Supplementary Information

Bayesian metamodeling of complex biological systems across varying representations

Barak Raveh[1,2,3,*], Liping Sun[4,*], Kate L. White[7,*], Tanmoy Sanyal[1,2,*], Jeremy Tempkin[1,2], Dongqing Zheng[8], Kala Bharat[1,2], Jitin Singla[7,9], ChenXi Wang[4,5,6], Jihui Zhao[4], Angdi Li[4,5,6], Nicholas A. Graham[8], Carl Kesselman[9], Raymond C. Stevens[4,5,7], Andrej Sali[1,2,3,10,#]

[1]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA 94158, USA

[2]Quantitative Biosciences Institute (QBI), University of California San Francisco, San Francisco, CA, 94158, USA

[3]School of Computer Science and Engineering, The Hebrew University of Jerusalem, 9190416, Israel

[4]iHuman Institute, ShanghaiTech University, Shanghai 201210, China

[5]School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

[6]University of Chinese Academy of Sciences, Beijing 100049, China

[7]Department of Biological Sciences, Bridge Institute, University of Southern California, Los Angeles, CA 90089, USA

[8]Mork Family Department of Chemical Engineering and Materials Science, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA

[9]Epstein Department of Industrial and Systems Engineering and Information Science Institute, the Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA

[10]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158, USA

[#]Correspondence: Andrej Sali, UCSF MC 2552, Byers Hall Room 503B, 1700 4th Street, San Francisco, CA 94143, USA. Tel +1 (415) 514-4227, email sali@salilab.org, web http://salilab.org.

Keywords: Bayesian metamodeling; integrative modeling; multiscale modeling; heterogeneous models; heterogenous representations; divide-and-conquer; multimodal data; whole-cell modeling; pancreatic ß-cell; insulin production pathway; glucose stimulated insulin secretion; incretins.

Supplementary Text

We first describe each of the eight input models and their corresponding surrogate models (Section 1), followed by outlining how surrogate models are coupled (Section 2).

1. Input and surrogate models

Input models. Our current goal is to illustrate metamodeling with a set of input models, without necessarily improving our understanding of pancreatic β-cell biology. Thus, we do not discuss in detail the validity of previously published input models (1). We do, however, highlight the main assumptions in constructing the previously unpublished input models (the pancreas, GSIS signaling, insulin metabolism, virtual screening of GLP1R, GI data, and GLP1 data models).

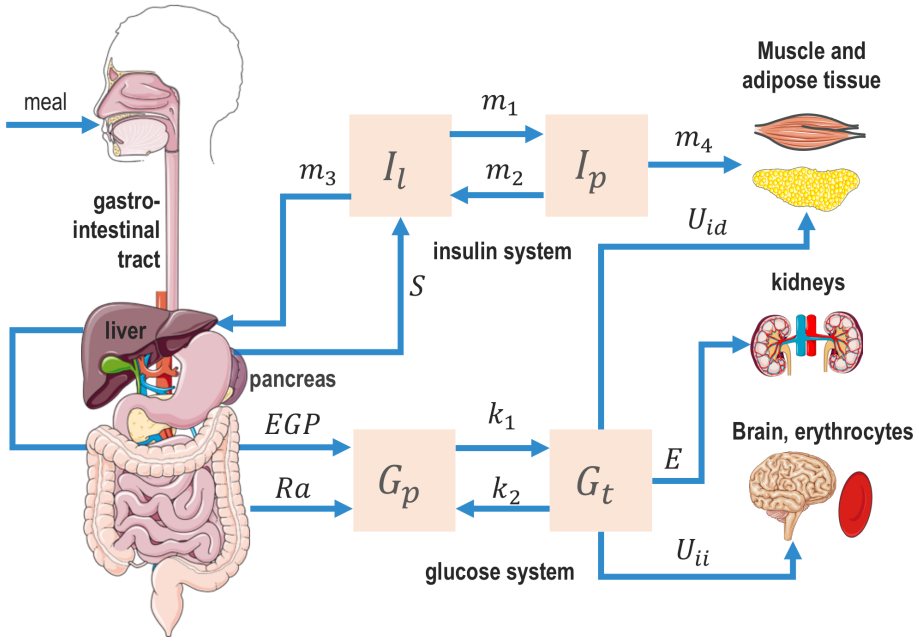
Surrogate models. We outline each surrogate model and how it was constructed from the input model, including modeling assumptions. We also provide a table listing the conditional PDF of its variables at different time slices of its DBN and their parameters.

DBNs of surrogate models. As described in the main text, a surrogate model is represented by a PDF over some variables of the corresponding input model and potentially additional auxiliary variables. A surrogate model aims to approximate statistical dependencies among the variables in the original model, potentially in a simplified form. Each surrogate model was represented by a DBN (2). Briefly, a DBN factorizes a PDF over a set of time-dependent variables by describing a conditional PDF for each random variable at time slice $t+\Delta t$ as a function of some random variables at time slice t and/or time slice $t+\Delta t$. In addition, a DBN describes a conditional PDF for all variables at an initial time slice t_0. If the values of some variables are observed, posterior conditional or marginal probabilities can be inferred for any subset of variables in the DBN at any time point. In addition, DBN parameters and topology can be learned from observed data.

DBN implementation. The topology of the DBN for each surrogate model is given in the top panel of Figure 2. The software implementation, input files, and sample output files are available at github.com/salilab/metamodeling. The conditional PDF of each random variable is a normal distribution whose mean is the weighted sum of some random variables (a linear Gaussian) in the current and/or previous time slice, with manually assigned standard deviations; the use of non-linear models is illustrated in a tutorial (github.com/tanmoy7989/bayesian_metamodeling_tutorial). The discrete time step Δt for the DBNs of all surrogate models was set to 1 min, although the original input models are constructed with their own time scales and time granularity. In fact, some models contain time-independent variables, which do not change over the timescale of a model. Time-independent variables may become time-dependent in the coupling stage, due to coupling with time-dependent variables from other surrogate models. To guarantee numerical stability, the conditional PDF of time-independent variables at each time slice is allowed to fluctuate slightly, by assigning them an arbitrary small standard deviation.

1.1 Postprandial response model

Input model. The postprandial response model describes insulin and glucose levels in the plasma and various body tissues (dependent variables) as a function of time, following a glucose-rich meal, in healthy and T2D subjects (Fig. S1) (1). The values of these variables are computed from the rate of glucose intake (independent variable), using a system of ODEs.



$$\dot{G}_p(t) = EGP(t) + Ra(t) - U_{ii}(t) - E(t) - k_1 G_p(t) + k_2 G_t(t)$$
$$\dot{G}_t(t) = U_{id}(t) + k_1 G_p(t) - k_2 G_t(t)$$
$$\dot{I}_l(t) = -[m_1 + m_3(t)] \cdot I_l(t) + m_2 I_p(t) + S(t)$$
$$\dot{I}_p(t) = -(m_2 + m_4) \cdot I_p(t) + m_1 I_l(t)$$
$$\vdots$$

Figure S1. The postprandial response input model. (Top) A schematic of the system of compartments and fluxes, described by a model consisting of 29 ODEs (1). The model takes into account interactions among different physiological systems and organs involved in glucose homeostasis. It was parameterized based on data from a cohort of 204 healthy subjects and 14 T2D subjects. These data include plasma insulin and glucose levels, endogenous glucose production, glucose rate of appearance, glucose utilization, and insulin secretion rate over 420 minutes at 20 minute intervals following a glucose-rich meal. (Bottom) A subset of 4 of the 29 ODEs in the complete model, indicating change in levels of plasma glucose (G_p), tissue glucose (G_t), liver insulin (I_l), and plasma insulin (I_p).

