revealed three quantitative proteomes for human and mouse platelets and HeLa cells that contained data on all the key proteins required for the PI cycle to function<sup>5,12,13</sup>. Further literature mining revealed that detailed sets of time-resolved data on complexes formed in the PI cycle pathway were also available for human platelets (Fig. 3A,C, Supplementary Figure S1 online). Using this information, we first developed a human platelet “Core Model” that focuses on G protein-coupled receptor (GPCR) signaling through Gαq, leading to PLCβ activation and production of inositol 1,4,5-trisphosphate (IP3) second messenger. Phosphatidylinositol (3,4,5)-trisphosphate (PIP3), produced via Gαi and PI3K, and other PIP3-derived PL were also monitored, but their highest levels were systematically two orders of magnitude lower than PL and Inositol (Ins)<sup>9,14</sup> (Fig. 2A, Supplementary Figure S1 and Supplementary Table S1 online). As such, we assumed they are unlikely to significantly alter the PLCβ results and conclusions and subsequently were excluded in the model we developed.

**The core model—the human platelet model.** We constructed an ODE model of the PI cycle in human platelets based on the model shown in Fig. 2A, which was solved in COPASI (v4)<sup>15</sup>. Full details on how the model was developed, parameterized and solved are presented in the Supplementary Information. Briefly, in order to accurately inform the full model parameterization and given the full model consisted of 35 parameter values, we decided to use an iterative approach of reduced models to both decrease the size of the parameter space being determined, whilst increasing confidence in the parameter values determined. We started with the most simplified model (iterations 1 and 2) shown in Fig. 2B, and via model-data fitting using the parameter estimation algorithm available in COPASI<sup>15</sup>, used this to inform the relevant parameter values. This model was then extended in the next iteration with a number of additional reactions, which were chosen so as to not greatly increase the size of the undetermined parameter set. Each model was informed where possible with the previously determined values, whilst new unknown parameter values were again determined. Iterative steps in building up the complexity of our model in this way, allowed us to inform the full model pathway as detailed in Supplementary Tables S1–S5, Supplementary Figure S2 and the Supplementary Information online.

**Phospholipid binding proteins stabilize phospholipids variations.** An important point that arose during the development of the Core Model, was that solely adapting the rates of production or recycling of different PL to switch from non-activated to activated states, and back again did not lead to a biologically realistic solution (Supplementary Figure S3 online). We concluded that the problem was linked to the availability of PL to the enzymes in addition to the rates at which they were being used in the cycle. It is known that PI4P and PI45P2 do not alter dramatically upon stimulation from their homeostatic levels<sup>2,6,16,17</sup>. We hypothesized that this might be due to the presence of PL binding proteins that would sequester membrane phospholipids through protein-lipid interactions, which have been shown to be important for numerous cell functions<sup>18</sup>.

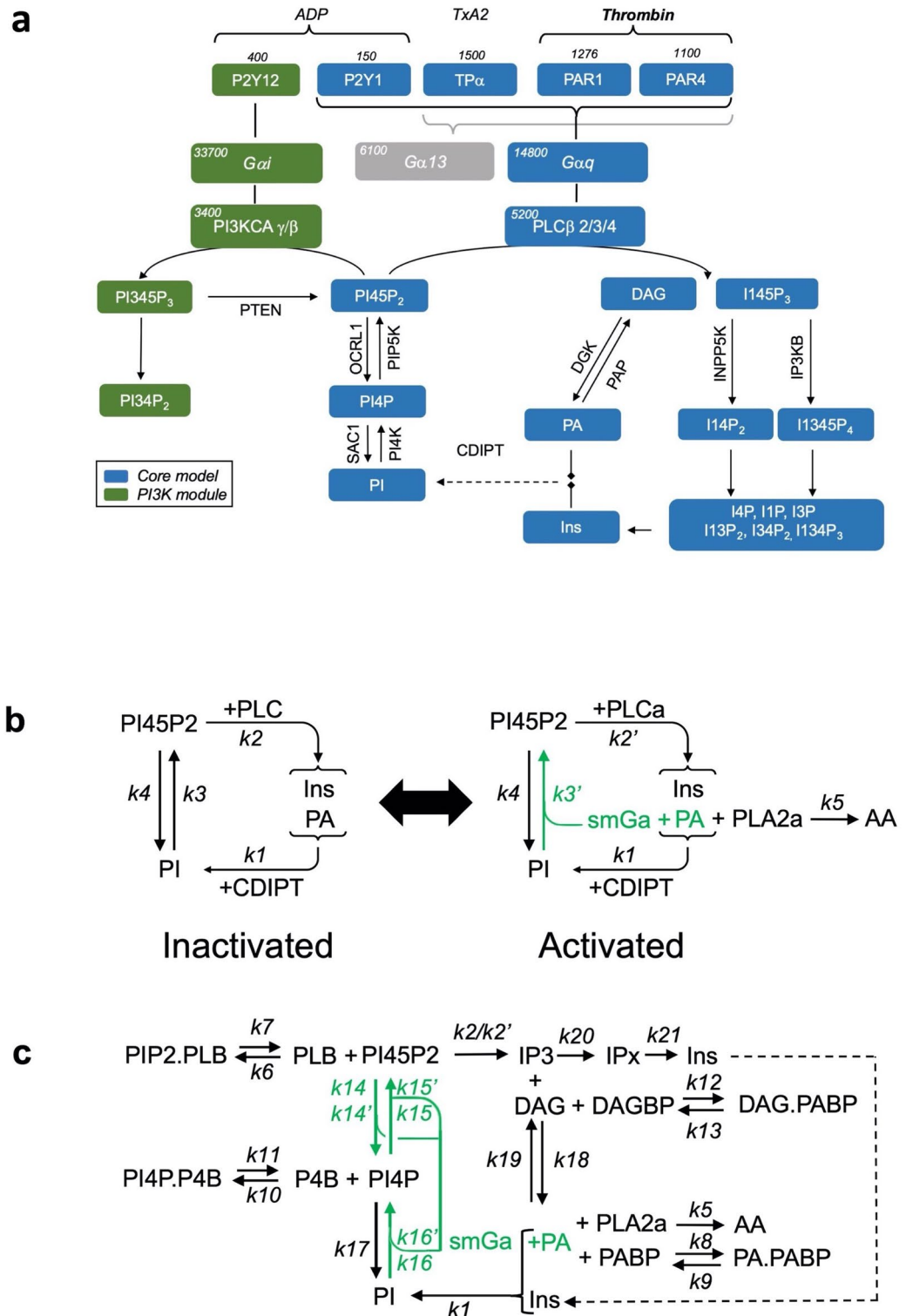
We first focused on PI45P2-binding proteins and added these to our model with reversible binding reactions with PL to control its use by PLCβ (Iteration 3, Fig. 2C, Supplementary Figs. 1–2 online). We performed a series of parameter scans of the amount of binding proteins and average ON/OFF binding rates which predicted that to correctly simulate temporal PI45P2 changes, the number of binding proteins should be around  $1.3 \times 10^6$  per cell. A detailed search for known PI45P2-binding proteins in the proteome dataset<sup>5</sup> revealed, in close agreement with our prediction, a value of  $1.12 \times 10^6$  per cell (Supplementary Table S6 online). Similar methodologies were used for PA, PI4P and DAG, and led to model results matching the experimental data (Fig. 3b,c, Supplementary Figure S3 online).

