



Pázmány Péter Katolikus Egyetem

Információs Technológiai és Bionikai Kar

Tutored Research Project - Report

Mathematical Modeling of Autophagy

Krisztián Szuppinger

Molecular Bionics Engineering BSc

2025

Supervisors: Bence Hajdú, Orsolya Kapuy

Internal Consultant: Attila I. Csikász-Nagy

CONTENTS

I	Introduction	1
II	Autophagy - Biological Background	1
II-A	Stages of Autophagy	1
II-B	Holistic Approach - The Crosstalk Between Autophagy and Apoptosis	2
III	Modeling Autophagy and Apoptosis	3
III-A	BCRN Model for Autophagy-Apoptosis Crosstalk	3
III-B	An Example through the mTOR Module	3
IV	Workflow and Rationale of the Statistical Analysis	4
V	Initial Condition Sampling Strategies	5
VI	Correlation Matrix Convergence	6
VII	Covariance Matrix and its Convergence	7
VIII	Conclusions	10
	References	10

Abstract

Autophagy is a conserved process in eukaryotic cells where vesicles form to degrade cellular components, helping to maintain energy balance and remove damaged molecules. Closely linked to apoptosis, autophagy can either promote or suppress tumor cell proliferation, thus influencing cell survival and rendering the process crucial in various physiological and pathological conditions. This report introduces a large-scale biochemical reaction kinetics model for basal state autophagy-apoptosis crosstalk. This is followed by a detailed statistical analysis of said computational model navigating its descriptive properties. Model performance is examined using several thousands of different initial conditions in order to determine the optimal sampling strategy yielding the most accurate predictions about the time course of key proteins in the autophagy-apoptosis signaling network. The empirical correlation and covariance matrices of the model output are shown to be convergent, allowing for the definition of a probability density function that characterizes the distribution of the initial distribution of all 84 variables.

I. INTRODUCTION

In this semester, I have assisted the research group of the Kapuy laboratory at Semmelweis University. The group aims to model the decision-making mechanisms governing autophagy and apoptosis, examining how these processes interact and determine cell fate decisions. Specifically, they focus on how various causal events drive the cell toward one phenotypic outcome over the other. Understanding the key determinants of autophagy can provide insights into various diseases in which autophagy plays a crucial role. Famous examples of pathologies that involve the dysfunction of autophagy span from neurodegenerative disorders like Alzheimer's through chronic inflammatory conditions such as IBD to tumorigenesis [1][2]. Understanding this cellular process could guide the design of new drugs that combat these diseases [3][4].

The aim of my research during the semester was to elaborate on the statistical properties of a homeostatic state biochemical reaction network (BCRN) model previously developed by Orsolya Kapuy's group [5]. This work serves the long-term goal of this research group, which is to fine-tune the current model such that it becomes applicable not only to basal, but also to non-homeostatic cellular conditions. The results that I will introduce here may facilitate this transition by informing our understanding of the intricate connections in this large-scale network.

In this document, I will provide a short summary of the biology of autophagy, focusing on the main pathways and networks that regulate this evolutionarily conserved process. This will be followed by introducing the BCRN and the computational optimization tool used for modeling the network. I will also present the results of the statistical analysis that I have performed along with its main conclusions and consequences for the model. Finally, future research plans will be described.

II. AUTOPHAGY - BIOLOGICAL BACKGROUND

Autophagy is a catabolic process responsible for cellular degradation and the recycling of proteins and organelles through lysosomal digestion [6]. In addition to helping cells adapt to stressful conditions, autophagy plays a crucial role in regulating cell growth, metabolism, and survival [7]. A basal level of autophagy is considered cytoprotective because it removes misfolded or unnecessary proteins, thereby maintaining cellular homeostasis [8]. Furthermore, autophagy is critical for mobilizing key nutritional elements such as carbohydrates (glycophagy), lipids (lipophagy), and minerals (ferritinophagy), promoting cell survival by recycling these nutrients [9]. It is also rapidly activated in response to starvation and various other forms of stresses, including e.g. hypoxia, or metabolic stress.