Surrogate model. The rationale behind the construction of the postprandial response surrogate model is described in Results (second example for Step 1 of metamodeling). It is relatively straightforward to construct a DBN for a system of ODEs, because a DBN can be considered a probabilistic, discretized generalization of a system of ODEs. The postprandial response surrogate model nonetheless simplifies the input model, still capturing key statistical relations among its variables (Fig. 2, Table S1) by fitting some parameters of the conditional PDFs listed

in Table 1 to reproduce its dependent parameters. As described above, all conditional PDFs are linear Gaussians.

Table S1. Random variables, corresponding conditional PDFs, and conditional PDF parameters in the postprandial response surrogate model.

Time-dependent variables and their conditional PDFs:					
Name	**Description**	**DBN time slice* [min]**	**Mean [healthy/T2D**]**	**Std-dev**	**Unit**
ΔG_d^{PR}	Rate of glucose intake from food digestion	t_0	0.0	10^{-2}	mM min^{-1}
		$t + \Delta t$	0.0	10^{-3}	
G_b^{PR}	Basal plasma glucose concentration	t_0	5.1/9.2	1.0	mM
		$t + \Delta t$	$G_b^{PR}(t)$	0.1	
G^{PR}	Plasma glucose concentration	t_0	5.1/9.2	0.1	mM
		$t + \Delta t$	$\Delta G_d^{PR}(t)\,\Delta t - k_1 I^{PR}(t)\,\Delta t + (1 - k_2)\,G^{PR}(t)\,\Delta t + k_3^{PR}\Delta G^{PR}(t)\,\Delta t$	0.01	
ΔG^{PR}***	Excess plasma glucose concentration compared with the basal concentration	t_0	$G^{PR}(t_0) - G_b^{PR}(t_0)$	0.1	mM
		$t + \Delta t$	$G^{PR}(t + \Delta t) - G_b^{PR}(t + \Delta t)$	0.01	
Y^{PR}	Provision of new insulin to the β-cells	t_0	0.00	0.1	pM min^{-1}
		$t + \Delta t$	$(1 - \alpha^{PR}\Delta t)Y(t) + \alpha^{PR}\beta^{PR}\Delta G^{PR}\Delta t$	0.01	
S_b^{PR}	Pancreatic insulin secretion rate	t_0	$S_b^{PR}(t_0)$	0.1	pM min^{-1}
		$t + \Delta t$	$Y^{PR}(t + \Delta t) + K^{PR}\Delta G_d^{PR}(t + \Delta t) + S_b^{PR}(t + \Delta t)$	0.01	
I^{PR}	Plasma insulin concentration	t_0	25.0/52.0	0.1	pM
		$t + \Delta t$	$(1 - \gamma^{PR}\Delta t)\,I^{PR}(t) + k_4^{PR}S^{PR}(t)\Delta t$	0.01	
Time-independent variables and their conditional PDFs:					
T2D	A boolean variable indicating whether a subject is diabetic or healthy	-	T2D is an observed variable, determining CPD parameters for CPDs of all other variables. In principle, a probabilistic prior could be defined for T2D (*e.g.*, population prevalence of T2D).	-	-
S_b^{PR}	Basal pancreatic insulin secretion rate	-	34.0/102.5	0.1/1	pM min^{-1}
Time-independent variables represented as conditional PDF parameters:					
α^{PR}	Delay between the glucose signal and insulin secretion	-	0.050/0.013	-	min^{-1}
β^{PR}	Pancreatic responsivity to glucose	-	120.0/115.0	-	pM min^{-1} mM^{-1}

Parameters of conditional PDFs:					
Name	**Description**	**DBN time slice* [min]**	**Mean [healthy/T2D**]**	**Std-dev**	**Unit**
γ^{PR}	Transfer rate between portal vein and liver	-	1.0/1.0	-	min^{-1}
k_1^{PR}	Coefficient for insulin reducing glucose concentration	-	0.00025/0.00013	-	min^{-1}
k_2^{PR}	Coefficient for glucose reducing glucose concentration	-	-0.001230/-0.000735	-	min^{-1}
k_3^{PR}	Coefficient for elevated glucose reducing glucose concentration	-	$-10^{-6}/-10^{-6}$	-	-
k_4^{PR}	Coefficient for insulin secretion accounting for the insulin degradation	-	0.7353/0.5073	-	-
K^{PR}	Pancreatic responsivity to the glucose rate of change	-	1000.0/400.0	-	Pmol L^{-1} mM^{-1}

*t_0 is the initial time slice in the DBN; $t + \Delta t$ is the time slice that follows time slice *t* by a time step of Δt.
**The Healthy/T2D variable in Figure 2 was implemented as different values for variables and parameters for the same PGM.
***ΔG^{PR} was explicitly used in the code only as an intermediate variable to compute the excess plasma glucose concentration (difference between variables G^{PR} and G_b^{PR}), and was omitted from Figure 2.

1.2 Pancreas model

Input model. The pancreas model is a simple linear model that relates the insulin secretion rate by individual cells (independent variable) to the insulin secretion rate by individual islets and an entire pancreas (dependent variables) (Fig. S2).

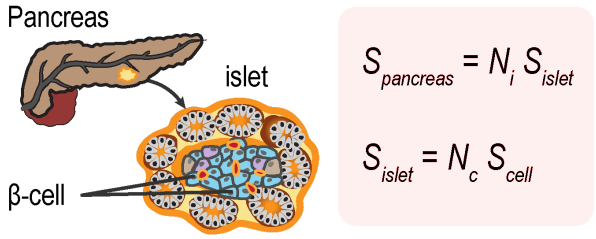


Figure S2. The pancreas input model. The insulin secretion rate of an islet S_{islet} is the sum of the secretion rates of N_c β-cells (S_{cell}) in the islet; similarly, the insulin secretion rate of the pancreas $S_{pancreas}$ is the sum of the secretion rates of N_i islets in the pancreas. The model is parameterized based on the estimated N_c =1140 β-cells in an islet (3) and N_i =3.2 million islets in a pancreas (4). Variable descriptions are indicated in Table S2. We made two simplifying assumptions in the construction of the pancreas input model. First, the secretion rates of all β-cells are identical and all islets contain the same

number of β-cells. Second, N_c and N_i were assigned identical values for both healthy and T2D subjects, although the proportion of β-cells in T2D islets is marginally decreased compared to normal islets (5).

Surrogate model. The DBN describing the pancreas surrogate model is a discretized, probabilistic version of the linear equations of the corresponding input model (Fig. 2, Table S2). It includes conditional PDFs corresponding to linear Gaussians describing the statistical dependencies between insulin secretion rates for β-cells, islets, and the pancreas. This surrogate model is essential for the subsequent coupling of the postprandial response, vesicle exocytosis, GSIS signaling, and GI data surrogate models (Results, Step 2). It includes only time-independent variables. However, it is implemented as a DBN, because its variables become time-dependent during the coupling stage, due to coupling with time-dependent variables of other surrogate models (Fig. 2).

