A comprehensive mechanistic model of adipocyte signaling with layers of confidence

William Lövfors[1,2], Cecilia Jönsson[1,3], Charlotta S. Olofsson[4], Elin Nyman[1,*,**], and Gunnar Cedersund[1,5, *,**]

[1]Department of Biomedical Engineering, Linköping University, Linköping, Sweden
[2]Department of Mathematics, Linköping University, Linköping, Sweden
[3]Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden
[4]Department of Physiology/Metabolic Physiology, Institute of Neuroscience and Physiology, The Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
[5]Center for Medical Image Science and Visualization (CMIV), Linköping University, Linköping, Sweden
[*]Shared last author
[**]Corresponding author

March 11, 2022

Abstract

Intracellular signaling in adipocytes in an expanding adipose tissue is central to type 2 diabetes. We have earlier developed detailed mathematical models for well-studied signaling pathways in adipocytes, normally and in type 2 diabetes. Compared to the whole intracellular activity, these models still only contain a fraction of signaling pathways. On the other hand, we have large-scale phosphoproteomic data for the insulin response of adipocytes as well as prior knowledge about their interactions associated with a confidence level. However, methods to combine detailed models with large-scale data with varying confidence are lacking. In our new method, we first establish a core model, based on current detailed knowledge on adipocyte cellular signaling with focus on: 1) lipolysis and fatty acid release, 2) glucose uptake, and 3) the release of adiponectin. We use the core model as input to the first layer of our comprehensive model and continuously add new layers with a parallel, pairwise approach with low computation time. We find that the first 15 layers (60 added phosphosites) with highest confidence can predict unseen inhibitor-data (70-90 % correct) and that this ability decrease when we add layers of decreasing confidence. We can in total add 60 layers (3926 phosphosites) and still keep predictive ability. Finally, we use the comprehensive adipocytes model to simulate systems-wide alterations in adipocytes in type 2 diabetes. This new method provide a tool to create large models that keeps track of diverse confidence.

Introduction

A major challenge within biomedical research, from clinical studies to drug development, is to handle and make optimal use of the increasing amounts of data. The handling of large data is today mostly done within the fields of bioinformatics, where data are mapped to static networks. The information in these networks are found in databases with known interactions of different level of confidence. Methods within bioinformatics cannot be used to simulate time-resolved scenarios, e.g. to predict the clinical effect of treatment or the change in signaling inside cells in response to new drugs. Such methods also lack the ability to integrate data from several sources into a consistent body of knowledge. This is instead the strength of systems biology models that uses ordinary differential equations (ODEs).

ODE based models are the most common framework to model biological systems were there is time-resolved data available [1]. These models include mechanistic details about e.g. signaling pathways, and need to be fed with values for included kinetic rate constants. These rate constants are model parameters that are usually unknown since they can often impossible to measure explicitly, and must instead be estimated based on indirect time-resolved measurements.

There are methods to handle semi-large protein data using ODE models [2–5]. These approaches use information rich multi-perturbation data to create data-driven ODE models. These methods have been used to predict the response of new drug candidates and rank their ability to overcome drug resistance. Even though these methods can handle a large number of parameters, for all possible interactions between states, they are limited to ∼200 proteins since all possible interactions are tested. Also, the information-rich perturbation data used in these methods is rare and costly.

The real challenge with large ODE models is to handle large amount of unknown parameters and estimate their values. This challenge has been approached using adjoint sensitivities [6] which improves scaling with the number of parameters, together with the use of a sparse linear solver. The authors [6] speed up the optimization process ∼40,000 times for a model with ∼4000 free parameters. This development is important to go from smaller to larger ODE models. However, methods developed to estimate parameters for large ODE models, still cannot handle the whole omics scale.

One approach to avoid the estimation problem is to use logical parameter-free models instead of ODE models. Such models have been used to study large protein-activity datasets [7, 8]. Logical models are useful in the search for biologically relevant interactions between proteins, and can thus be used for hypothesis generation when it comes to new interactions. However, logical models cannot be used to predict the effect of new drugs/perturbations or to simulate scenarios, another important aspect of biomedical research.

Another solution to the large estimation problem is to divide the problem into smaller sub-problems, adding a single protein/gene (and corresponding data) at a time, preferably starting from an established core model. The additions of new proteins/genes can then be done in parallel, substantially reducing the computation time. We have earlier developed a method to do such an iterative model expansion [9]. However, this method does not take the confidence of included interactions into account and can only add a single layer of new interactions.

For methods that divide the problem into sub-problems and adding a single interaction or molecule at a time, the establishment of a core model is key. Within adipocyte signaling,

there are several different models for different aspects of the adipocyte, such as glucose uptake, lipolysis, and adiponectin secretion [10–15]. The adipocytes of an expanding adipose tissue are central to understand the mechanisms behind type 2 diabetes, and several of the mentioned models are developed using not only data from adipocytes from non-diabetics, but also from type 2 diabetic patients [10–12, 15]. However, there exist no core model including all these different aspects of adipocyte signaling.

Here, we present a method to create comprehensive mechanistic models with layers of confidence, starting with a core of detailed and well-studied interactions. From there, we add large-scale data in layers starting from high confidence interactions to purely data-driven interactions (Fig. 1). When adding these layers, we use a parallel, pairwise approach with a manageable computation time. The biological system we focus on is adipocyte cellular signaling, due to the relevance for type 2 diabetes.

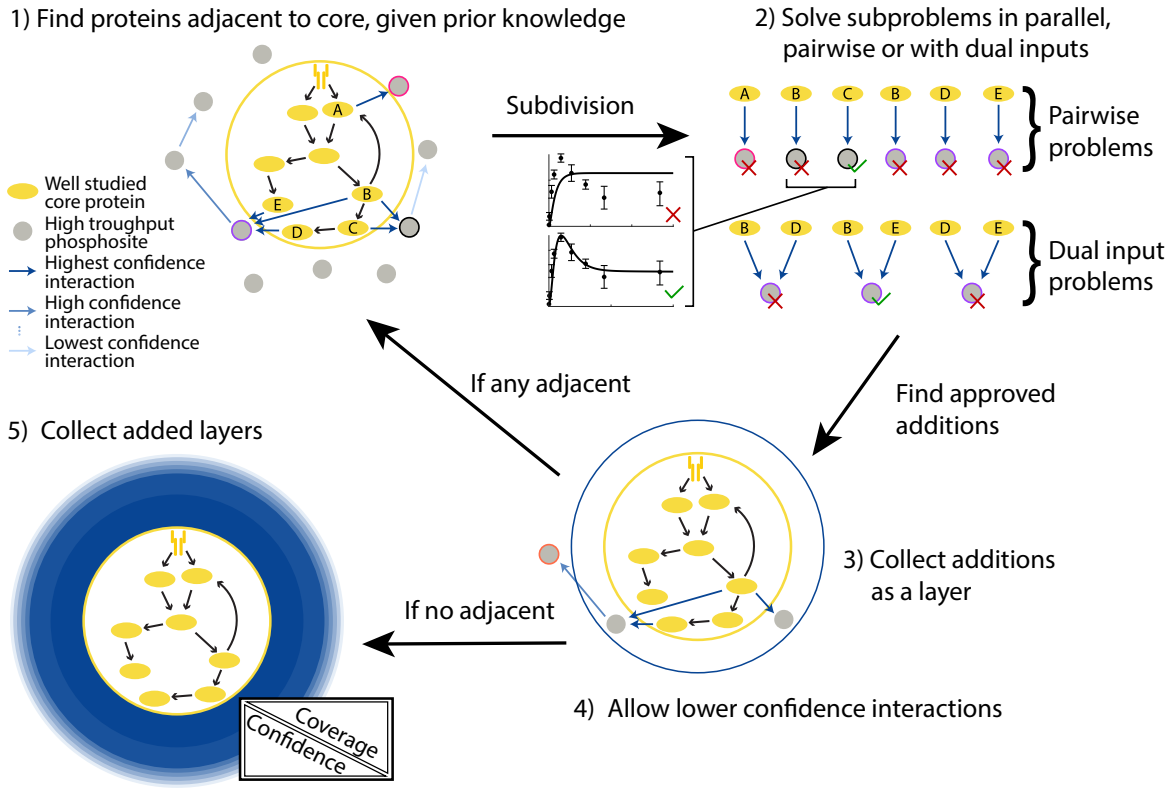

Figure 1: **Overview of the developed approach.** At the the core of the method is our connected model. Outside of the core model, there exists phosphoproteomic data (grey circles) not covered by the model. Using lists of interactions and our automatic model expansion algorithm we are able to add parts of this phosphoproteomic data for proteins adjacent to the core model to an expanded model. These additions are then subsequently used to further add proteins to the expanded model. The core combined model has the highest degree of confidence, but has the lowest coverage of the phosphoproteome. For each added layer, a larger portion of the phosphoproteome is covered but with a lower confidence.