A. Stages of Autophagy

The molecular machinery underlying autophagy is highly complex and involves numerous kinases, phosphatases, and GTPases, encoded by autophagy-related genes (ATG) [10]. The autophagy process can be divided into five distinct steps: initiation, autophagosome nucleation, expansion, completion, and docking and fusion with the lysosome. During the initiation of autophagy, proteins or organelles are enclosed by the autophagosome, which subsequently fuses with a lysosome. This fusion forms the autolysosome, allowing lysosomal enzymes to digest the contents of the vacuoles to their biological building blocks, which are then released back into the cytoplasm by permeases to be reused by biosynthetic pathways [11][12].

The autophagic program involves various signalling cascades, such as calcium or inositol pathways, and

is closely interconnected with other cellular processes, notably apoptosis [13]. The relationship between autophagy and apoptosis is of particular interest, as both are triggered by cellular stress but lead to opposing outcomes: autophagy supports cell survival, whereas apoptosis results in programmed cell death [14].

B. Holistic Approach - The Crosstalk Between Autophagy and Apoptosis

It is well understood that autophagy removes dysfunctional components within the cell, while apoptosis leads to programmed cell death [15][16]. However, the mechanisms by which cells evaluate damage and determine whether to initiate autophagy or apoptosis, thus making a ‘life’ or ‘death’ decision, remain unclear [17]. Stress signals, such as hypoxia, deprivation of amino acids or growth factors, can trigger either process. Currently, the dominant opinion is that autophagy is a mostly cytoprotective process that is triggered by such stresses [18]. However, if the harmful conditions are prolonged, then at some point the affected cells of a multicellular organism enter a controlled death program in order to prevent the formation of necrotic lesions [14]. A better understanding of the interplay between autophagy and apoptosis is crucial, as they are both often abnormally regulated during the development of tumors and their acquisition of resistance [19]. Uncovering the intricate connections between autophagy and apoptosis is precisely what a data-driven computational approach may facilitate [20], thus in the next section, I will describe the model employed in our research.