**Model predictions of the regulation of PI cycle enzymes.** PI45P2 and PI4P homeostatic levels in human platelets are similar, at around  $1.2–1.8 \times 10^6$  molecules per cell respectively, while the estimated amount of plasma membrane PI is around  $6 \times 10^6$  molecules per cell (Supplementary Table S1 online). This suggests unbal-

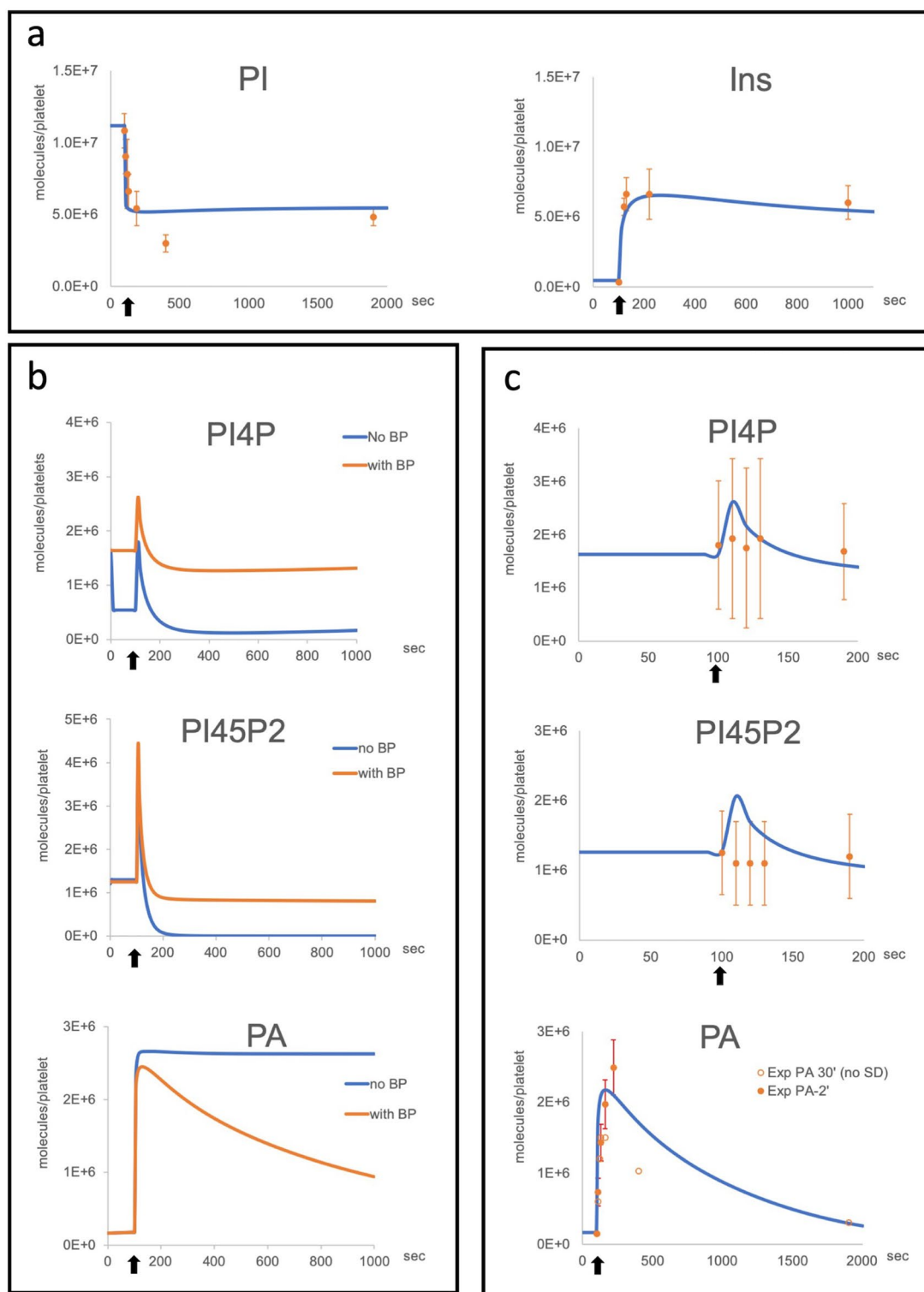


**Figure 1.** Graphical summary of the different modelling steps. **(a)** The core model was produced using proteomics and signaling data (obtained with a single GPCR ligand, Thrombin) obtained for human platelets and used to generate predictions regarding PI cycle driven cell signaling. **(b)** Mouse platelet and nucleated cell proteomic data (Cell type B) were used to populate our Core Model, generate output predictions and analyse the impact of using data from different origins in a single model.

anced phosphorylation/dephosphorylation reactions between PI and PI4P and balanced reactions between PI4P and PI45P2 in inactivate cells. The addition of PI4P, its binding proteins and its metabolizing enzymes PI4K,



**Figure 2.** Schematic illustration of the PI cycle in mammalian cells and model iterations. (a) Schematic of the GPCR ligands and reactions in platelets covered by our model (blue). PI3K-dependent (green) and *Gα13*-dependent (grey) reactions were not included in the core model. *PI* Phosphatidylinositol, *Ins* Inositol, *PIP3* Phosphatidylinositol-3,4,5-trisphosphate, *PI45P2* Phosphatidylinositol-4,5-bisphosphate, *PI34P2* Phosphatidylinositol-3,4-bisphosphate, *PI4P* Phosphatidylinositol-4-phosphate, *IP3* Inositol trisphosphate, *DAG* Diacylglycerol, *PA* Phosphatidic Acid. (b) Early iterations used a simplified Core Model to estimate the phospholipids synthesis rates in inactivated cells (+ PLC) and activated cells (+ PLCα). Rate labels are indicated as either basal rates (*k*) or activated rates (*k'*). (c) The completed Core Model schematic including lipid binding proteins. *k14*=PIP5K, *k15*=OCRL1, *k16*=PI4K, *k17*=SAC1. *k14* to *k17* represent rates before activation, *k14'* to *k16'* represent the rates after activation. *k17* remains unchanged after activation.”



**Figure 3.** Comparison of experimental results and model simulations. **(a)** PI and Ins experimental results (orange dots with SD) and simulations (blue curves) after Thrombin stimulation. Secondary signaling by other, secreted, GPCR ligands is taken into account and assumed to happen immediately. **(b)** Early model iterations simulations showing the stabilizing impact of lipid binding proteins (BP) on the dynamics of PI45P2, PI4P and PA before and after activation with Thrombin. The numbers of BP and the ON/OFF rates of their binding to their relevant PL were determined using parameter scans. **(c)** The final simulations in our completed core model are shown together with the experimental results. Experimental results are shown with standard deviations when available. (References, SD and model parametisation procedures in “Methods” and Supplementary Tables S1–S5). Experimental data:  $n = 10$ , SEM are shown. Activation is indicated by arrow.

OCRL1 and SAC1, could only achieve the correct dynamics of PI4P, PI45P2 and PI observed both before and after activation when the rates of the PI4K and OCRL1 were regulated in a manner similar to PIP5K i.e. all three enzymes have similar low kinetic levels before activation, which are increased upon GPCR activation. In contrast, the SAC1 kinetic rate constant needed to remain unchanged after GPCR activation to recapitulate the correct experimentally observed PL dynamics i.e. a short and synchronised burst for PI4P and PI45P2 within the first 20 s of the activation, followed by a return to their initial levels while PI level fall substantially to half its initial level within seconds and is not replenished significantly for the next 30 min (Fig. 3a,c).