Results

We first established a core model of adipocyte signaling (Fig. 2), based on our earlier work [11, 14, 15]. The included earlier works are the following: 1) a lipolysis model [15] that describes the release of fatty acids and glycerol in response to α- and β-adrenergic receptor agonists and inhibitors, as well as the anti-lipolytic effect of insulin. The lipolysis model includes intracellular signaling intermediaries and data from both non-diabetic and diabetic patients. 2) a glucose uptake model [11] that includes major insulin signaling pathways of adipocytes. The glucose uptake model also include data from both non-diabetic and diabetic patients. 3) an adiponectin-release model [14] that unravel the mechanisms of adiponectin release in response to epinephrine, a β3-adrenergic receptor agonist, as well as intracellular signaling intermediaries. This adiponectin-release model includes patch-clamp experiments in the 3T3-L1 adipocyte cell line, as well as confirming studies in adipocytes from non-diabetic patients.

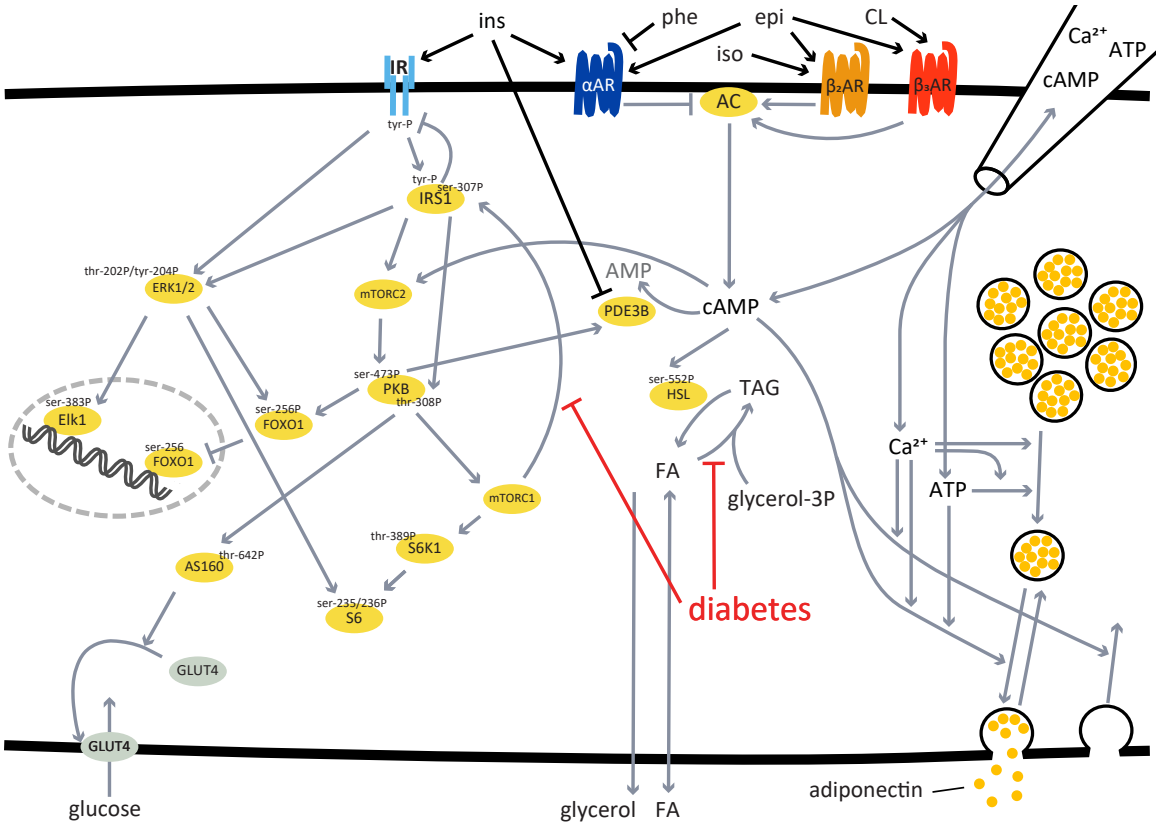

Figure 2: **The core model of adipocyte signaling**. The core model includes crosstalk between glucose uptake in response to insulin (ins), fatty acid (FA) and glycerol release in response to α- and β-adrenergic receptor signaling and adiponectin secretion in response to cAMP, ATP and Ca2+ (added inside the cell with a pipette). The model includes key differences between signaling normally and in type 2 diabetes. phe - phentolaline, iso - isoprenaline, epi - epinephrine, CL - CL 316243.

The crosstalk between different pathways in adipocyte signaling

To connect the models, we looked for crosstalk in between the included signaling pathways (Fig. 2). There are clear crosstalk between glucose uptake and lipolysis, through the signaling pathways of both insulin and α- and β-adrenergic receptors. A central node for this crosstalk is cAMP. cAMP is stimulated by β-adrenergic receptors and inhibited by α-adrenergic receptors, and in turn affects the phosphoinositide 3-kinase inhibitor (PI3Kα) [16]. PI3Kαis also involved in the insulin – glucose uptake signaling network. Furthermore, stimulation with insulin (at physiological concentrations) will lead to an activation of phosphodiesterase 3B (PDE3B), which directly breaks down cAMP. cAMP is also a central mediator for adiponectin secretion [13].

The creation of a core model from previous models and data

The three separate models [11, 14, 15] were connected in two steps. Firstly, we connected the lipolysis model [15] with the glucose uptake model [11]. This was done by replacing the Ins_1 effect and the protein kinase B (PKB) equations from the lipolysis model with the insulin receptor and PKB from the glucose uptake model, and by having cAMP from the lipolysis model activate the mammalian target of rapamycin complex 2 (mTORC2) in the glucose uptake model. Secondly, we connected the newly combined lipolysis–glucose uptake model with the adiponectin model [14]. In essence, we combined the model by letting the β3-adrenergic receptors from the adiponectin model activate adenylate cyclase (AC) from the lipolysis model instead of directly activating cAMP. All details regarding the connection of the model and the full connected core-model is found in all details in *S2 Model equations and parameter values*.

The core model can reproduce all data

We tested the core model by separating the experimental data used in the previous works [11, 14, 15] into a training set and a validation set. To be able to compare the combined core model to the ingoing individual models, we divided the data in the same way as done in the previous works. In practice, this meant that we used the data for the β3-adrenergic receptor agonist CL 316243 and ATP stimulated adiponectin release [14], and the data for the phosphorylation of HSL in response to stimulation with insulin and isoprenaline [15] for model validation, and thus removed this validation data from the set of training data. After estimating the model parameters using the training data, we found the best agreement to be acceptable, supported by a χ^2-test ($v^*_{train} = 616.50 < \chi^2(p = 0.05, dgf = 582) = 639.23$). We gathered parameter sets where the cost to the original estimation data was below the threshold of the χ^2-test. Of the collected parameter sets, the best parameter set with respect to the validation data was ($v^*_{train} < \chi^2(p = 0.05, dgf = 582)$). We found that the model agreement to the validation data was acceptable ($v^*_{val} = 20.65 < \chi^2(p = 0.05, dgf = 16) = 26.30$). The full description and graphical representations of the model agreement to data is given in the supplemental file *S1 Recreation of the model agreement to data and validations*.