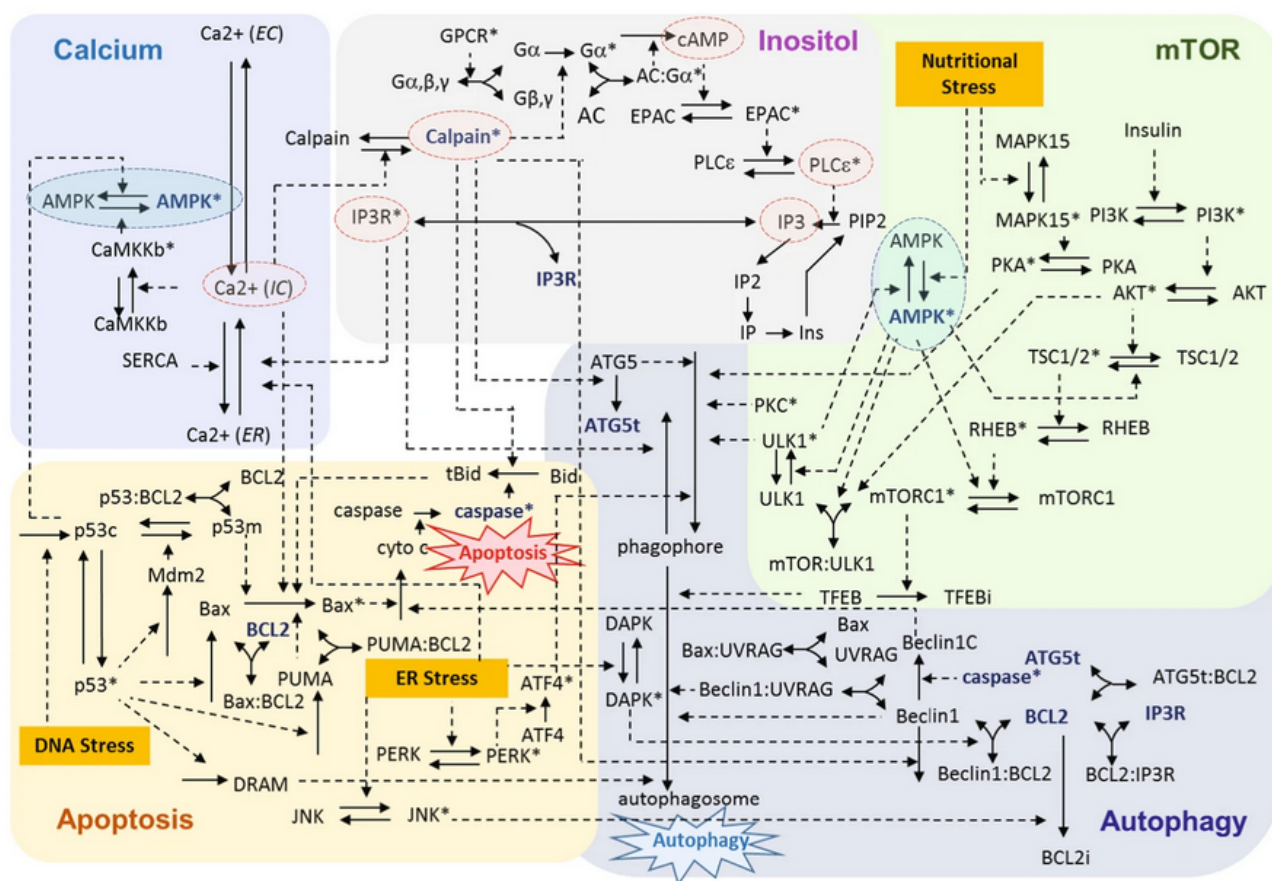


Figure 1: The diagram illustrates a detailed network of interactions among proteins, ions, and metabolites, organized into five interconnected modules: calcium, inositol, mTOR, apoptosis, and autophagy [17]. Each module is visually distinguished by a unique background color. Solid arrows denote physical interactions such as binding, dissociation, or translocation, while dashed arrows represent chemical reactions. Critical mediators of cellular responses, such as AMPK, IP3R, Bcl-2, Bax, and Atg5, are displayed in multiple modules for better visualization of their roles. Key interactions and components identified as pivotal to cellular processes are highlighted in red or blue ellipses. The three yellow blocks highlight cellular stresses and represent the inputs of the model. The starburst shapes (encompassing either autophagy or apoptosis) constitute the main output of the system, which could either be a decision for autophagy, or for apoptosis.

III. MODELING AUTOPHAGY AND APOPTOSIS

The computational model used in our research utilizes a (bio)chemical reaction network (CRN) framework, in which the network of proteins is represented as a graph with the species (i.e., proteins) as the nodes of the graph and the defined reactions as the edges. Each edge is weighted with the corresponding k rate constant as defined by mass action kinetics. The result of such a framework is a system of ordinary differential equations (ODE), each ultimately characterized by the corresponding k constant.

In this section I will briefly introduce the model, as well as provide a simple example (1) demonstrating how mass action kinetics is applied in this biological setting.

A. BCRN Model for Autophagy-Apoptosis Crosstalk

The network implemented in the computational model includes the key signaling pathways initiating autophagy and apoptosis in response to different stress stimuli. The system is composed of five main units. These are the autophagy, the apoptosis, the mTOR, the inositol and the Ca^{2+} modules. Each of these modules includes elements of one or more major signaling pathways that govern cell fate decisions and are activated by genotoxic, nutritional, or ER stress signals. The wiring diagram of the model is depicted in figure 1.