Table S2. Random variables, corresponding conditional PDFs, and conditional PDF parameters in the pancreas surrogate model.

Time-independent variables and their conditional PDFs:				
Name	**Description**	**Mean**	**Std-dev**	**Unit**
S_{cell}^{Pa}	Insulin secretion rate of a cell	$S_{cell}^{C}(t)$	10^{-10}	pM min^{-1}
S_{is}^{Pa}	Insulin secretion rate of an islet	$N_c S_{cell}^{Pa}(t)$	10^{-6}	pM min^{-1}
S_{pa}^{Pa}	Insulin secretion rate of the pancreas	$N_i S_{is}^{Pa}(t)$	0.1	pM min^{-1}
Parameters of conditional PDFs:				
N_c	Number of β-cells in an islet	1140	-	-
N_i	Number of islets in a pancreas	3.2×10^6	-	-

*Legend: t_0 - the initial time slice in the DBN; $t + \Delta t$ - the time slice that follows time-slice *t* by a time step of Δt.

1.3 Vesicle exocytosis model

Input model. The vesicle exocytosis model describes coarse-grained spatiotemporal trajectories of vesicle exocytosis in pancreatic β-cells (dependent variables) after glucose stimulation, given an initial cell configuration (independent variables) (Fig. S3) (6).

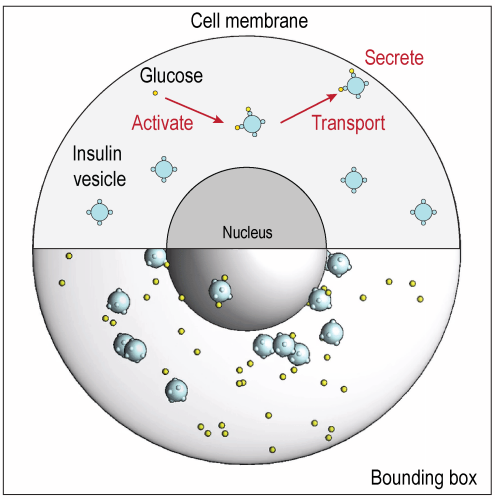


Fig. S3. The vesicle exocytosis input model. The scheme indicates model components (top) and a snapshot from an actual coarse-grained Brownian dynamics simulation trajectory (bottom) (6). The model includes the cell membrane (light gray sphere), the nucleus (dark gray sphere), hundreds of insulin vesicles (light blue spheres), and thousands of glucose molecules (yellow spheres). The components are rescaled for visualization. Brownian dynamics simulations are restrained by various experimental data, including soft X-ray tomograms of the cell (7). 48 different 200 ms trajectories were computed for each of the following three treatment conditions (Table S3): (i) no glucose stimulation, (ii) 25 mM glucose stimulation, and (iii) 25 mM glucose stimulation with 10 nM Exendin-4 treatment. Exendin-4 is a peptide agonist of the glucagon-like peptide (GLP) receptor that attenuates postprandial plasma glucose (8).

Table S3. Model parameters of data-driven Brownian dynamics simulations under different conditions.

Treatment	Number of runs	Simulation time [ms]	Time step [s]	Radius of the cell [Å]	Number of glucose molecules	Number of insulin vesicles	Force coefficient [kcal mol^{-1} Å^{-1}]	Number of activation patches	Vesicle diffusion coefficient [Å^2 fs^{-1}]	Glucose diffusion coefficient [Å^2 fs^{-1}]
0 mM Glucose	48	200	10^{-8}	65,000	1000	306	For each condition, 10^{-4}, 10^{-5}, 10^{-6}, and 10^{-7}	For each condition, 6 and 12	10^{-5}	10^{-7}
25 mM Glucose	48	200	10^{-8}	65,000	4000	791			10^{-5}	10^{-7}
25mM Glucose + 10 nM Exendin-4	48	200	10^{-8}	65,000	4000	508			10^{-5}	10^{-7}

Surrogate model. The rationale behind the construction of the vesicle exocytosis surrogate model is described in Results (first example for Step 1 of metamodeling). This surrogate model includes a subset of the input model variables and new variables that are computed from input model variables. For example, total insulin secretion rate in the surrogate model is computed from insulin vesicle coordinates; these coordinates are in turn omitted from the surrogate model for practical reasons of dimensionality reduction. The conditional PDF parameters are fitted manually to the Brownian dynamics simulations to recapitulate the insulin secretion rates of the β-cell for different simulation conditions. The time step in the surrogate model (1 min) is longer

than the entire time scale of the vesicle exocytosis model (200 ms); thus, a single time slice in the DBN of the surrogate model represents a single simulation trajectory of vesicle exocytosis. The surrogate vesicle exocytosis model simplifies the corresponding input model in three ways. First, the cell is secreting at a constant rate across one minute. Second, the surrogate model interpolates a linear Gaussian relationship from a discrete set of treatment conditions. Third, the surrogate model describes instantaneous rather than second phase insulin secretion occuring 30 minutes after glucose stimulation.

Table S4. Random variables, corresponding conditional PDFs, and conditional PDF parameters in the vesicle exocytosis surrogate model.

Time-dependent variables and their conditional PDFs:					
Name	**Description**	**DBN time slice* [min]**	**Mean**	**Std-dev**	**Unit**
G^{VE}	Intracellular glucose concentration	t_0	2.55	0.1	mM
		$t + \Delta t$	2.55	0.01	
k^{VE}_t	Effective rate of vesicle trafficking towards the cellular periphery	t_0	$5.0 + \alpha^{VE} S^{VE}(t_0)/2$	1.0	m s^{-1}
		$t + \Delta t$	$5.0 + \alpha^{VE} S^{VE}(t + \Delta t)/2$	0.1	
N^{VE}_v	Number of insulin vesicles in one β-cell	t_0	$\beta^{VE} S^{VE}(t_0)$	1	-
		$t + \Delta t$	$\beta^{VE} S^{VE}(t + \Delta t)$	0.1	
N^{VE}_{patch}	Number of activation patches per vesicle	t_0	6	0.1	-
		$t + \Delta t$	$1.0\, N^{VE}_{patch}(t)$	0.01	
N^{VE}_{ins}	Amount of insulin in one vesicle	t_0	1.8×10^{-6}	10^{-7}	pmol
		$t + \Delta t$	$1.0\, N^{VE}_{ins}(t)$	10^{-8}	
S^{VE}	Insulin secretion rate of one β-cell	t_0	$9.32 * 10^{-9}$	10^{-10}	pM min^{-1}
		$t + \Delta t$	$k^{VE}_G G^{VE}(t) + k^{VE}_p N^{VE}_{patch}(t + \Delta t) + k^{VE}_{ins} N^{VE}_{ins}(t + + k^{VE}_D D^{VE}_v(t + \Delta t) + k^{VE}_R R^{VE}_{cell}(t + \Delta t)$	10^{-11}	
Time-independent variables represented as conditional PDF parameters:					
R^{VE}_{cell}	Radius of the β-cell	t	6	0.1	μm
D^{VE}_v	Diffusion coefficient of insulin vesicles in the beta cell	t	0.0032	10^{-4}	A^2 fs^{-1}