After the model validation, we included both the training and validation data into an extended set of data. We tested if the model could explain the extended set of data sufficiently well. We found the model agreement to the extended data to be acceptable ($v^*_{tot} = 654.81 < \chi^2(p = 0.05, dgf = 598) = 656.00$) and thus used this model as a core for further extensions. The model agreement to the extended data is shown in Figs. 3 to 5

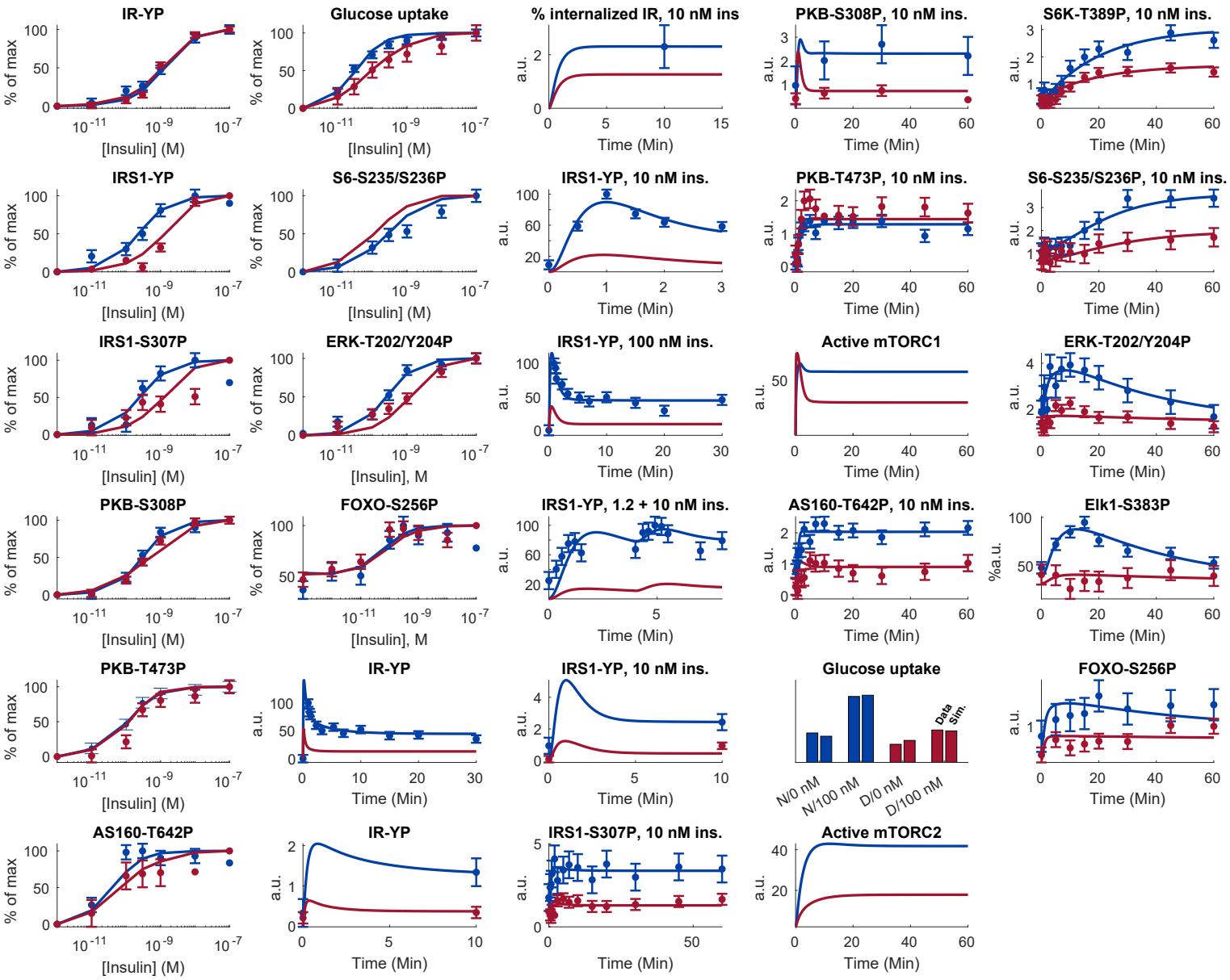

Figure 3: **Model agreement to the data from the previous insulin signaling work.** Data comes from isolated human adipocytes stimulated with insulin in different doses and for different times [11]. In all panels, lines represent the model simulation with the best agreement to data and experimental data points are represented as mean values with error bars (SEM). Simulations and experimental data in red corresponds to experiments under diabetic conditions, and in blue under non-diabetic conditions.