This framework was developed by Liu et al. [17], however, the original model failed to reproduce even the basal conditions as it provided outputs that indicated elevated levels of autophagy and/or apoptosis even in stress-free, homeostatic conditions [5]. The system of equations governing the reaction network was, therefore, redesigned and the influential kinetic parameters were optimized to correct the shortcomings of the prior model. The current computational framework uses the FOCTOPUS algorithm and the CANTERA solver within the Optima++ environment to estimate and optimize the parameters and solve the system of differential equations. Cantera is a popular solver for problems in chemical reaction kinetics [21], Optima++ is a simulation environment originally developed for describing and optimizing parameters of large-scale combustion kinetics networks [22][23].

B. An Example through the mTOR Module

Here, I will introduce the basic concept of mass action kinetics when applied in a systems biological setting. We will consider the mTOR module from fig. 1. The mammalian target of rapamycin (i.e., mTOR) is a well-studied master regulator molecule in various cell growth mechanisms [24][25][26][27]. The mTOR module in our framework centers on mTORC1 - the active complex; mentioned simply as mTOR* from here on out - which inhibits autophagy by binding ULK1. Under stress, however, AMPK suppresses mTOR* and activates ULK1, promoting autophagy [28]. In the computational model of the basal state kinetics, there are four ODEs that govern the activity of mTOR. These are depicted below in eq. (1).

$$\begin{aligned}
 \text{ULK1}^* + \text{MTOR}^* &\xrightleftharpoons[k_{-1}]{k_1} \text{MTOR}^* \rightarrow \text{ULK1} &\Rightarrow -\frac{d[\text{MTOR}^*]}{dt} &= k_1[\text{ULK1}^*][\text{MTOR}^*] - k_{-1}[\text{MTOR}^* \rightarrow \text{ULK1}] \\
 \text{ULK1}^* + \text{MTOR}^* &\xrightleftharpoons[k_{-2}]{k_2} \text{MTOR} \rightarrow \text{ULK1}^* &\Rightarrow -\frac{d[\text{MTOR}^*]}{dt} &= k_2[\text{ULK1}^*][\text{MTOR}^*] - k_{-2}[\text{MTOR} \rightarrow \text{ULK1}^*] \\
 \text{AMPK}^* + \text{MTOR}^* &\xrightleftharpoons[k_{-3}]{k_3} \text{AMPK} \rightarrow \text{MTOR} &\Rightarrow -\frac{d[\text{MTOR}^*]}{dt} &= k_3[\text{AMPK}^*][\text{MTOR}^*] - k_{-3}[\text{AMPK} \rightarrow \text{MTOR}] \\
 \text{RHEB}^* + \text{MTOR} &\xrightleftharpoons[k_{-4}]{k_4} \text{RHEB}^* + \text{MTOR}^* &\Rightarrow \frac{d[\text{MTOR}^*]}{dt} &= k_4[\text{RHEB}^*][\text{MTOR}] - k_{-4}[\text{RHEB}^*][\text{MTOR}^*] \\
 \text{MTOR}^* &\xrightarrow{k_5} \text{MTOR} &\Rightarrow -\frac{d[\text{MTOR}^*]}{dt} &= k_5[\text{MTOR}^*]
 \end{aligned} \tag{1}$$

In this small example alone, there are already seven rate constants (i.e., k_i , and k_{-i}) and three non-linear differential equations (introduced by the inhibition kinetics). The equations essentially capture the basics of mTOR kinetics in a basal state without any stresses. It is important, however, to note here that the kinetics of mTOR are not simply influenced by these four equations in (1). In fact, the entire rationale of systems biology lies in the idea of looking at a holistic picture instead of just the direct interactions [29]. Other key mediators for example (that govern mTOR activity *indirectly*) include Atg5, Beclin-1, UVRAG, PKA and PKC as inhibitors. Furthermore, cross-regulation links apoptosis and autophagy: UVRAG inhibits Bax while promoting autophagy, Bcl-2 suppresses autophagy via Beclin-1, and activated caspases cleave Beclin-1 to shift toward apoptosis. AMPK itself (and thus mTOR as well) is regulated by $\text{CaMKK}\beta$ from the calcium module, which responds to cytoplasmic Ca^{2+} levels that is also influenced by the inositol module which involves IP_3 -triggered Ca^{2+} release from the ER [17]. The multitude of connections makes analysis untractable for humans alone, thus requiring a computational approach.