Parameters of conditional PDFs:					
Name	**Description**	**DBN time slice* [min]**	**Mean**	**Std-dev**	**Unit**
α^{VE}	Correlation between insulin secretion rate and the force coefficient of vesicle transport	-	1.073×10^{9}	-	$m\ s^{-1}\ pM^{-1}\ min$
β^{VE}	Correlation between insulin secretion rate and the number of insulin vesicles in the cell	-	3.219×10^{10}	-	$pM^{-1}\ min$
k_p^{VE}	Coefficient for the number of activation patches on the vesicle surface accelerating insulin secretion rate	-	2.55×10^{-10}	-	$pM\ min^{-1}$
k_G^{VE}	Coefficient for the intracellular glucose concentration stimulating the insulin secretion	-	3.4×10^{-9}	-	$pM\ min^{-1}\ mM^{-1}$
k_{ins}^{VE}	Coefficient for the number of insulin molecules in each vesicle determining insulin secretion rate	-	0.01	-	$pM\ min^{-1}$
k_D^{VE}	Coefficient for the vesicle diffusion promoting insulin secretion rate	-	10^{-7}	-	$pM\ min^{-1}\ A^{-2}\ fs$
k_R^{VE}	Coefficient for the cell radius reducing insulin secretion rate	-	-3.2×10^{-9}	-	$pM\ min^{-1}\ \mu m^{-1}$

*t_0 is the initial time slice in the DBN; $t + \Delta t$ is the time slice that follows time slice *t* by a time step of Δt.

1.4 GSIS signaling model

Input model. The GSIS signaling model describes the dynamics of the molecular signaling network leading to insulin secretion in β-cells following glucose stimulation and a GLP1 hormone signal (Fig. S4). The model computes insulin secretion rates as well as concentrations

of various signaling molecules, such as ATP, cAMP, and Ca^{2+} over time (dependent variables), for a given starting condition (independent variables), using a system of linear ODEs.

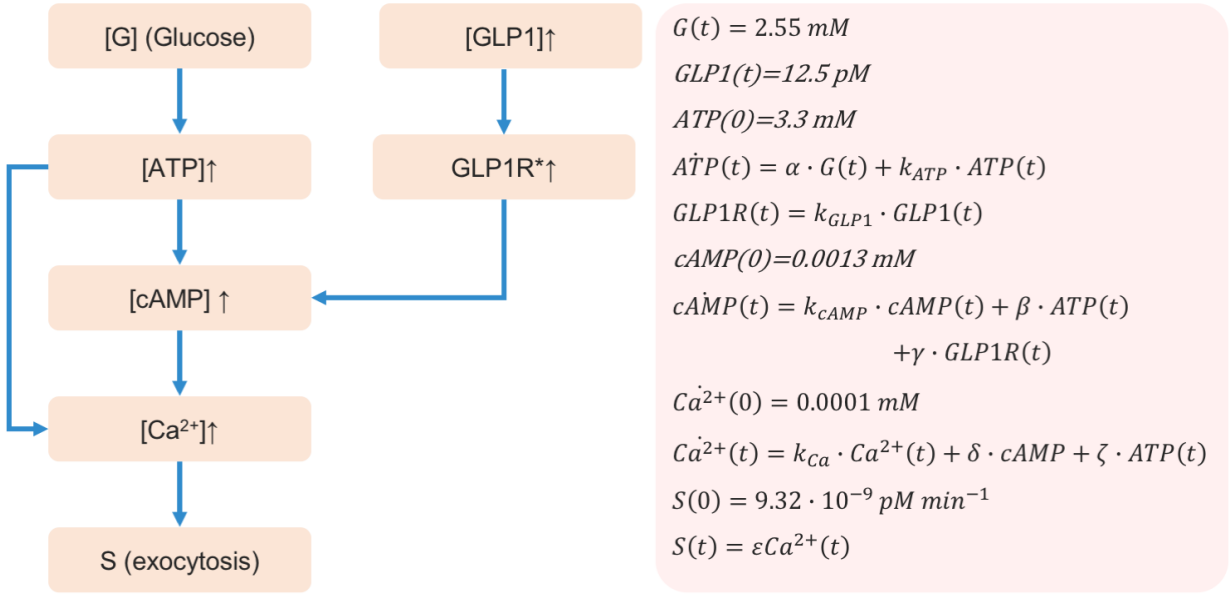


Figure S4. The GSIS signaling input model. The signaling network topology (left) is a combination of the linear pathways leading from glucose stimulation to insulin secretion, based on pathway hsa04911 in the KEGG database (9) and the signaling pathway leading to cAMP-dependent enhancement of insulin secretion following activation of the GLP1R receptor by the peptide hormone GLP1. GLP1R* is a variable that indicates GLP1R relative activity levels. Linear ODEs (right) describe the time evolution of the network component, leading to insulin secretion, *S(t)*. ODE coefficient values are identical to the values of the corresponding parameters of the conditional PDFs (Table S5). The following simplifying assumptions were made during model construction: (1) the system dynamics are described by linear equations; (2) the signaling network of pancreatic β cells is simplified, for example, by merging alternative pathways through which cAMP and ATP modulate calcium release from various cellular and extracellular compartments and omitting feedback loops in the network; and (3) identical parameters are used for both healthy and T2D subjects.

Surrogate model. The DBN describing the GSIS signaling surrogate model is a discretized, probabilistic version of the linear ODEs of the corresponding input model (Fig. 2, Table S5). As with other surrogate models, standard deviations were added to reflect our uncertainty in variable values.

Table S5. Random variables, corresponding conditional PDFs, and conditional PDF parameters in the GSIS signaling surrogate model.

Time-dependent variables and their conditional PDFs:					
Name	**Description**	**DBN time slice* [min]**	**Mean**	**Std-dev**	**Unit**
G^{Sg}	Intracellular glucose concentration	t_0	2.55	0.1	mM
		$t + \Delta t$	2.55	0.01	
ATP^{Sg}	Intracellular ATP concentration	t_0	3.3	0.1	mM
		$t + \Delta t$	$\alpha^{Sg} G^{Sg}(t) + k_{ATP}^{Sg} ATP^{Sg}(t)$	0.01	
$GLP1^{Sg}$	Plasma GLP1 concentration	t_0	12.5 (basal levels, see GLP1 model)	0.1	pM
		$t + \Delta t$	$1.0\ GLP1^{Sg}(t)$	0.01	
$GLP1R^{Sg}$	GLP1R activity	t_0	1.0	0.1	-
		$t + \Delta t$	$k_{GLP1}^{Sg} GLP1^{Sg}(t)$	0.01	
$cAMP^{Sg}$	Intracellular cAMP concentration	t_0	1.3 x 10^{-3}	10^{-4}	mM
		$t + \Delta t$	$k_{cAMP}^{Sg} cAMP^{Sg}(t) + \gamma^{Sg} GLP1R^{Sg}(t) + \beta^{Sg} ATP^{Sg}(t)$	10^{-5}	
Ca^{2+Sg}	Intracellular Ca^{2+} concentration	t_0	0.0001	10^{-5}	mM
		$t + \Delta t$	$k_{Ca}^{Sg} Ca^{2+Sg}(t) + \delta^{Sg} cAMP^{Sg}(t) + \zeta^{Sg} ATP^{Sg}(t)$	10^{-6}	
S^{Sg}	Insulin secretion rate of a single β-cell. Concentration is relative to total plasma volume.	t_0	$9.32*10^{-9}$	10^{-10}	pM min^{-1}
		$t + \Delta t$	$\epsilon^{Sg} Ca^{2+Sg}(t)$	10^{-11}	
Parameters of conditional PDFs:					
α^{Sg}	Coefficient of glucose-dependent ATP production	-	0.65	-	-
k_{ATP}^{Sg}	Percentage of unconsumed ATP between adjacent timeslices	-	0.5	-	-
k_{GLP1}^{Sg}	Coefficient for GLP1 activating GLP1R	-	0.08	-	pM^{-1}
β^{Sg}	Coefficient for cAMP production through ATP	-	0.00013	-	-
γ^{Sg}	Coefficient for cAMP production through GLP1R activation	-	0.000438	-	-