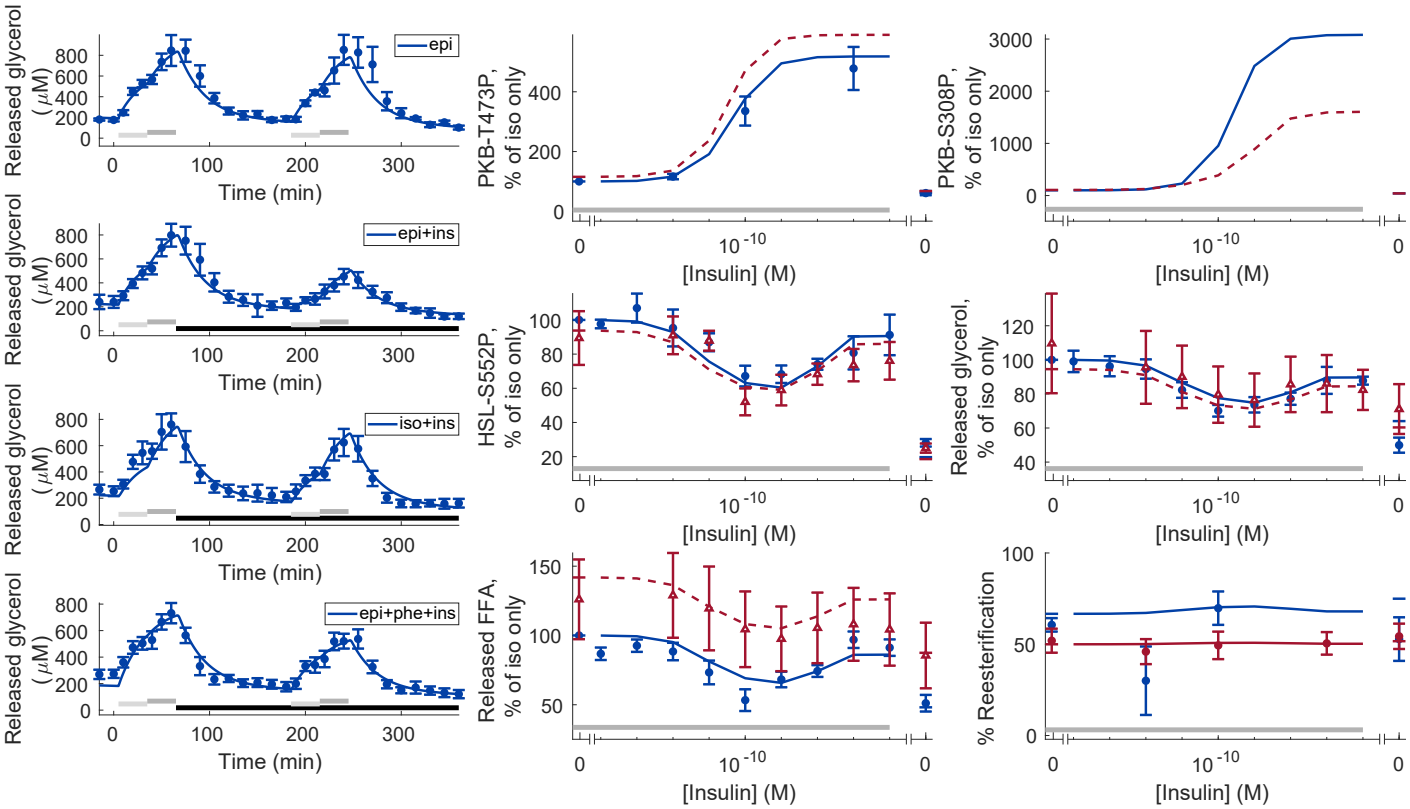

Figure 4: **Model agreement with the data from the previous lipolysis work.** Released glycerol (left) was measured using microdialysis in the adipose tissue in situ [17]. All other data (middle, right) was measured in isolated human adipocytes [16]. In all panels, lines represent the model simulation with the best agreement to data and experimental data points are represented as mean values with error bars (SEM). Simulations and experimental data in red corresponds to experiments under diabetic conditions, and in blue under non-diabetic conditions. Dashed lines and experimental data with open triangles were not used to estimate the model parameters. epi - epinephrine, ins - insulin, iso - isoprenaline, FFA - released fatty acids

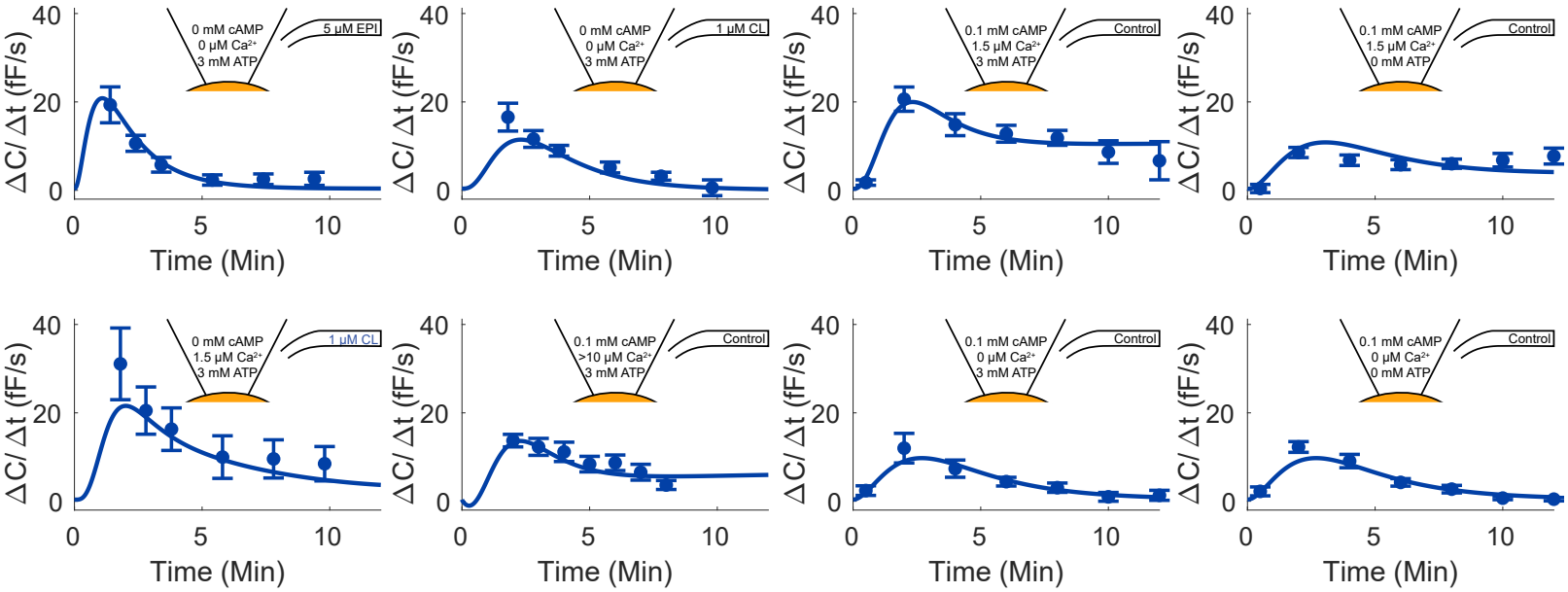

Figure 5: **Model agreement with the data from previous adiponectin release modeling work.** Data represent patch-clamp capacitance recordings in 3T3-L1 adipocytes [15]. In all panels, lines represent the model simulation with the best agreement to data and experimental data points are represented as mean values with error bars (SEM). EPI - epinephrine, CL - β3-adrenergic receptor agonist CL 316243, Ca - Calcium.

From the core model to the comprehensive mechanistic model of adipocyte signaling

To go from the well-studied core model to the comprehensive adipocyte, we use phosphoproteomic data for the time-resolved insulin response in 3T3-L1 adipocytes [18]. Prior knowledge on possible protein–protein interactions was collected using the python package OmniPathDB [19], which compiled a list of interactions from several sources, such as Reactome and PhosphoSitePlus. The list includes a confidence level in the sense of number of primary sources that have studied each interaction.

We divided the phosphoproteomic data into two sets as done by the authors in [18]: one set with sites responding to insulin stimulation and another set with sites not responding to insulin stimulation, based on if the sites are effected by inhibitors of PI3K and PKB. We refer to these data sets as *responders* or *nonresponders*. For our model expansion we started with the responder data.

Using a pairwise approach, we expand the core model towards the comprehensive mechanistic model in multiple layers (Fig. 1). We refer to the additions of the adjacent sites as the addition of a *layer* of sites to the model. For the first layer, we start with the core model and add phosphorylation sites by identifying the sites that are adjacent to the core model given prior knowledge about interactions. We start with interactions of highest confidence, i.e. with the most (20) primary sources. From that level of confidence, we test in a random order if the phosphorylation sites can be added using interaction from the core model as input. Once no more adjacent proteins could be added to the layer, we used the last added layer as inputs to find new adjacent sites. This layered addition of sites was continued until no more sites could

be added to the model.

Once no more sites could be added to the model, we also included the interactions with the next to highest number of primary sources and attempted to add adjacent sites. We continued to include interactions with a lower confidence in a stepwise manner until no more sites could be added even with the interactions with the lowest number of primary sources. Using the responder data and all interactions, we added 189 sites to the model.

We then used all of the data (including the nonresponders), and again started with the interactions with the highest number of primary sources. Again, we step-wise included interactions with a lower number of primary sources. Using all data and all interactions we could add 2146 additional sites, to a total of 2335 sites. When no more interactions could be added using all the data and all the interactions, we continued to add sites not yet included in the model using a data-driven approach. Here, we created a new list of potential interactions, based on the agreement between the simulations of the sites in the expanded model with the experimental data of the sites that had not been added. Using the data-driven approach we could add 1591 additional sites, to a total of 3926 sites.

Figure 6: **The comprehensive model of adipocyte signaling.** The network structure of the expanded model, color coded by the order addition for the 60 added layers in blue. The core model is shown in yellow. Additions with high confidence is shown in dark blue, and additions with lower confidence in light blue. The full network is also available in *S4 Cytoscape file of the expanded model*. The simulations and experimental data for all phosphosites in the model is available in *S3 Time series of all sites added to the model*.