IV. WORKFLOW AND RATIONALE OF THE STATISTICAL ANALYSIS

The aim of this work was to elucidate the statistical properties of the existing basal state model for autophagy-apoptosis crosstalk. Following this description of the model, the goal is to introduce various stresses (e.g., starvation or rapamycin treatment), then assess and perfect model performance. In the full model there are 113 similar equations to the ones in (1), each with their own rate constants. This results in a highly complex system for which the determining rate constants are not well-known. Consequently, a data-driven, statistical approach was employed to fine-tune model parameters. The first step thus, involved reviewing the literature to set the species' concentration ranges to reasonable values. Details of this can be found in [5]. All in all, ranges were 'confidently' set for 34 out of 84 species. These serve as the primary predictors of model performance. Abbreviated names of the included molecules are given in figure 6, full names are found in the *Supplementary Material*. Literature search is followed by the first stage of the actual modeling, which is the selection of the initial conditions (IC) for each species. To choose these values, a uniform random sampling process was utilized, that generated ICs from the $[0.5 \cdot \bar{x}, 1.5 \cdot \bar{x}]$ range, where \bar{x} depicts the empirical mean as revealed by the literature search. Another sampling strategy was also utilized, where bounds were set to the whole theoretical range. In both cases, $|\sigma_i|$ was calculated as $1/8$ of the sampling range. The ICs are then saved into an xml file, which will serve as the input of the Optima++ environment. For basal state modeling homeostatic concentration values for the 34 followed species are also generated, saved into the xml, and serve as the 'measurements' which Optima++ tries to approximate. These datapoints constitute a 24 hour time series (i.e., one day of measurement data) and consist of the mean values for all 34 of the predictor species. The generated xmls can be interpreted as slightly different experimental homeostatic conditions. After the simulation, the acceptability of the output is assessed for each input condition (i.e., xml). If the concentrations of the 34 followed variables are within the established theoretical bounds, then the input conditions are approved, and the output can be used for statistical analysis. This workflow process is summarized in fig. 2.