Parameters of conditional PDFs (cont.)					
Name	**Description**	**DBN time slice* [min]**	**Mean**	**Std-dev**	**Unit**
k_{cAMP}^{Sg}	Percentage of unconsumed cAMP between adjacent timeslices	-	0.333	-	-
δ^{Sg}	Coefficient from cAMP-dependent change to Ca^{2+} levels	-	0.0385	-	-
k_{Ca}^{Sg}	Percentage of unconsumed Ca^{2+} between adjacent timeslices	-	0.5	-	-
ϵ^{Sg}	Coefficient for Ca^{2+} facilitating insulin secretion	-	0.0011333	-	pM min^{-1} mM^{-1}
ζ^{Sg}	Coefficient for ATP-dependent change to Ca^{2+} levels		0.0		

*t_0 is the initial time slice in the DBN; $t + \Delta t$ is the time slice that follows time slice *t* by a time step of Δt.

.

1.5 Insulin metabolism model

Input model. The insulin metabolism model predicts activation of cellular metabolic pathways (dependent variables) for different treatment conditions (independent variables), based on experimental measurements of metabolomic signatures obtained using liquid chromatography-mass spectrometry.

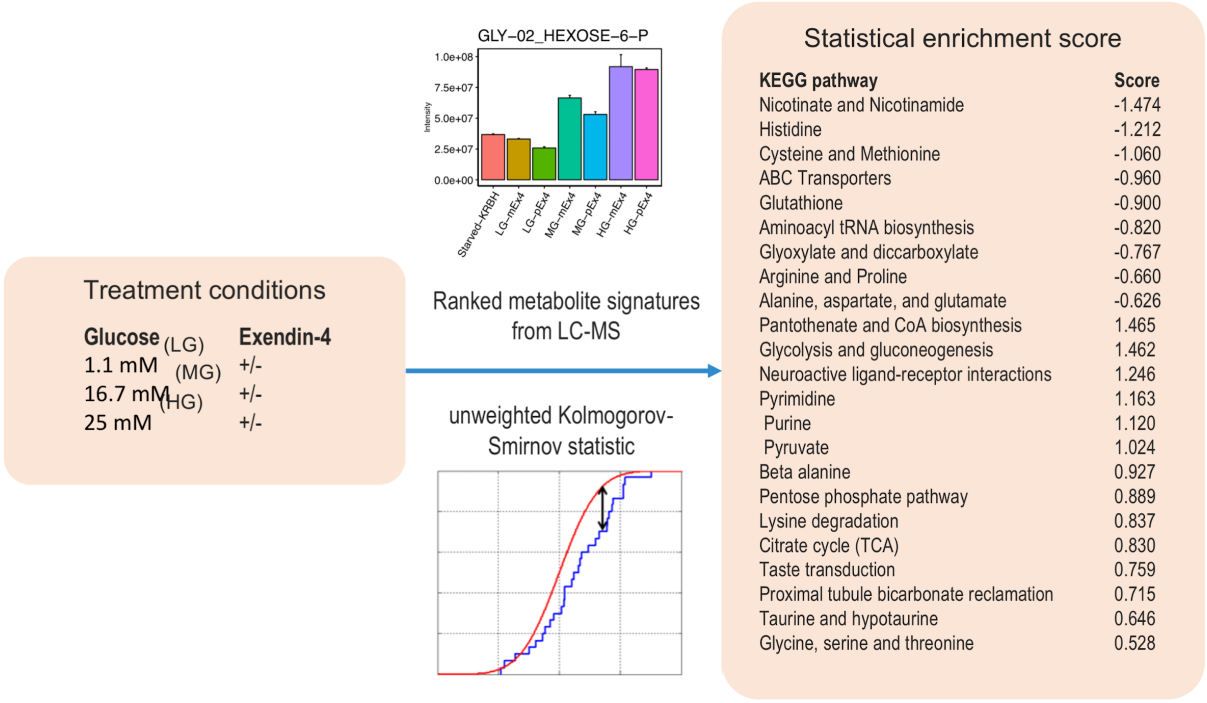


Figure S5. The insulin metabolism input model. Enrichment analysis for functional metabolic pathways is informed by metabolite quantities measured by liquid chromatography-mass spectrometry (LC-MS) measurements on INS-1E cells under different glucose and Exendin-4 treatments. Based on these measurements, a statistical model was constructed that outputs normalized enrichment scores for 25 different metabolic pathways from the KEGG pathways database (116), in response to different treatment conditions. Metabolomic signatures were ranked based on log2-transformed fold change observed for a given perturbation (*e.g.*, LG/MG/HG co-stimulated with Exendin-4 compared with incretin-free LG/MG/HG treated INS-1E cells). Metabolites were annotated using KEGG COMPOUND ID (*e.g.,* D-Glucose: C00031). The enrichment analyses were run with unweighted Kolmogorov-Smirnov statistic using the Broad Institute's GSEA java applet against a library containing all KEGG metabolic pathways. Normalized enrichment scores (NES) were calculated to determine if metabolic pathways were overrepresented at the top or the bottom of the given rank lists. Statistical significance scores were assessed by 5,000 permutations of the ranked lists. The permutation-based P-value is defined by the fraction of randomly permuted lists resulting in the NES values greater than or equal to the observed NES. The KEGG metabolic pathway library (c2_kegg_gene_cpd_set.gmt) was constructed by scraping the KEGG API. NES scores for each KEGG pathway at each treatment conditions are provided in the Excel file https://github.com/salilab/metamodeling/blob/master/data/072919-INS1e-30min-Enrichment-analysis-cleaned-summary.xlsx.

Surrogate model. The insulin metabolism surrogate model describes a parametrized linear relation between the treatment conditions (the independent variables), the TCA cycle enrichment score (one of the dependent variables), and the concentrations of intracellular ATP (an additional variable estimated from enrichment of TCA). The insulin metabolism surrogate model simplifies the corresponding input model in three ways. First, it summarizes a narrow aspect of the input model that is of particular interest. Second, it interpolates a linear Gaussian relationship from a discrete set of treatment conditions. Third, it assumes that the change in TCA cycle activation and ATP concentration occurs instantaneously rather than 30 minutes after glucose stimulation. The model parameters were manually fitted to reproduce the empirical relation between independent and dependent variables in the corresponding input model. While this model describes INS-1e cells rather than primary β-cells, differences in pathway enrichment among cell types are accounted for during the coupling stage (Discussion).