Predicting inhibitions

To test the validity of the expanded model, we simulated the change in the insulin response with a PKB inhibitor and with a PI3K inhibitor and compared to experimental data for the insulin response in the presence of the PKB inhibitor MK2206 and the PI3K inhibitor LY294002 from [18]. Since the model was expanded using both interactions and data with different levels of confidence, we evaluate the model's predictive capabilities on a layer-to-layer basis. The simulated time series of all phosphorylation sites and the effect of the inhibitors can be found in *S3 Time series of all sites added to the model*. The model is able to correctly predict the direction of inhibition with MK and LY in a majority of all sites, in all expansion layers, shown in Fig. 7. However, it should be noted that the model predictions are the most accurate when the most confident interactions and data is used, and that the accuracy of the predictions fall

as the confidence is decreased. Since the model expansion is done using the most confident interactions and data first - and then gradually decrease in confidence - it is clear that the prior knowledge of high confidence helps guide the model to the most accurate predictions. Both the decrease in confidence of interactions (left to right in Fig. 7A,B,D,E) and the inclusion of nonresponder data (Fig. 7A,B vs Fig. 7D,E) leads to a decrease in prediction accuracy.

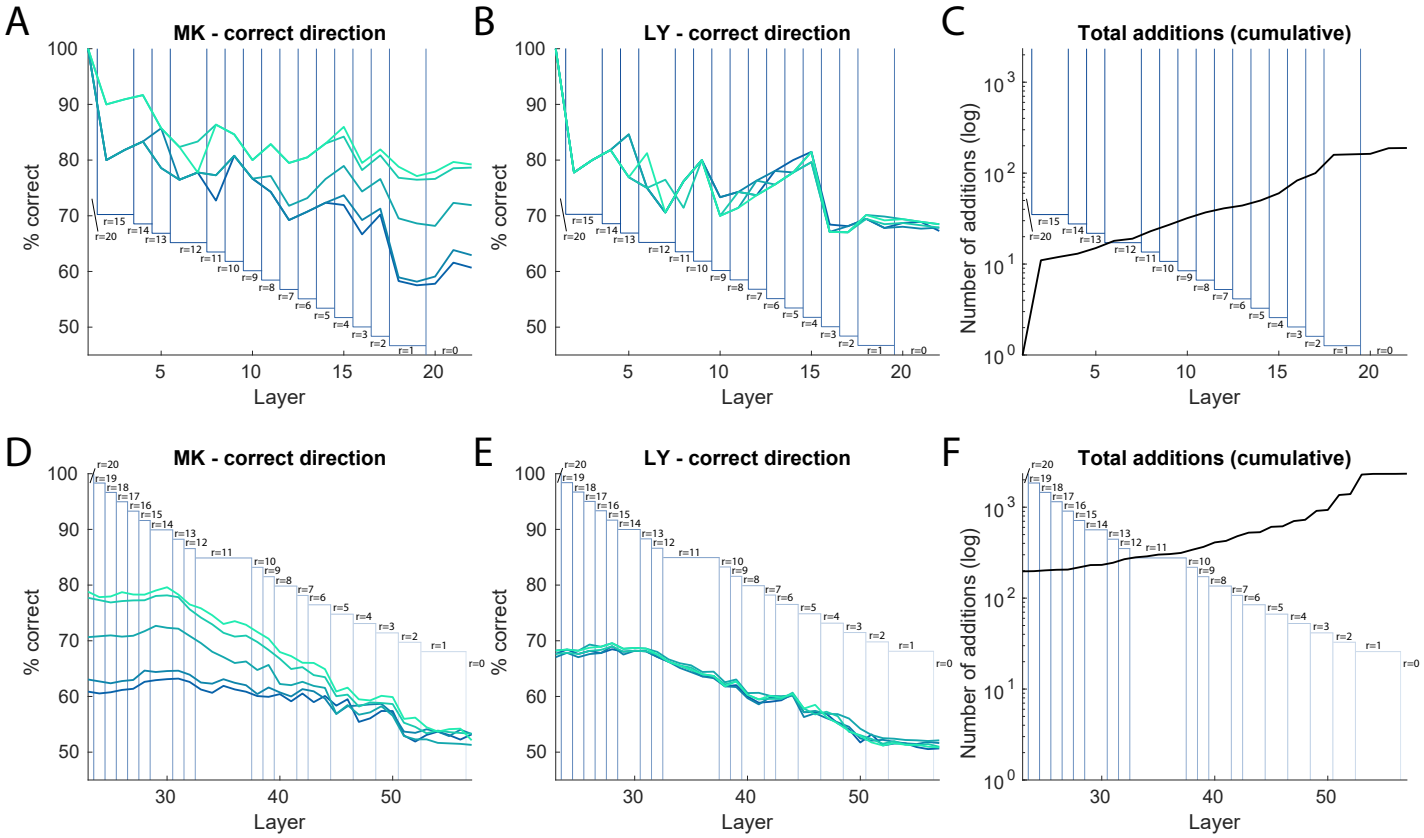

Figure 7: **The predictive ability of the comprehensive model.** For each added layer, we compute the predictive ability of the model using independent inhibitor data. The left column shows to the ability of the model to predict the correct direction of the response for added phosphosites at t=20 min for the PKB inhibitor MK2206. The middle column shows to the ability of the model to predict the correct direction for of the response added phosphosites at t=20 min for the PI3K inhibitor LY294002. The right column shows the total number of added phosphosites. The top row corresponds to layers where only the responder data were included and the bottom row to layers where all experimental data were used. Horizontal graded blue lines correspond to when additional interactions were allowed, where r=x corresponds to x number of primary sources being needed for the inclusion of an interaction (ranging from 20 to 0 number of primary sources).

Simulating diabetes

Since the combined model used as a core for the automatic model expansion could simulate diabetes, we can now simulate how to the effect of type 2 diabetes would spread to a large portion of the phosphoproteome. We evaluated the effect of diabetes by simulating the same

insulin stimulated experiment as done in the phosphoproteomic data under both normal and diabetic conditions. The time-resolved comparison between normal and diabetes is explicitly shown for the first 5 layers in Fig. 8. We quantified the effect by finding the fold change at $t = 20$ minutes and visualized for the whole model in a network view in Fig. 9. The simulated time series of all phosphorylation sites and the effect of diabetes can be found in *S3 Time series of all sites added to the model*.