In conclusion, our results suggest that the regulation of the PI cycle in both inactive and activate cells is achieved by mechanisms differentially controlling the enzymes involved. PI4K and PIP5K have been suggested to be scaffolded by proteins at the membrane, exchanging PL almost directly<sup>19</sup>. Our simulations suggest that OCRL1 is also part of this complex or co-localises in the plasma membrane, being then regulated in concert with the kinases. SAC1, however, is regulated differently than the other PI cycle enzymes which leads to the hypothesis that it may not co-localize with them.

The final step of our model development (Iteration 4, see Supplementary Information for details and Supplementary Figure S2D online) was to reintroduce PLC $\beta$  products, IP3 and DAG, as intermediates for Ins and PA. While DAG and PA levels, regulated by two simple reactions and their respective binding proteins could be easily modeled, cytosolic IP3 and Ins levels could not be correctly simulated by a single direct reaction. We were only able to match their respective dynamics by adding an intermediate step which removes IP3 rapidly, reflecting the production of IP4 by the kinase IP3Kb and IP2 by the PL phosphatase INPP5<sup>20</sup>. This is followed by slower production of Ins by a complex set of reversible reactions<sup>21</sup> which we simplified in our model and wrote as a single reaction (Fig. 2a,c, Supplementary Figure S2D online).

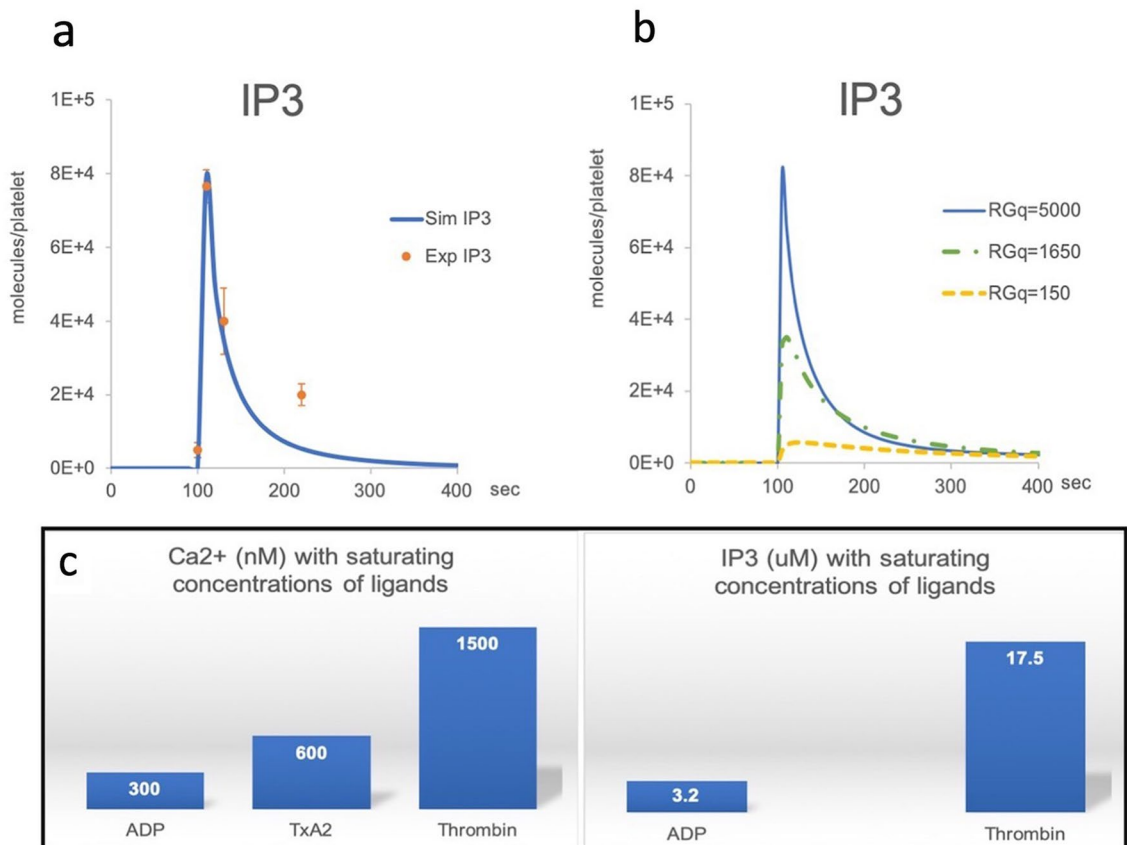
**Gaq-coupled receptor number governs the strength of IP3 production.** We chose to model the PI cycle because of its central role in signalling cascades triggered by a variety of GPCR agonists. One issue with this approach is that the strength of the signalling is highly variable depending on the type of receptor involved, with some researchers arguing that differences in molecular identity, biological activation processes and post-activation recycling of the receptors are responsible for this variability. Kinetic studies, however, demonstrate that GPCR signalling is primarily and rapidly down-regulated at the level of the receptors by phosphorylation, and by inactivation of their direct partners, the G proteins, and the RGS and PLC $\beta$  proteins<sup>22,23</sup>. Together the evidence suggests that the strength of the GPCR signalling is actually a function of receptor abundance<sup>1,24</sup>.

The experimental data we used to inform our model kinetic rate constants were obtained after Thrombin activation of the platelets. However, the experimental methodologies used by the different authors suggest that secondary GPCR signalling through secreted ADP and TxA2 release was also activated<sup>6–11,16</sup>. We originally designed our Core Model to account for the activation of all their respective G $\alpha$ q-coupled receptors, namely PARI/4 for Thrombin, P2Y1 for ADP and TP $\alpha$  for Thromboxane (Fig. 4a,b, RGq = 5,000, Supplementary Table S7 online). By solely reducing the receptor numbers to simulate only ADP activation via the P2Y1 receptor (Fig. 4b, RGq = 150), we were able to simulate the experimentally observed IP3 output supporting the claim that, in the case of G $\alpha$ -coupled receptors, the strength of the signalling is indeed related to the number of receptors being activated<sup>25</sup> (Fig. 4b, Supplementary Table S7 online).

We could not find data regarding IP3 levels after TxA2 activation, however, IP3 triggers calcium cellular mobilization and both experimental and mathematical models have shown a relationship between cytosolic IP3 and calcium levels<sup>17,26</sup>. Quantification of calcium in platelets shows that TxA2-mediated activation leads to half of calcium being mobilized compared to Thrombin. ADP activation only triggers a fraction of IP3 and calcium release compared to Thrombin and TxA2<sup>27,28</sup> (Fig. 4c). Interestingly, the number of TP $\alpha$  and P2Y1 receptors which would be involved in a TxA2 primary activation followed by the secreted ADP secondary activation is just under half the number of the full G $\alpha$ q-coupled receptor complement (Supplementary Table S7 online). When modeling the likely IP3 output following the activation of TP $\alpha$  and P2Y1 receptors, we obtained a predicted peak value for IP3 roughly half that obtained with Thrombin (Fig. 4b, RGq = 1,650). This supports the hypothesis that IP3 levels regulate the intensity of calcium release in a GPCR receptor number dependent manner.

**Applying the core model to other cell types.** Given the universality of GPCR signaling and the PI cycle in mammalian cells we assume that the network structure and kinetic rate constants do not vary greatly between platelets and other cell types. Temporal expression profiles of PI45P2 and IP3 in human Purkinje cells, mouse N1E-115 neuroblastoma neural cells or chicken B cells DT-40 have been published and the similarities between them and human platelets suggest those profiles are common features in vertebrate cells<sup>22,17,29</sup>. We thus hypothesised that we should be able to simulate the PI cycle and IP3 production in other cells and obtain similar output dynamic patterns by using cell specific protein initial concentration values, whilst leaving our model structure and kinetic rates unchanged (Fig. 1b).