Table S6. Random variables, corresponding conditional PDFs, and conditional PDF parameters in the insulin metabolism surrogate model.

Time-independent variables and their conditional PDFs:					
Name	**Description**	**DBN time slice* [min]**	**Mean**	**Std-dev**	**Unit**
G^{Mb}_{ex}	Extracellular glucose concentration	*t*	25	0.01	mM
$Ex4^{Mb}_{ex}$	Extracellular Ex-4 concentration	*t*	0.0	0.01	nM
P^{Mb}_{INS1E}	Normalized enrichment score of the TCA pathway in INS1E cells	*t*	$k1^{Mb} G^{Mb}_{ex}(t) + k2^{Mb} Ex4^{Mb}_{ex}(t)$	0.01	-
ATP^{Mb}_{INS1E}	Intracellular metabolite concentration (here, ATP) in INS1E cells	*t*	$k3^{Mb} * P^{Mb}_{INS1E}(t)$	0.01	mM
Parameters of conditional PDFs:					
$k1^{Mb}$	Coefficient for the extracellular glucose stimulating the metabolic pathway	-	0.0372	-	mM^{-1}
$k2^{Mb}$	Coefficient for the extracellular Ex-4 stimulating the metabolic pathway	-	0.5	-	nM^{-1}
$k3^{Mb}$	Coefficient for the TCA metabolic pathway producing ATP	-	3.3	-	mM

*Legend: t_0 - the initial time slice in the DBN; $t + \Delta t$ - the time slice that follows time-slice *t* by a time step of Δt.

Cell cultures. INS-1E *Rattus Norvegicus* insulinoma cells were obtained from the Cell Culture Core of the Raymond Stevens lab (Bridge Institute USC). Cells were maintained in modified RPMI 1640 supplemented with 5% horse serum, 1 mM sodium pyruvate, 50 μM β-mercaptoethanol, 2 mM glutamine, and 10 mM HEPES in monolayer. All cells were grown in a humidified incubator at 5% CO_2 and 37˚C. They were used between 30-50 passages of thawing. Cell counting and viability were assessed using trypan blue staining with a TC20 automated cell counter (BioRad).

Cell stimulation and intracellular metabolites extraction. Cells were plated on 6-well plates at a density of 7,000 cells/cm^2. When cell densities reached 70%, the media was removed, cells were washed twice with 2 mL of PBS, followed by adding 5 mL of KRBH buffer with 0 mM glucose to cells. The cells were starved for 30 min prior to treatment. Following starvation, KRBH buffers were removed and the cells were treated with 1.1 mM, 16.7 mM, and 25 mM of glucose without or with 10 nM of Exendin-4. After 30 min, supernatants were collected, spun-down, and assayed using Mercodia Rat Insulin ELISA kit (10-1250-01) according to the manufacturer protocol. The culture plates were cooled on ice, and the cells were washed with 1 mL of cold ammonium acetate. The methanol cell suspensions were scraped and transferred to Eppendorf tubes, followed by centrifugation at 4˚C. The supernatants were transferred to new Eppendorf tubes, and the pellets were re-extracted with another 350 μL of -80˚C methanol. The second methanol extraction was spun-down, and the supernatants were pooled with the first extraction. Metabolites were speed-vac'ed to dryness, resuspended in LC-MS grade water, and submitted to LC-MS.

Liquid chromatography-mass spectrometry (LC–MS) metabolomics. Samples were randomized and analyzed on a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer coupled to an UltiMate 3000 UHPLC system (Thermo Scientific). The mass spectrometer was run in polarity switching mode (+3.00 kV/-2.25 kV) with an m/z window ranging from 65 to 975. Mobile phase A was 5 mM NH4AcO, *p*H 9.9, and mobile phase B was acetonitrile. Metabolites were separated on a Luna 3 μm NH2 100 Å (150 × 2.0 mm) column (Phenomenex). The flow rate was 300 μL/min, and the gradient was from 15% A to 95% A in 18 min, followed by an isocratic step for 9 min and re-equilibration for 7 min. All samples were run in biological triplicate. Metabolites were detected and quantified by the area-under-the-curve, based on retention time and accurate mass (≤ 8 ppm) using the TraceFinder 3.3 (Thermo Scientific) software. Intracellular data was normalized to the cell number at the time of extraction.

1.6 Virtual screening model

Input model. The virtual screening model describes the increase in the activity level of GLP1R (dependent variable) for different concentrations of various GLP1 agonist compounds, based on their rank in a library of compounds (independent variables).

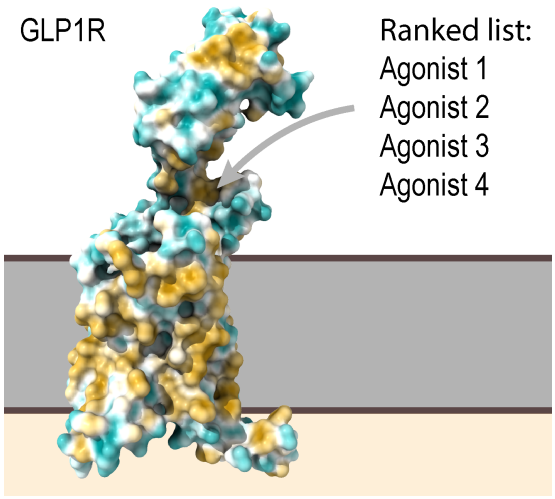


Figure S6. The virtual screening input model. The model is based on computational docking of 5,689 potential agonists against an atomic structure of GLP1R (10). These potential agonists were ranked by their predicted affinity to the allosteric site of GLP1R. We selected the top four compounds. The model computes the increase in GLP1R activity as a function of the rank of the compound *A* and its concentration *C*: $\frac{1-e^{-k\cdot A\cdot C}}{1+e^{-k\cdot A\cdot C}} \cdot F$, where maximal relative activation *F* is set to 310.0 fold and a concentration normalization coefficient *k* was set to 1.1 pM^{-1}.

Surrogate model. The virtual screening surrogate model is a linear approximation of the corresponding input model, obtained through manual fitting. It simplifies the corresponding input model in two ways: First, activity is related to the rank and concentration *via* a linear Gaussian. Second, identical variable and parameter values are used for both healthy and T2D subjects.

Table S7. Random variables, corresponding conditional PDFs, and conditional PDF parameters in the virtual screening surrogate model.