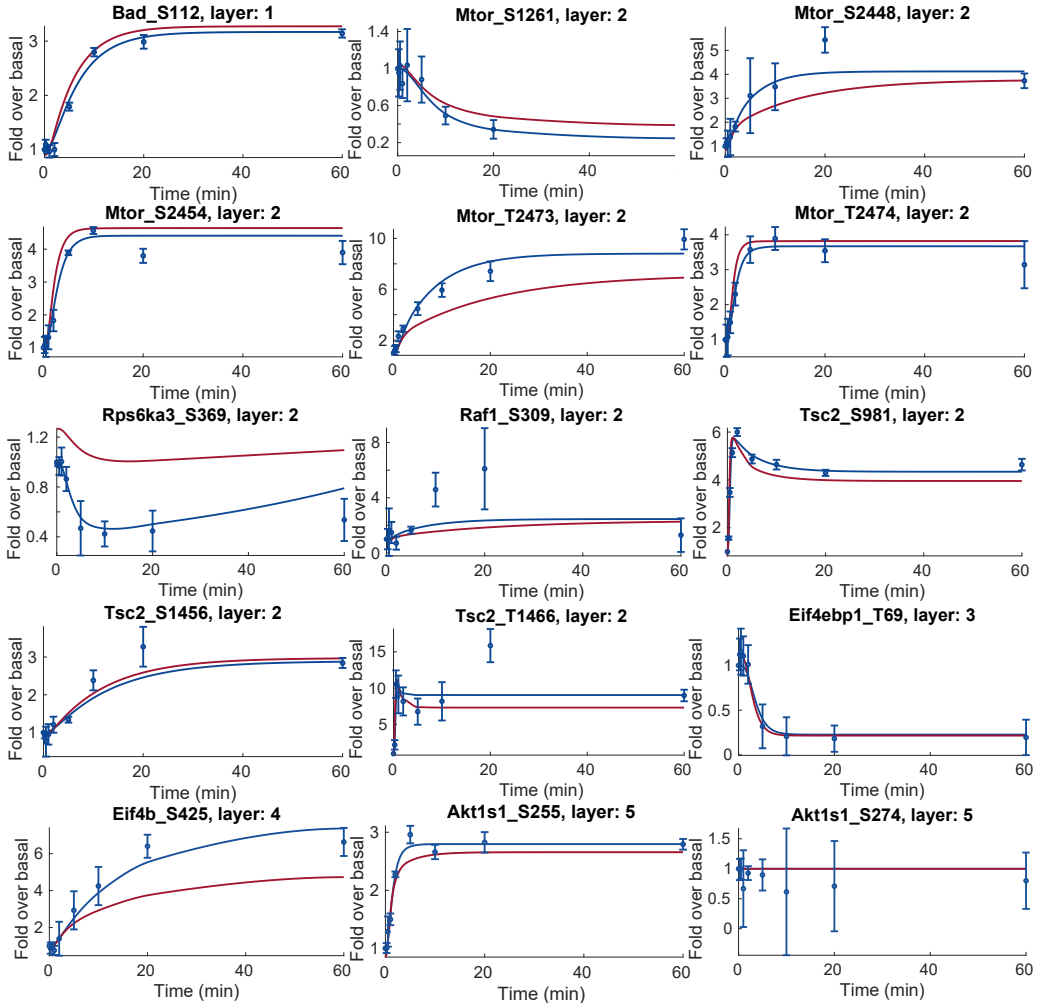

Figure 8: **Simulations and data for the first 5 layers of the comprehensive model.** The first 5 layers contain 15 added sites. In all panels, blue lines represent model simulation with the best agreement to data and experimental data points are represented as mean values with error bars (SEM). Red lines are simulations of the diabetic conditions, and these simulations are model predictions without corresponding data. All model simulations of the effect of diabetes are available in *S3 Time series of all sites added to the model*.

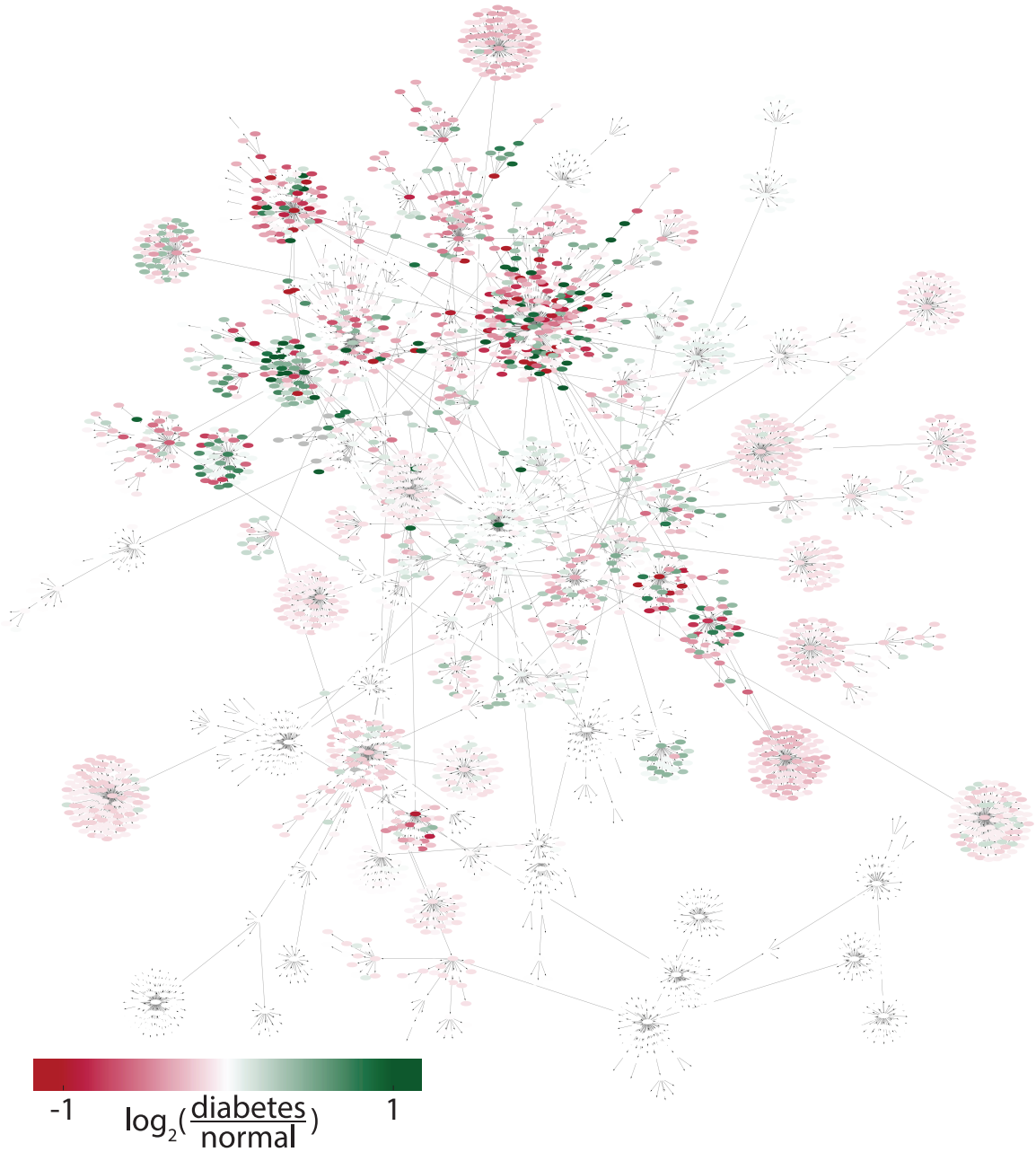

Figure 9: **The effect of type 2 diabetes propagated to all sites in the expanded model.** The color code goes from red that corresponds to decreased signaling to green that corresponds to increased signaling in type 2 diabetes. White corresponds to no or low effect. The full network is also available in *S4 Cytoscape file of the expanded model*. The simulations and experimental data for all phosphosites in the model is available in *S3 Time series of all sites added to the model*.

Discussion

In this work, we present a new methodology for automatic model expansion from a dynamic core model to include large-scale data and keep track of the level of confidence of both data

and interactions (Fig. 1). In the methodology, the new data is added in parallel in layers, to reduce computation time and allow for thousands of added time-resolved measurements. To create a well-studied core for adipocyte signaling, we have connected major events related to lipolysis, glucose uptake, and adiponectin release into a connected model (Fig. 2). This connected model, developed herein, agrees with numerous observations from adipocytes (Figs. 3 to 5). From the core model, we use our new methodology to add layers of experimentally measured phosphosites [18] based on a list of possible interactions, including level of confidence, gathered using OmniPathDB [19]. The developed methodology allows for combining detailed mechanistic models, common in the field of systems biology, with large-scale data, traditionally analyzed in the field of bioinformatics, thus acting as a bridge between such methods.

The created comprehensive model has a layered level of confidence based on both interactions and data (Fig. 6). Most of the data from the insulin signaling models [11, 15] comes from isolated primary adipocytes from non-diabetic and diabetic patients, where protein levels and activities have been measured using antibodies and the western blot technique. The adiponectin release and phosphoproteome data instead comes from the 3T3-L1 adipocyte cell line, where protein activities have been measured with mass-spectrometry and the release of adiponectin with patch-clamp technique. Furthermore, the small-scale data (western blot and patch-clamp) are typically more reliable since there are more repeats (typically $n = 5 - 15$) compared to the large-scale phosphoproteomic data ($n = 0-3$) per measurement. The additions in the expanded model is based on interactions with varying degrees of confidence. During the expansion, we first start with the interactions with the highest confidence, and then incrementally allow interactions with lower confidence (Fig. 7).