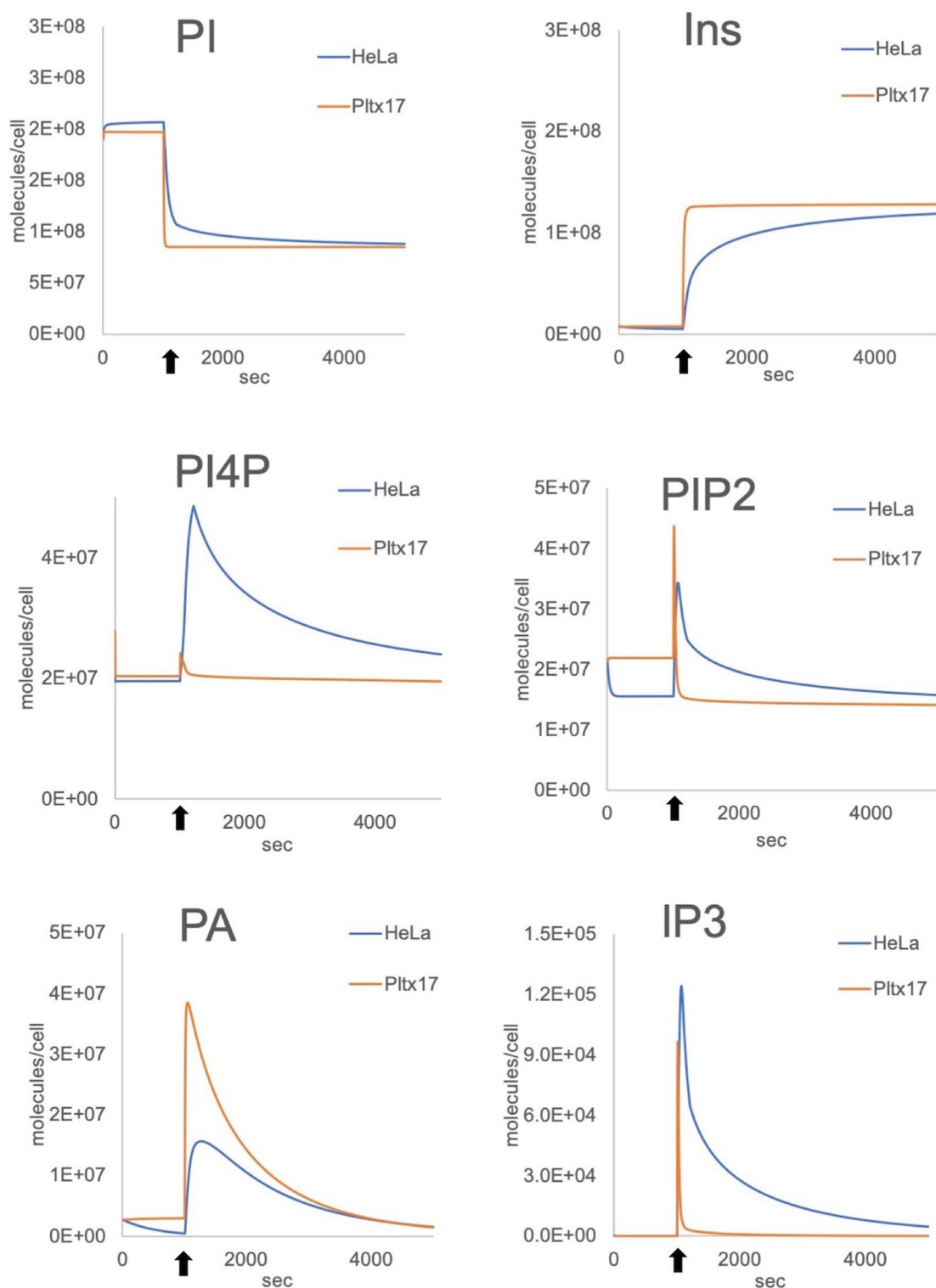
**Application to the mouse platelet using proteome data.** Before proceeding to consider how applicable our model was in nucleated cells, we first used data from the mouse platelet proteome, to check the model behavior<sup>12</sup>. We corrected for the difference in size of the reaction compartments (25% of those in human platelets, Supplementary Table S7 online) but left the ODEs untouched. Despite the protein concentrations being sometimes very different from their human counterparts, the simulations show almost identical temporal behaviors of the different PL and IP as for human platelets and levels in line with the initial PL abundance (Supplementary Figure S4 online).



**Figure 4.** IP3 simulations with differential receptor numbers. **(a)** IP3 experimental results (orange dots with SD) and simulations (blue curves) after Thrombin activation. **(b)** simulations of the impact of Gαq-coupled receptor numbers (RGq) on IP3 production. RGq numbers reflect primary and secondary GPCR ligand activation (Thrombin followed by TxA2 and ADP: RGq = 5,000; TxA2 followed by secreted ADP: RGq = 1,650; ADP alone: RG1 = 150). Activations occur at 100 s. **(c)** comparison of known IP3 and calcium mobilization results in platelets. In absence of experimental data on IP3 production after TxA2 activation, we are using known cytosolic calcium concentrations as a proxy<sup>27</sup>. The simulations of IP3 release from a receptor number corresponding to a TxA2/ADP activation (RGq = 1,650) show an intermediate response as is also seen for calcium mobilization experimental data.

**Application to nucleated cells using cell-specific data.** To demonstrate the wider utility of our model we next created a generic simulated nucleated cell, using an average volume of 2000 fl and calculated the size of the reaction compartments as described earlier. PL, IP and lipid-binding protein numbers were adapted from the human platelet data (Supplementary Table S7 online). We first simulated the behavior of an enlarged platelet by populating our nucleated PI cycle model with the protein copy numbers of a human platelet scaled to match the new reaction volumes (labelled pltx17). Next, protein copy numbers from the epithelial adenocarcinoma HeLa human cell line proteome dataset were used to populate the same model<sup>13</sup>. In order to compare the results and in the absence of common data regarding GPCRs, the activation in these simulations were performed using 85,000 molecules of GPCRs, i.e. corresponding to the same concentration of receptors for a normal human platelet. Given that HeLa cell protein numbers are drastically different from the enlarged platelet simulation, we expected a radically different series of outputs. HeLa cell simulations did indeed show some differences in the peak concentrations of the PLs and IPs we surveyed compared to the large platelet simulation, but overall the temporal dynamics were similar (Fig. 5), suggesting our model solutions are robust to changes in protein concentration and exhibit the correct PLCβ-dependent IP3 signaling response.

**The effect of using mosaic proteomic datasets for informing model parameters.** After demonstrating the portability of our PI cycle pathway model to other cell types and its robustness when parameterised with cell specific protein numbers, we wanted to investigate its response when informed by a mosaic dataset. Using our nucleated cell model, we used a proteomic dataset from the bone osteosarcoma epithelial U2OS human cell lines with only partial data regarding the PI cycle enzymes and missing the concentration of Gαq, IP3 processing enzymes (IP3Kb and INPP5), OCRL1, PI4K and cPLA2 proteins<sup>30</sup> (Supplementary Table S7 online). HeLa protein concentrations were used to inform the missing values. The resulting simulations led to PI, Ins and PA concentration profiles similar to the HeLa simulations although the concentration of PI45P2, PI4P and IP3 were much lower than expected for a cell of this size. IP3 production is significantly affected and unlikely to lead



**Figure 5.** Application to other cell types. GPCR-PLC $\beta$  outputs simulations using HeLa proteome dataset (blue line) and compared to a hypothetical large platelet (pltx17, orange line). Reactions and kinetic parameters are unchanged from the original core model; reaction volumes, initial PL numbers and their binding proteins were adapted as described in “Methods” and Supplementary Table S7 online. Simulations are run for 5,000 s with activation occurring at 1,000 s (arrows).

to a realistic outcome (Fig. 6). We then utilised a series of parameters scans, to inform those protein concentrations for which values were not available. Whilst we were able to find protein copy numbers which meant Core



Model simulations could replicate the previously modelled HeLa cell behaviour, the use of partial data from a different cell type would not necessarily provide the correct outputs (Fig. 6, Supplementary Table S7 online).