Time-dependent variables and their conditional PDFs:					
Name	**Description**	**DBN time slice* [min]**	**Mean**	**Std-dev**	**Unit**
C^{VS}	Compound concentration of the GLP1 analogs	t_0	0	0.01	pM
		$t + \Delta t$	$C^{VS}(t)$	10^{-3}	
$GLP1R^{VS}$	GLP1R activity level relative to GLP1R activity upon binding to GLP1 at its basal concentration	t_0	0	0.01	-
		$t + \Delta t$	$k1^{VS}A^{VS}(t) + k2^{VS}C^{VS}(t)$	10^{-3}	
Time-independent variables and their conditional PDFs:					
$A^{VS}{}_{**}$	Rank of different GLP1 agonists with increasing affinities	t	Prior: 0 Observed values: 1/2/3/4	0.01	-

Parameters of conditional PDFs:					
Name	**Description**	**DBN time slice* [min]**	**Mean**	**Std-dev**	**Unit**
$k1^{VS}$	Coefficient for the GLP1 agonists activating GLP1R	-	40	-	-
$k2^{VS}$	Coefficient for the compound concentration affecting GLP1R activation	-	2	-	pM^{-1}

*Legend: t_0 - the initial time slice in the DBN; $t + \Delta t$ - the time slice that follows time-slice *t* by a time step of Δt.

** This variable was implemented in the code through additional nodes in a mathematically equivalent fashion.

1.7 Glucose intake data model

Input model. The GI data model tabulates data on the rate of glucose intake from food digestion, ΔG_d^{GS} (Table S8). The GI data model illustrates how real-world data can be integrated through metamodeling and coupled with other models.

Table S8. GI data model. The data was simulated at 1 min interval using a sigmoid postprandial response model. The postprandial response model itself is in turn based on empirical data that are typically measured by glucose sensors as the rate of appearance from the intestine (1).

Time [min]	**ΔG_d [mM min^{-1}]**
0	0.047
1	0.057
2	0.069
3	0.083
4	0.099
5	0.118
6	0.140
7	0.166
8	0.195
9	0.226
10	0.261
11	0.298
12	0.337
13	0.375
14	0.411
15	0.443

16	0.470
17	0.489
18	0.498
19	0.498
20	0.489
21	0.470
22	0.443
23	0.411
24	0.375
25	0.337
26	0.298
27	0.261
28	0.226
29	0.195
30	0.166
31	0.140
32	0.118
33	0.099
34	0.083
35	0.069
36	0.057
37	0.047
38	0.039
39	0.032
40	0.026
41	0.022
42	0.018
43	0.015
44	0.012
45	0.010
46	0.008
47	0.007
48	0.005
49	0.004
50	0.004
51	0.003
52	0.002
53	0.002
54	0.002

55	0.001
56	0.001
57	0.001
58	0.001
59	0.000
60	0.000

Surrogate model. The GI data surrogate model relies on the rate of glucose intake in the input model, with standard deviations reflecting data uncertainties.

Table S9. Conditional probability distributions for Gaussian variables in the GI data model.

Variable name	**Description**	**Values**	**Unit**
ΔG_d^{GS}	Rate of glucose intake from food digestion	Table S8	mM min^{-1}

1.8 GLP1 data model

Input model. The GLP1 data model defines classes for four discrete values of the plasma concentration of GLP1 (Table S10); the classification is the same for both normal and T2D patients.

Table S10. The GLP1 data model.

Level	**GLP1 [pM]**
low	400
medium	800
medium-high	1200
high	1600

Surrogate model. The GLP1 data surrogate model is the observation made in the corresponding model. It becomes probabilistic only during the coupling stage, in which a conditional PDF of the $GLP1^{GL}$ variable relates it to a coupling variable $GLP1^{C}$, as described in SI Appendix: Supplementary Text 2.

Table S11. Conditional probability distributions for Gaussian variables in the GL data model.

Variable name	Description	Values	Unit
$GLP1^{GL}$	Plasma GLP1 concentration	400 (low) 800 (medium) 1200 (medium-high) 1600 (high)	pM

2. Coupling and coupling variables

In the coupling stage, coupling variables were introduced as random variables with linear Gaussian conditional PDFs, relating them to variables from surrogate models (Table S10). In addition, the conditional PDFs of some variables from surrogate models were modified to include dependencies on coupling variables (Table S11). When variables from several models inform a coupling variable, our prior assigns even confidence to information from all variables. For instance, the coupling variable S^{C}_{cell} is statistically coupled to S^{VE} and S^{Sg} with equal weights. When additional prior information is available, it is used to inform the coupling. For example, intracellular glucose concentration has been estimated to be half of the plasma glucose concentration (11).

Table S12. Conditional PDFs for Gaussian variables of the coupling variables.

Coupler	Description	Time slice, [min]	Mean	Std-dev	Unit
G^{C}_{pl}	Plasma glucose concentration	t	$G^{PR}(t)$	10^{-3}	mM
G^{C}_{cell}	Intracellular Glucose concentration	t	$G^{C}_{pl}(t)/2$	10^{-3}	mM
$GLP1R^{C}$	GLP1R activity	t	$1.0\ GLP1R^{VS}(t)$	10^{-4}	-
ATP^{C}_{INS1E}	Intracellular ATP concentration	t	$1.0\ ATP^{Mb}_{INS1E}(t)$	10^{-3}	mM
S^{C}_{pa}	Insulin secretion rate of the pancrea	t	$S^{Pa}_{pa}(t)$	10^{-3}	pM min^{-1}
S^{C}_{cell}	Insulin secretion rate of one β-cell	t	$S^{Sg}(t)\ /2\ + S^{VE}(t)/2$	10^{-11}	pM min^{-1}
ΔG^{C}_{d}	Rate of glucose intake from food digestion	t	$1.0\ \Delta G^{PR}_{d}$	10^{-6}	mM min^{-1}
$GLP1^{C}$	Plasma GLP1 concentration	t	$1.0\ GLP1^{Sg}(t)$	10^{-3}	pM

Table S13. Modified conditional PDFs for variables in surrogate models that become dependent on coupling variables after the coupling stage.

Variable name	Description	Time slice, [min]	Mean	Std-dev	Unit
S^{PR}	Pancreatic insulin secretion rate	$t + \Delta t$	$(Y^{PR}(t + \Delta t) + K^{PR}\Delta G_d^{PR}(t) + 1.0\, S_b^{PR}(t + \Delta t))/3 + 2/3\, S_{pa}^{C}(t + \Delta t)$	10^{-3}	pM min^{-1}
S_{cell}^{Pa}	Insulin secretion of a cell	t	$S_{cell}^{C}(t)$	10^{-10}	pM min^{-1}
G^{VE}	Basal intracellular glucose concentration	t	$1.0\, G_{cell}^{C}(t)$	10^{-3}	mM
G^{Sg}	Intracellular glucose concentration	t	$1.0\, G_{cell}^{C}(t)$	10^{-3}	mM
ATP^{Sg}	Intracellular ATP concentration	$t + \Delta t$	$0.9\,(\alpha^{Sg} G_{ic}^{Sg}(t) + k_{ATP}^{Sg} ATP^{Sg}(t)) + 0.11\, ATP_{INS1E}^{C}(t + \Delta t)$	10^{-4}	mM
$GLP1R^{Sg}$	GLP1R activity	t	$k_{GLP1}^{Sg} GLP1^{Sg}(t) + GLP1R^{C}(t)$	10^{-3}	-
ΔG_d^{GI}	Rate of glucose intake from food digestion	t	$1.0\, \Delta G_d^{C}(t)$	10^{-6}	mM min^{-1}
$GLP1^{GL}$	Plasma GLP1 concentration	t	$1.0\, GLP1^{C}(t)$	10^{-3}	pM