The development of new mechanistic ODE models for biological systems is a long and iterative process. First, you start with some data and hypothesis and then you refine the model in many steps, while using the model to make predictions for new data to gather. The development of the models included in the core model here, as well as the gathering of corresponding data, have taken us and our collaborators decades of work. Thus, to cover the entire phosphoproteome with consistent high quality data and mechanistic ODE models would not be feasible. The developed automated methodology therefore provides a reasonable trade-off between confidence in the model and the coverage of the phosphoproteome (Fig. 1).

The biological relevance of a comprehensive model of adipocytes signaling is related to type 2 diabetes. Insulin resistance is a hallmark of type 2 diabetes, and the core of the model describe in detail the intracellular insulin resistance of adipocytes (Fig. 3). A major function of adipocytes is to act as energy reserve for other organs, i.e. to store and release fatty acids. The release of fatty acids is increased in diabetes, due to a decrease in the reesterification (Fig. 4). Also, circulating adiponectin has been associated with a reduced risk of type 2 diabetes, and an increased insulin sensitivity. All of these processes have already been established in detail in models, and here we provide a connection including crosstalk that allow for their simultaneous simulations. Even more important – we provide a link between these highly established processes to the whole phosphoproteome, where the model allow anyone to simulate, e.g. the effect of a drug. We provide a type 2 diabetes version of the model (Figs. 8 and 9), therefore the model can be used to simulate the effect of anti-diabetic drugs throughout the adipocyte.

Other approaches to model human adipocytes on a large scale is the use of genome-scale metabolic models [20]. Such models provide another view on adipocyte changes in obesity and type 2 diabetes. These models integrate genomics, transcriptomics, and proteomics data

and apply to a fixed metabolic map based on known human metabolic reactions. Such models are complementary to the work herein since they do not consider the signaling pathways, and our approach does so far not include metabolic reactions. We are currently gathering labelled metabolite data for metabolic flux analyses from human adipocytes. These data will be integrated in the next version of the comprehensive adipocyte.

Method

Data collection and filtration

We used publicly available data, both data previously used to develop our models of glucose uptake, lipolysis, and adiponectin release[10, 11, 14, 15] and from other sources [18].

For the newly used phosphoproteomic data [18], we excluded sites where any time point had less than two repeats. For the inhibition data, sites with less than two repeats, or with an unclear effect of the inhibition (such that the uncertainty covered both an increase and a decrease) were ignored when the prediction accuracy was calculated. Furthermore, we divided the data into two groups (*responders* and *nonresponders* to insulin) based on the original classification done by the authors of the phosphoproteome work [18].

Quantifying the model agreement to experimental data

In order to evaluate the model's performance, we quantified the model agreement to data using a function typically referred to as a cost function. In this case, we used the normalized sum of squared residual as cost function (Eq. (1)).

$$v(\theta) = \sum_t \left(\frac{y_t - \hat{y}_t(\theta)}{SEM_t} \right)^2 \tag{1}$$

Here, $v(\theta)$ is the cost, equal to the sum of the normalized residual over all measured time points, t; p is the parameters; y_t is the measured data and $\hat{y}_t(\theta)$ is the model simulations; SEM_t is the standard error of the mean for the measured data.

Statistical analysis

To reject models, we used the χ^2-test with a significance level of 0.05. We used 582 degrees of freedom for the original training data leading to a threshold for rejection of $\chi^2(p = 0.05, dgf = 582) \approx 639.23$. For the extended set of experimental data including all the experimental data from the combined model we used 598 degrees of freedom, resulting in a threshold for rejection of $\chi^2(p = 0.05, dgf = 598) \approx 656.00$. Any combination of model and parameter set that results in a cost (Eq. (1)) above the threshold must be rejected. If no parameter set exists for a model that results in a sufficiently low cost, the model structure must be rejected.

For the automatic model expansion we allowed additions to the model where the cost for the tested phosphorylation site was below $\chi^2(0.05, 8) \approx 15.5$.

Optimization and software

We used MATLAB R2021a (MathWorks, Natick, MA) and the IQM toolbox (IntiQuan GmbH, Basel, Switzerland), a continuation of [21], for modelling. The parameter values were estimated using the enhanced scatter search (eSS) algorithm from the MEIGO toolbox [22]. We allowed the parameter estimation to freely find the best possible combinations of parameter values, within boundaries. Both the optimal parameter values and the bounds for the parameter values are given in *S2 Model equations and parameter values*.

Model and code availability

The model as well as the complete code for data analysis and modelling will be freely available upon submission for peer review. The model equations and parameter values are available in *S2 Model equations and parameter values*.

Automatic model expansion

In essence, we create an expanded model by iteratively adding proteins to the core model by identifying the proteins that are closest to, but not yet in, the model, and test if they can be added to the model in a pairwise fashion. This was done using phosphoproteomic data collected using mass-spectrometry [18], and a list of protein-protein interactions compiled using the python package OmniPathDB [19]. In detail, we find all proteins that are adjacent to the model, i.e. having at least one direct interaction in the list of interactions going from the model to the proteins. Using the list of adjacent proteins and corresponding prior knowledge interactions, we test if the proteins can be added to the model one by one (in parallel). In practice, each addition of an adjacent protein is tested using three different types of interactions. Firstly, we test if the interaction corresponds to a phosphorylation of the adjacent protein. Secondly, if the interaction is a dephosphorylation. Thirdly, if the interaction is a saturated phosphorylation. Fourthly, if the interaction must be a double interaction with two different inputs (double phosphorylation, double dephosphorylation or one phosphorylation with one dephosphorylation). Lastly, we test if the interaction results in a phosphorylation leading to a subsequent secondary state (such as an internalization) before returning to the unphosphorylated state. The adjacent proteins and interactions are tested by comparing the model simulation of the adjacent protein with the available experimental data. All adjacent proteins where the model simulations agree with the data sufficiently well are added to the model. We refer to the addition of the adjacent proteins that agree with data sufficiently well as the addition of a *layer* to the model. Once one such layer have been added to the model, we repeat the process of finding and adding another layer of adjacent proteins. For subsequent addition of layers, proteins in the previous layers are also used to find the new adjacent proteins. The algorithm for the model expansions is given as pseudocode in Algorithm 1

Algorithm 1 Automatic model expansion

```
I ← prior knowledge of protein-protein interactions
P_M ← all proteins in model
P_A ← all proteins adjacent to model, given I and P_M
valid_interactions = ∅
layer ← 0
while P_A ≠ ∅ do
    for p_A in P_A do
        I_A ← interactions from P_M to p_A in I
        for i in I_A do                    ▷ Interactions already tested are not tested again
            valid ← test_pairwise(i)
            if valid then
                valid_interactions = [valid_interactions; i], return
            end if
        end for
        if not valid then
            i_1 ← best i from pairwise loop
            for i_2 in I_A do
                valid ← test_double(i_1, i_2)
                if valid then
                    valid_interactions = [valid_interactions; i_1; i_2], return
                end if
            end for
        end if
        if not valid then
            for i in I_A do                ▷ Interactions already tested are not tested again
                valid ← test_extra_state(i)
                if valid then
                    valid_interactions = [valid_interactions; i], return
                end if
            end for
        end if
        if valid then
            P_M = [P_M; p_A],
        end if
    end for
    P_A ← all proteins adjacent to model, given I and P_M
    layer = layer + 1
end while
```