We then considered random combinations of all the protein concentrations from the U2OS and HeLa datasets. The simulations showed no consistency for any PL or IP we surveyed, with 75% leading to incorrect behaviours (Fig. 7). Whilst 25% of the results produced the correct behaviour for IP3, these simulations did not always lead to correct outputs for the other PLs we surveyed such as PI4P or PA. Ultimately this suggests that combinations of protein concentration values may lead to incorrect model approximations of the underlying protein concentrations.

Together these results suggest that combining data from different cell types does not necessarily lead to results that are consistent in simulating PI cycle dynamics. Indeed, it is highly likely that in the majority of cases, model simulations are unlikely to agree with experimentally observed results for specific cell types. To further test this result we undertook a sensitivity analysis to determine the influence of initial protein levels on the production of IP3. This revealed both commonalities and specific patterns for each cell type (Supplementary Figure S5 online). Specifically, similar changes in the concentration of PLC $\beta$ , PIP5K and OCRL1 lead to distinctive IP3 outputs in the different simulated cells.

Thus, we conclude that although our model can be populated with protein concentrations sourced from different cell types, to describe functional outputs, it can lead to erroneous conclusions.

## Discussion

Conventionally, mathematical models of cell signalling pathways have been informed by data taken from a range of cell types. This is often a result of data not being available for a specific cell type to inform all kinetic rate constants and concentrations. Whilst to the seasoned mathematical modeller this might seem intuitive, we demonstrate here that the consequences should not be ignored. Here we developed a biological model of the PI cycle entirely based on a single quantitative proteome and multiple sets of experimental data generated under the same conditions for a single cell type, the human platelet. This allowed us to focus on determining in situ kinetic rate parameters of the reactions governed by Gq-coupled receptors. In addition, and in contrast to most other published cell signalling models, we have considered how the system behaviour varies from an inactive to active state. This has allowed us to reveal a number of mechanisms involved in the maintenance of the observed steady-state before activation.

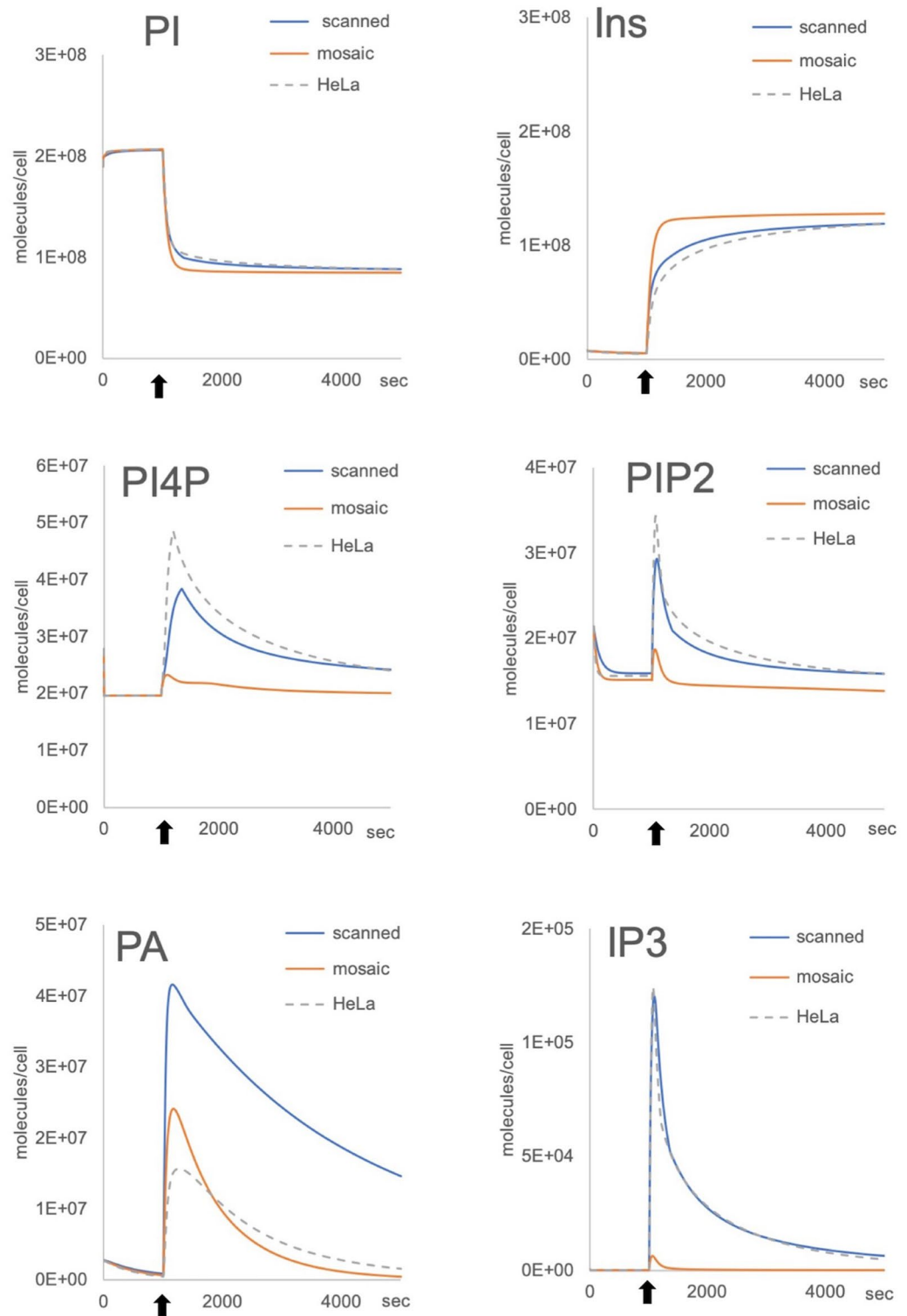
Previous mathematical models have largely ignored the cyclic generation of PI, assuming instead that it was available at all times<sup>1–3</sup> or suggested the presence of an “open” PI cycle where some PLs produced externally can recapitulate both steady-state and activated experimental results<sup>31</sup>. We postulated that signalling events lead to the depletion of PI on the plasma membrane and termination of the signalling, while leaving a pool of PI on ER and Golgi membranes. Using our model, we were able to show that while there is a constant exchange of PI between the different membranes, the rate of plasma membrane PI replenishment does not seem to be modified by activation events. This leads to the conclusion that PI distribution on the different cell membranes, and the rate at which it can be replenished at the plasma membrane, is a major way of regulating the duration of cell signalling.

While it has been suggested that sequestering of membrane lipids by binding proteins is an integral part of cell homeostasis and activation<sup>18,31,32</sup>, the involvement of PL-binding proteins in cell signalling regulation is, however, usually considered only after cell stimulation. We demonstrated that PL-binding proteins also play a stabilising role prior to signalling events, functioning as cellular sinks and limiting the availability of lipids to modifying enzymes such as PLC and PI3K; effectively inhibiting signalling processes. After signalling is triggered, PL-binding proteins also regulate the maximum amount of available PL for reactions leading to intermediates such as IP3 and DAG. Simulations of PI45P2-binding proteins correspond to the peak quantity of PI45P2 monitored experimentally and we predict that this is likely to be true for all other PL-binding proteins.