Additional Supplementary Figures and Movie

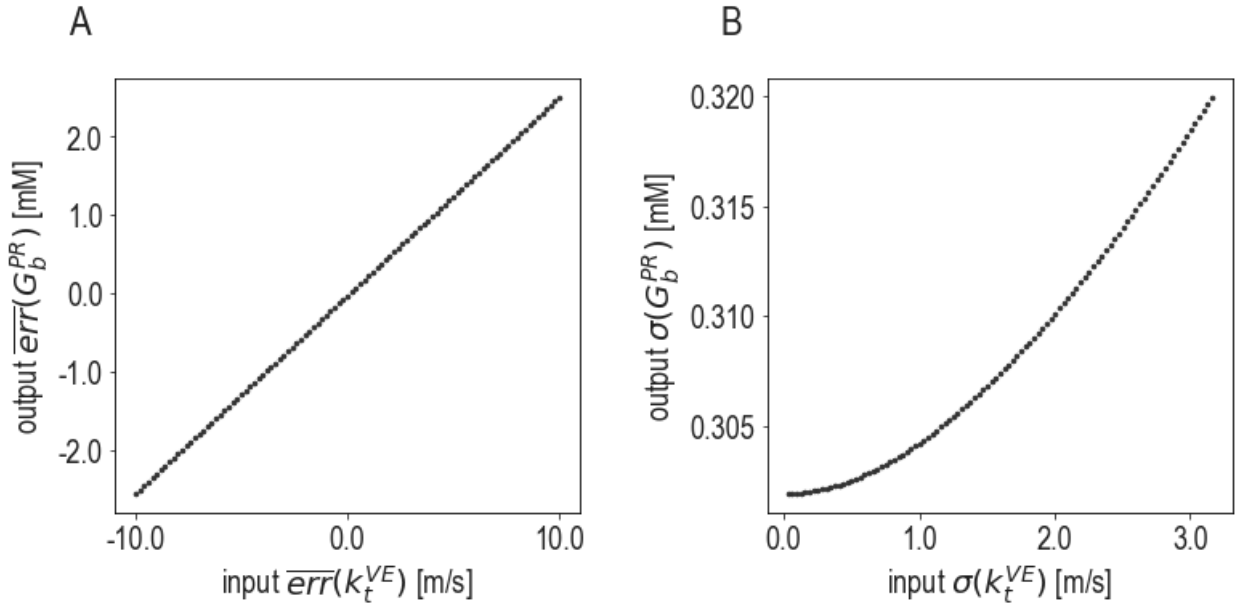
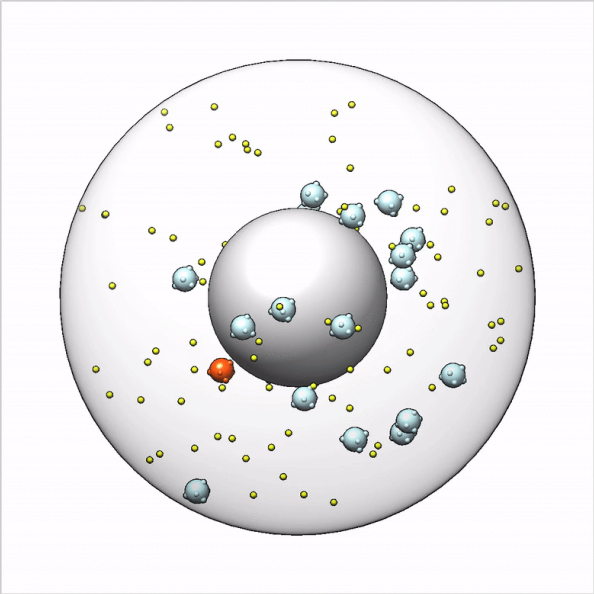


Figure S14. Effect of metamodeling on model accuracy and precision. (A) Statistical dependency of the output systematic error ($\overline{err}$) of the variable G_b^{PR} in the postprandial response model on the input systematic error of the variable k_t^{VE} in the vesicle exocytosis model. The input systematic error for G_b^{PR} is 0.0. The coupling coefficient corresponds to the slope of the line. (B) Statistical dependency of the output random error (σ) of k_t^{VE} on the input random error of G_b^{PR}. Input k_t^{VE} was presented as evidence with equal values for all time steps. The input random error for G_b^{PR} is 1.0. For both (A) and (B), the reference values used for computing the systematic and random errors of G_b^{PR} and k_t^{VE} are 5.1 mM and 10.0 m/s, respectively. All output values are at *t* = 100 min.



Movie S1. A sample coarse-grained Brownian dynamics trajectory contributing to the vesicle exocytosis input model. A 32.8 μs trajectory is shown. See legend of Figure S3. A single insulin vesicle is colored in red, highlighting the spatiotemporal trajectory of a single secretion event, including glucose activation, transport, and secretion. Following secretion, the vesicle is “reborn” inside the cytoplasm, modeling a new vesicle biogenesis event. The simulation time step is 10 nanoseconds, and the animation shows snapshots at 200 nanosecond intervals.

References

1. C. Dalla Man, R. A. Rizza, C. Cobelli, Meal simulation model of the glucose-insulin system. *IEEE Trans. Biomed. Eng.* **54**, 1740–1749 (2007).

2. D. Koller, N. Friedman, F. Bach, *Probabilistic Graphical Models: Principles and Techniques* (MIT Press, 2009).

3. A. Pisania, *et al.*, Quantitative analysis of cell composition and purity of human pancreatic islet preparations. *Lab. Invest.* **90**, 1661–1675 (2010).

4. C. Ionescu-Tirgoviste, *et al.*, A 3D map of the islet routes throughout the healthy human pancreas. *Sci. Rep.* **5**, 14634 (2015).

5. L. Marselli, *et al.*, Are we overestimating the loss of beta cells in type 2 diabetes? *Diabetologia* **57**, 362–365 (2014).

6. L. Sun, A. Sali, Data-driven Brownian dynamics simulations: Application to insulin secretory pathway in pancreatic β-cells. *to be submitted*.

7. K. L. White, *et al.*, Visualizing subcellular rearrangements in intact β-cells using soft X-ray tomography. *Science Advances* (2020).

8. S. S. Thazhath, *et al.*, The glucagon-like peptide-1 (GLP-1) receptor agonist, exenatide, inhibits small intestinal motility, flow, transit and absorption of glucose in healthy subjects and patients with type 2 diabetes: a randomised controlled trial. *Diabetes*, db150893 (2015).

9. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

10. T. Redij, R. Chaudhari, Z. Li, X. Hua, Z. Li, Structural Modeling and in Silico Screening of Potential Small-Molecule Allosteric Agonists of a Glucagon-like Peptide 1 Receptor. *ACS Omega* **4**, 961–970 (2019).

11. M. Fehr, S. Lalonde, I. Lager, M. W. Wolff, W. B. Frommer, In vivo imaging of the dynamics of glucose uptake in the cytosol of COS-7 cells by fluorescent nanosensors. *J. Biol. Chem.* **278**, 19127–19133 (2003).