References

1. Hübner, K., Sahle, S. & Kummer, U. Applications and Trends in Systems Biology in Biochemistry. *The FEBS Journal* **278,** 2767–2857. ISSN: 1742-464X. https://doi.org/10.1111/j.1742-4658.2011.08217.x (2022) (Aug. 1, 2011).
2. Korkut, A. *et al.* Perturbation Biology Nominates Upstream–Downstream Drug Combinations in RAF Inhibitor Resistant Melanoma Cells. *eLife* **4** (ed Airoldi, E. M.) e04640. ISSN: 2050-084X. https://doi.org/10.7554/eLife.04640 (Aug. 2015).
3. Miller, M. L. *et al.* Drug Synergy Screen and Network Modeling in Dedifferentiated Liposarcoma Identifies CDK4 and IGF1R as Synergistic Drug Targets. *Science Signaling* **6,** ra85–ra85. https://www.science.org/doi/abs/10.1126/scisignal.2004014 (2013).
4. Yuan, B. *et al.* CellBox: Interpretable Machine Learning for Perturbation Biology with Application to the Design of Cancer Combination Therapy. *Cell Systems* **12,** 128–140.e4. ISSN: 2405-4712. https://www.sciencedirect.com/science/article/pii/S2405471220304646 (Feb. 17, 2021).
5. Nyman, E. *et al.* Perturbation Biology Links Temporal Protein Changes to Drug Responses in a Melanoma Cell Line. *PLOS Computational Biology* **16,** e1007909. https://doi.org/10.1371/journal.pcbi.1007909 (July 15, 2020).
6. Fröhlich, F. *et al.* Efficient Parameter Estimation Enables the Prediction of Drug Response Using a Mechanistic Pan-Cancer Pathway Model. *Cell Systems* **7,** 567–579.e6. ISSN: 2405-4712. https://doi.org/10.1016/j.cels.2018.10.013 (2022) (Dec. 26, 2018).
7. Terfve, C. D. A., Wilkes, E. H., Casado, P., Cutillas, P. R. & Saez-Rodriguez, J. Large-Scale Models of Signal Propagation in Human Cells Derived from Discovery Phosphoproteomic Data. *Nature Communications* **6,** 8033. ISSN: 2041-1723. https://doi.org/10.1038/ncomms9033 (Sept. 10, 2015).
8. Eduati, F. *et al.* Patient-Specific Logic Models of Signaling Pathways from Screenings on Cancer Biopsies to Prioritize Personalized Combination Therapies. *Molecular Systems Biology* **16,** e8664. ISSN: 1744-4292. https://doi.org/10.15252/msb.20188664 (2022) (Feb. 1, 2020).
9. Magnusson, R. *et al.* LASSIM—A Network Inference Toolbox for Genome-Wide Mechanistic Modeling. *PLOS Computational Biology* **13,** e1005608. ISSN: 1553-7358. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005608 (2021) (June 22, 2017).
10. Brännmark, C. *et al.* Insulin Signaling in Type 2 Diabetes: Experimental and Modeling Analyses Reveal Mechanisms of Insulin Resistance in Human Adipocytes. *Journal of Biological Chemistry* **288,** 9867–9880. ISSN: 0021-9258. https://doi.org/10.1074/jbc.M112.432062 (2021) (Apr. 5, 2013).
11. Rajan, M. R., Nyman, E., Kjølhede, P., Cedersund, G. & Strålfors, P. Systems-Wide Experimental and Modeling Analysis of Insulin Signaling through Forkhead Box Protein O1 (FOXO1) in Human Adipocytes, Normally and in Type 2 Diabetes. *Journal of Biological Chemistry* **291,** 15806–15819. ISSN: 0021-9258, 1083-351X. http://www.jbc.org/content/291/30/15806 (2020) (July 22, 2016).

12. Nyman, E. *et al.* A Single Mechanism Can Explain Network-wide Insulin Resistance in Adipocytes from Obese Patients with Type 2 Diabetes. *Journal of Biological Chemistry* **289,** 33215–33230. ISSN: 0021-9258, 1083-351X. http://www.jbc.org/content/289/48/33215 (2020) (Nov. 28, 2014).

13. Brännmark, C. *et al.* Mathematical Modeling of White Adipocyte Exocytosis Predicts Adiponectin Secretion and Quantifies the Rates of Vesicle Exo- and Endocytosis. *Journal of Biological Chemistry* **292,** 20032–20043. ISSN: 0021-9258, 1083-351X. http://www.jbc.org/content/292/49/20032 (2020) (Dec. 8, 2017).

14. Lövfors, W. *et al.* A Systems Biology Analysis of Adrenergically Stimulated Adiponectin Exocytosis in White Adipocytes. *Journal of Biological Chemistry* **297.** ISSN: 0021-9258, 1083-351X. https://www.jbc.org/article/S0021-9258(21)01024-3/abstract (2021) (Nov. 1, 2021).

15. Lövfors, W. *et al.* A Systems Biology Analysis of Lipolysis and Fatty Acid Release from Adipocytes in Vitro and from Adipose Tissue in Vivo. *PLOS ONE* **16,** e0261681. ISSN: 1932-6203. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0261681 (2022) (Dec. 31, 2021).

16. Jönsson, C., Castor Batista, A. P., Kjølhede, P. & Strålfors, P. Insulin and β-Adrenergic Receptors Mediate Lipolytic and Anti-Lipolytic Signalling That Is Not Altered by Type 2 Diabetes in Human Adipocytes. *Biochemical Journal* **476,** 2883–2908. ISSN: 0264-6021. /biochemj/article/476/19/2883/220490/Insulin-and-adrenergic-receptors-mediate-lipolytic (2020) (Oct. 15, 2019).

17. Stich, V. *et al.* Activation of A2-Adrenergic Receptors Blunts Epinephrine-Induced Lipolysis in Subcutaneous Adipose Tissue during a Hyperinsulinemic Euglycemic Clamp in Men. *American Journal of Physiology-Endocrinology and Metabolism* **285,** E599–E607. ISSN: 0193-1849. https://www.physiology.org/doi/full/10.1152/ajpendo.00502.2002 (2020) (Sept. 1, 2003).

18. Humphrey, S. J. *et al.* Dynamic Adipocyte Phosphoproteome Reveals That Akt Directly Regulates mTORC2. *Cell Metabolism* **17,** 1009–1020. ISSN: 1550-4131. https://www.cell.com/cell-metabolism/abstract/S1550-4131(13)00152-6 (2021) (June 4, 2013).

19. Türei, D., Korcsmáros, T. & Saez-Rodriguez, J. OmniPath: Guidelines and Gateway for Literature-Curated Signaling Pathway Resources. *Nature Methods* **13,** 966–967. ISSN: 1548-7105. https://www.nature.com/articles/nmeth.4077 (2021) (12 Dec. 2016).

20. Mardinoglu, A. *et al.* Integration of Clinical Data with a Genome-Scale Metabolic Model of the Human Adipocyte. *Molecular Systems Biology* **9,** 649. ISSN: 1744-4292. https://doi.org/10.1038/msb.2013.5 (2022) (Jan. 1, 2013).

21. Schmidt, H. & Jirstrand, M. Systems Biology Toolbox for MATLAB: A Computational Platform for Research in Systems Biology. *Bioinformatics* **22,** 514–515. ISSN: 1367-4803. https://academic.oup.com/bioinformatics/article/22/4/514/183389 (2020) (Feb. 15, 2006).

22. Egea, J. A. *et al.* MEIGO: An Open-Source Software Suite Based on Metaheuristics for Global Optimization in Systems Biology and Bioinformatics. *BMC Bioinformatics* **15,** 136. ISSN: 1471-2105. https://doi.org/10.1186/1471-2105-15-136 (2020) (May 10, 2014).

The following files are available as supplemental files.

S1 Recreation of the model agreement to data and validations

S2 Model equations and parameter values

S3 Time series of all sites added to the model

S4 Cytoscape file of the expanded model

S5 Excel spreadsheet with the impact of diabetes