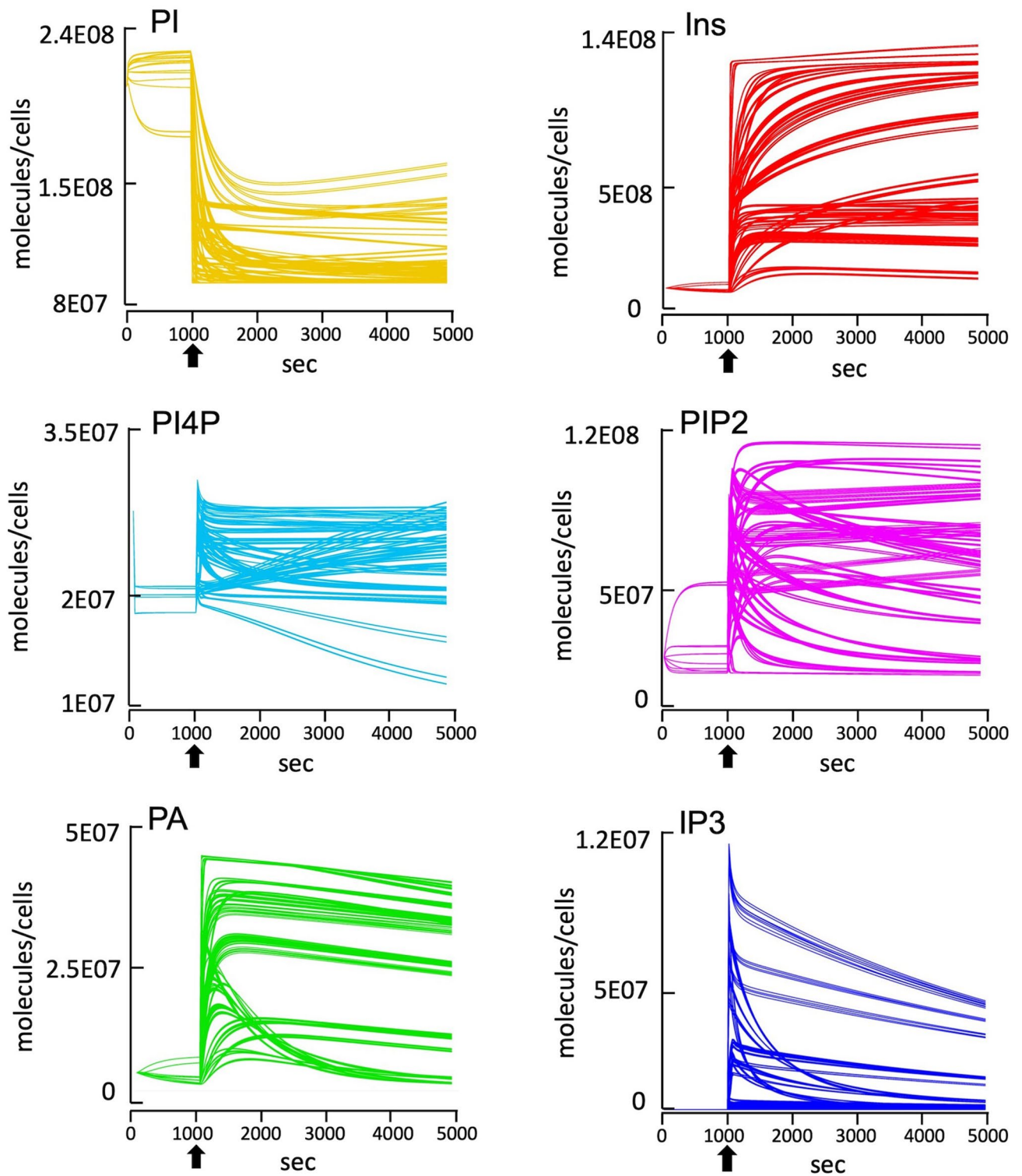
Next, our model, formulated using thrombin-stimulated platelet data, was able to recapitulate known IP3 outputs for other GPCR ligands by simply replacing the receptor numbers for each ligand with their respective known values. These results support the still-debated hypothesis that the number of GPCRs, rather than their molecular identity, regulates the intensity of this type of signalling<sup>1,24</sup>.

We also used our model to produce a series of predictions regarding the regulation of the major enzymes involved in the central PI45P2 -PI4P -PI axis of the PI cycle. We predicted that three out of the four enzymes, namely PI4K, PIP5K and OCRL1 have their activities up- and down-regulated in concert. The co-location of PI4K and PIP5K has already been shown to occur experimentally<sup>19</sup> and the use of a protein complex containing PIP5K and PI4K shown to successfully replicate experimental observation in mathematical modelling<sup>3</sup>. We now suggest adding OCRL1 to this complex. In contrast, to explain the differential abundance of PI and PI4P/PI45P2 as well as the distinct response of the phosphatase SAC1 to the activation signal, we predict this enzyme to be isolated from its counterparts. SAC1 has been shown to be restricted to discrete regions of contact between the plasma and ER membranes<sup>33</sup>. Interestingly, CDPIT, otherwise known as PI-synthase, is also shown a lack of up-regulation after signalling<sup>34</sup> and is known to be located at cell membrane contact points between the ER and plasma membranes. It is likely that the PI4K-PIP5K-OCRL1 complex is recruited by either the receptors or some of their downstream effectors. There is no physical connection between the location of the complex and the membrane contact regions containing the different enzymes regulating the regeneration of PI and its location on the different membranes.

Due to the lack of detailed protein concentration and reaction rate constant values for individual cell types, the use of data obtained from multiple cell types is common in cell signalling models. Kinetic rates are often sourced from experiments performed with purified enzymes or adapted from previous modelling attempts. While mathematical models informed using ‘mosaic data’ have been crucial in understanding mechanisms that underlie processes within cells, combining data from different mathematical models of the same pathway are



**Figure 6.** Simulations with mosaic proteome datasets. GPCR-PLC $\beta$  outputs simulations comparing HeLa (dashed line) and 2 U2OS datasets (“scanned” or “mosaic”). Because the values for PI4K, OCRL1, G $\alpha$ q, cPLA2 and IP3 modifying enzymes are missing from the U2OS proteome, we performed a series of scans to estimate the missing protein numbers (“scanned”, see Supplementary Table S7 for details) that allow results similar to those obtained with the HeLa dataset (“HeLa”) and compared them to a mosaic dataset created by using HeLa protein numbers to replace the missing U2OS numbers (“mosaic”). The results show that the mosaic dataset generates outputs for PI45P2 (PIP2), PI4P and IP3 are unlikely to lead to a cell signaling response. Simulations are run for 5,000 s with activation occurring at 1,000 s (arrows).



**Figure 7.** Simulation results of systematic “mix-and-match” of protein numbers between HeLa and U2OS cells proteomic data. The copy numbers of the 12 key proteins surveyed in our model (Gaq, smG, PLC $\beta$ , the combined IP3 modifying enzymes, DGK, LPP, CDIPT, OCRL1, PI4K, PIP5K, SAC1 and cPLA2) were taken from either the HeLa or U2OS proteome (12, 34), or from the calculated numbers for U2OS missing data shown in supplementary Table S7, and systematically mixed using Parameter Scans in COPASI 15. 75% of the 4,096 simulations obtained lead to incorrect dynamic behaviours for the outputs monitored with our model. Simulations are run for 5,000 s with activation occurring at 1,000 s (arrows), #molecules per cell, s seconds.

often difficult. This is because model formulations often differ, meaning parameter dimensions do as well. This is further compounded by the fact that data is not always available to inform all model parameters, meaning uninformed cell specific parameters are often determined using similar processes in other cell types or are simply estimated. We hypothesised that fully informing a mathematical model of a pathway using cell type-specific data would lead to more accurate predictions of the pathway dynamics. We demonstrated that our PI model of a human platelet demonstrated similar behaviour when informed by mouse platelet data, where the kinetics were assumed the same, but protein concentrations varied. In extending the model to that applicable to a nucleated HeLa mammalian cell, similar observations were made. However, maintaining the same kinetic rate constant values but using protein concentrations obtained from two different cell types led to widely varying dynamical predictions for PL and IP3. Our sensitivity analysis of the impact of protein levels on the production of IP3 suggested that each cell has a specific balance of PL/IP modifying enzymes working together to lead to the correct behaviour. With only a small number of proteomic datasets available at this time, we have been unable to identify how this balance is established. We believe that further information from other nucleated cells will be needed to understand it.

The long-term goal of biological modelling is to recapitulate processes that govern cell behaviour. Our experience in modelling the PI cycle was that there was too much variability in kinetic reaction rate constants and protein concentration published from prior partial models to build a comprehensive model of the pathway. In order for parameterisation of mathematical models for specific cell-types to be more consistent, the collection of comprehensive experimental data informing the concentration of cellular components and their dynamic behaviour needs to occur. Currently, such values are generally informed by traditional methodologies such as western blotting or lipid chromatography, both of which are limited in the number of molecules they can simultaneously characterise. They also suffer from a lack of uniformity and reproducibility. Omics technologies are now available to process high numbers of molecules using highly standardised protocols, which should allow for such issues to be overcome<sup>35</sup>. Extensive quantitative proteomic, metabolomic and lipidomic datasets for commonly available cell lines already provide a valuable resource for modellers. Given the number of signalling pathways in eukaryotic cells that are being considered for mathematical modelling, the rapid expansion of cell-specific omics generated data sets, are likely to become a central part of ensuring mathematical models of signalling pathways are quantitatively well informed.

## Methods

Further details of the biological rationale and of the description of the different iterations of the model are available in the Supplementary Information document.

**Estimation of compartment sizes.** Our core model consists of 3 compartments: the plasma membrane, the cytosol and the organelles. The volume occupied by the platelet plasma membrane, including the open canicular system (OCS) inside the cytoplasm, was calculated to be around 1 fl. Based on published observations, the size of the cytoplasm was reduced as the observed volume of platelets is filled by organelles such as vesicles, ER remnant, mitochondria and the OCS estimated to occupy up to 40% of the internal volume. Most of the remaining volume contains cytoskeletal proteins and glycogen granules and is calculated to occupy around 1 to 2 fl<sup>37,41,42</sup>. We estimated the mouse platelet plasma membrane and cytosol volumes to be around 0.25 fl each while the nucleated cells average plasma membrane and reaction cytosol volumes at around 17 fl each based on a 2000 fl overall cell volume.

**Initial parameters.** Our model takes into account the concentration of protein and lipids relevant for each reaction. Data for protein copy numbers were obtained from a human platelet proteome<sup>5</sup> (Supplementary Table S2 online) while initial and post-activation time-resolved data for the Phospholipids (PL) and Inositol Phosphates (IP) were collated from several publications<sup>6–10,14,16,25,43–48</sup> (Fig. 3, Supplementary Table S1 and Supplementary Figure S1 online). For collated pools of proteins, the UniProt database was first mined using the PL names, followed by a search of the quantitative proteome using the UniProt codes (Supplementary Table S2, S6 and S7 online). Although their binding affinities for the different PL are likely to be variable, we parameterised the different on/off rates on the assumption of average kinetics.

**Estimating kinetic rate constants.** We used the free software COPASI (v4)<sup>15</sup> to build our model starting from a reduced basic network of reactions. All molecules were considered as “well-mixed” inside each compartment. Mass Action kinetics were assumed in modelling the respective reactions. The governing equations were solved using the deterministic method (LSODA) solver. Reaction enzymes were not written as modifiers as their concentrations were important for our model. Activation by ligands were written as events after the start of the simulation. All enzymatic activations were terminated by either the production of an inactive protein (inactivation) or by the return to the initial basal activity state (reset). For each reaction the default rate was first selected then a series of parameter scans were performed, starting at  $\pm 4$  orders of magnitude until the values were producing time-resolved curves for each and every PL and IP output in the model matching the experimental data using a fit-by-eye. Steady-state analysis and Time Series Sensitivity Analysis were performed on all reactions that occur before the activation event, and the protein and PL/IP concentrations adapted accordingly. Parameters sets that led to a deviation of more than 20% from the experimental data were rejected and the full analysis restarted. Schematic diagrams of the reactions and the parameters are shown in Table S3. The computed initial molecule numbers are listed in Table S4. IP3 Time Series Sensitivity Analysis were performed using the in-built “Sensitivities” function in COPASI<sup>15</sup>, looking at the scaled variation of IP3 with respect to the variation of the protein initial concentrations, with the delta factor set at 0.2.

## Data availability

The model was deposited in the BioModel database<sup>49</sup> under the accession number MODEL2006300001.

Received: 2 June 2020; Accepted: 24 July 2020

Published online: 06 August 2020

## References

- Purvis, J. E., Chatterjee, M. S., Brass, L. F. & Diamond, S. L. A molecular signaling model of platelet phosphoinositide and calcium regulation during homeostasis and P2Y1 activation. *Blood* **112**, 4069–4079 (2008).
- Xu, C., Watras, J. & Loew, L. M. Kinetic analysis of receptor-activated phosphoinositide turnover. *J. Cell Biol.* **161**, 779–791 (2003).
- Oliveira, D. V. *et al.* A mathematical model of the phosphoinositide pathway. *Sci. Rep.* **8**, 3904 (2018).
- Zotter, A., Bäuerle, F., Dey, D., Kiss, V. & Schreiber, G. Quantifying enzyme activity in living cells. *J. Biol. Chem.* **292**, 15838–15848 (2017).
- Burkhardt, J. M. *et al.* The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways. *Blood* **120**, e73–82 (2012).
- Wilson, D. B., Neufeld, E. J. & Majerus, P. W. Phosphoinositide interconversion in thrombin-stimulated human platelets. *J. Biol. Chem.* **260**, 1046–1051 (1985).
- Bell, R. L. & Majerus, P. W. Thrombin-induced hydrolysis of phosphatidylinositol in human platelets. *J. Biol. Chem.* **255**, 1790–1792 (1980).
- Broekman, M. J., Ward, J. W. & Marcus, A. J. Phosphatidylinositol and phosphatidic acid in stimulated platelets. *J. Biol. Chem.* **256**, 8271–8274 (1981).
- Kucera, G. L. & Rittenhouse, S. E. Human platelets form 3-phosphorylated phosphoinositides in response to a-thrombin, U46619, or GTPγS\*. *J. Biol. Chem.* **265**, 5345–5348 (1990).
- Mujalli, A. *et al.* Profiling of phosphoinositide molecular species in human and mouse platelets identifies new species increasing following stimulation. *BBA Mol. Cell Biol. Lipids* **1863**, 1121–1131 (2018).
- Nakashima, S., Suganuma, A., Matsui, A. & Nozawa, Y. Thrombin induces a biphasic 1,2-diaclyglycerol production in human platelets. *Biochem. J.* **275**, 355–361 (1991).
- Zeiler, M., Moser, M. & Mann, M. Copy number analysis of the murine platelet proteome spanning the complete abundance range. *Mol. Cell. Proteom.* **13**, 3435–3445 (2014).
- Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548–548 (2011).
- Morris, J. B., Hinchliffe, K. A., Ciruela, A., Letcher, A. J. & Irvine, R. F. Thrombin stimulation of platelets causes an increase in phosphatidylinositol 5-phosphate revealed by mass assay. *FEBS Lett.* **475**, 57–60 (2000).
- Hoops, S. *et al.* COPASI—a COMplex PATHway Simulator. *Bioinformatics* **22**, 3067–3074 (2006).
- Rittenhouse, S. E. & Sasson, J. P. Mass changes in myoinositol trisphosphate in human platelets stimulated by thrombin Inhibitory effects of phorbol ester. *J. Biol. Chem.* **260**, 8657–8660 (1985).
- Brown, S.-A., Morgan, F., Watras, J. & Loew, L. M. Analysis of phosphatidylinositol-4,5-bisphosphate signaling in cerebellar purkinje spines. *Biophys. J.* **95**, 1795–1812 (2008).
- McLaughlin, S., Wang, J., Gambhir, A. & Murray, D. PIP2 and proteins: Interactions, organization, and information flow. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 151–175 (2002).
- Choi, S. *et al.* Agonist-stimulated phosphatidylinositol-3,4,5-trisphosphate generation by scaffolded phosphoinositide kinases. *Nat. Cell. Biol.* **18**, 1324–1335 (2016).
- Pattni, K. & Banting, G. Ins(1,4,5)P3 metabolism and the family of IP3-3Kinases. *Cell. Signal.* **16**, 643–654 (2004).
- Berridge, M. J. & Irvine, R. F. Inositol phosphates and cell signalling. *Nature* **341**, 197–205 (1989).
- Ryu, S. H. *et al.* Feedback regulation of phospholipase C-β by protein kinase C. *J. Biol. Chem.* **265**, 17941–17945 (1990).
- Rajagopal, S. & Shenoy, S. K. GPCR desensitization: Acute and prolonged phases. *Cell. Signal.* **41**, 9–16 (2018).
- Lohse, M. J. *et al.* Kinetics of G-protein-coupled receptor signals in intact cells. *Br. J. Pharmacol.* **153**, S125–S132 (2008).
- Daniel, J. L., Dangelmaier, C. A., Selak, M. & Smith, J. B. ADP stimulates IP3 formation in human platelets. *FEBS Lett.* **206**, 299–303 (1986).
- Finch, E. A. & Augustine, G. J. Local calcium signalling by inositol-1,4,5-trisphosphate in Purkinje cell dendrites. *Nature* **396**, 753–756 (1998).
- Bye, A., Gibbins, J. M. & Lydford, S. Development of platelet calcium flux assay using Fura-2-am on flexstation 3 reader. *Mol. Dev.* **2**, 2 (2015).
- Murray, R. & FitzGerald, G. A. Regulation of thromboxane receptor activation in human platelets. *Proc. Natl. Acad. Sci. USA* **86**, 124–128 (1989).
- Xu, C. *et al.* A PIP2-derived amplification loop fuels the sustained initiation of B cell activation. *Sci. Immunol.* **2**, 787 (2017).
- Beck, M. *et al.* The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, 549–549 (2014).
- Suratekar, R., Panda, A., Raghu, P. & Krishna, S. Evidence of sinks and sources in the phospholipase C-activated PIP2 cycle. *FEBS Lett.* **592**, 962–972 (2018).
- Martin, T. F. J. PI(4,5)P2-binding effector proteins for vesicle exocytosis. *Biochim. et Biophys. Acta Mol. Cell Biol. Lipids* **1851**, 785–793 (2015).
- Del Bel, L. M. & Brill, J. A. Sac1, a lipid phosphatase at the interface of vesicular and nonvesicular transport. *Traffic* **19**, 301–318 (2018).
- Kim, Y. J., Guzman-Hernandez, M.-L., Wisniewski, E. & Balla, T. Phosphatidylinositol-phosphatidic acid exchange by Nir2 at ER-PM contact sites maintains phosphoinositide signaling competence. *Dev. Cell* **33**, 549–561 (2015).
- Fischer, H. P. Mathematical modeling of complex biological systems. *Curr. Rev.* **31**, 11 (2008).
- van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **9**, 112–124 (2008).
- Platelets and their factors*. vol. 126 (Springer, Berlin, 1997).
- Hammond, G. R. V. *et al.* PI4P and PI(4,5)P2 are essential but independent lipid determinants of membrane identity. *Science* **337**, 727–730 (2012).
- Van den Bout, I. & Divecha, N. PIP5K-driven PtdIns(4,5)P2 synthesis: regulation and cellular functions. *J. Cell Sci.* **122**, 3837–3850 (2009).
- Peng, B. *et al.* Identification of key lipids critical for platelet activation by comprehensive analysis of the platelet lipidome. *Blood* **132**, e1–e12 (2018).
- Van Pannerden, H. *et al.* The platelet interior revisited: Electron tomography reveals tubular alpha-granule subtypes. *Blood* **116**, 1147–1156 (2010).
- Neumüller, J., Ellinger, A. & Wagner, T. Transmission Electron Microscopy of Platelets From Apheresis and Buffy-Coat-Derived Platelet Concentrates. In *The transmission electron microscope—theory and applications* (ed. Maaz, K.) (InTech, Berlin, 2015). <https://doi.org/10.5772/60673>.

43. Siegl, A. M., Smith, J. B., Silver, M. J., Nicolaou, K. C. & Ahern, D. Selective binding site for [3H]Prostacyclin on platelets. *J. Clin. Investig.* **63**, 215–220 (1979).
44. Baurand, A. *et al.* Inhibition of platelet function by administration of MRS2179, a P2Y1 receptor antagonist. *Eur. J. Pharmacol.* **412**, 213–221 (2001).
45. Ohlmann, P. *et al.* The platelet P2Y12 receptor under normal and pathological conditions. Assessment with the radiolabeled selective antagonist [3H]PSB-0413. *Purinerg. Signal.* **9**, 59–66 (2013).
46. Ramström, S., Öberg, K. V., Åkerström, F., Enström, C. & Lindahl, T. L. Platelet PAR1 receptor density-correlation to platelet activation response and changes in exposure after platelet activation. *Thromb. Res.* **121**, 681–688 (2008).
47. Nurden, P. Immunolocalization of P2Y1 and TPalpha receptors in platelets showed a major pool associated with the membranes of alpha-granules and the open canalicular system. *Blood* **101**, 1400–1408 (2003).
48. Vanags, D. M., Lloyd, J. V., Rodgers, S. E. & Bochner, F. ADP, adrenaline and serotonin stimulate inositol 1,4,5-trisphosphate production in human platelets. *Eur. J. Pharmacol.* **358**, 93–100 (1998).
49. Chelliah, V. *et al.* BioModels: Ten-year anniversary. *Nucleic Acids Res.* **43**, D542–D548 (2015).

## Acknowledgements

We thank Dr J. Rudge and Prof P. Dash for useful discussions and advice during the development of the model. This work was funded by the British Heart Foundation (PG/16/20/32074).

## Author contributions

F.M., M.J.T., J.M.G., M.J.F. designed the work and participated to the writing of the manuscript. F.M. conducted the research including data collection and creation of the model.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-70215-7>.

**Correspondence** and requests for materials should be addressed to F.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020