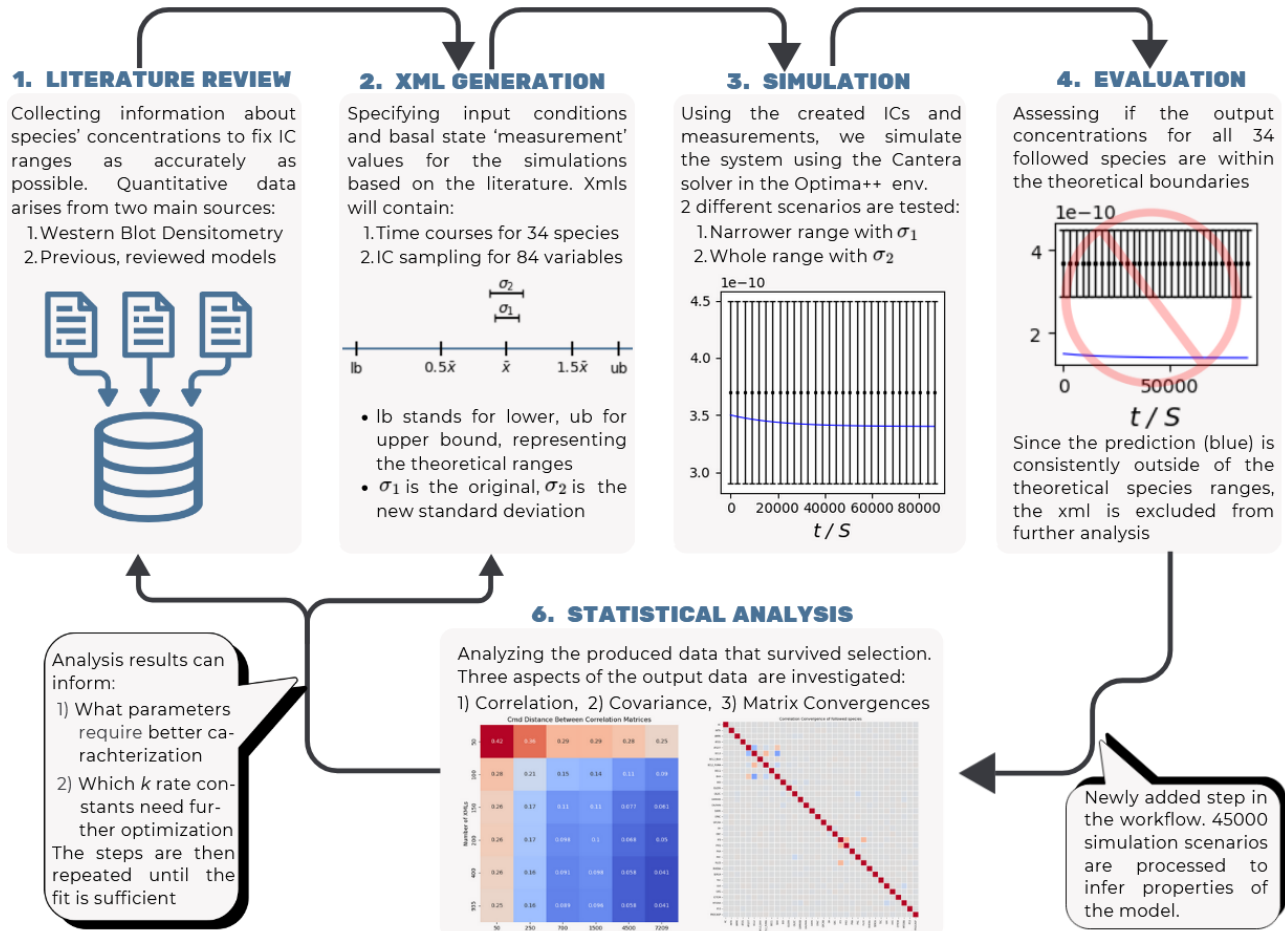


Figure 2: Summary of the applied workflow. Listed steps follow each other in a pipeline-like fashion.

The specific questions asked in this work can be grouped into three categories. These are as follows:

- 1) How sensitive is the output of the current model to the chosen initial condition sampling strategy? Does a wider sampling range with higher variance lead to a better or worse model output in terms of the concentrations of the followed 34 species?
- 2) What can be said about the convergence of the correlation matrix of the output for all 84, as well as for the 34 followed species of the model? How accurately do these correlations resemble known biological connections?
- 3) How can we assess the convergence of the covariance matrix of the output? What can we use the covariance matrix for, given that it converges?

V. INITIAL CONDITION SAMPLING STRATEGIES

The first question we asked concerned the sampling process of the initial conditions. As demonstrated in fig. 2 (in the 'XML GENERATION' step), the original sampling range is narrower with a smaller standard deviation (std) in $|\sigma_1^{(s)}| = 1/8 \cdot ||0.5 \cdot \bar{x}_s, 1.5 \cdot \bar{x}_s||$, where s stands for the species in question. The new approach used the whole literature inspired range from the lower bound (lb) to the upper bound (ub). The corresponding standard deviation, therefore, was also different as $|\sigma_2^{(s)}| = 1/8 \cdot ||lb_s, ub_s||$. In both cases ranges were checked only for the 34 predictor variables, using 15000 different xmls for each sampling strategy. For the new approach, we also produced 15000 xmls for which ICs were chosen from the whole range, but the std. was kept at the narrower σ_1 value.

The performance of the sampling strategies was quantified simply by the percentage of xmls (i.e., initial conditions) that produced results which stayed within a 4 std. radius of the corresponding species' mean. For the original sampling strategy this efficiency percentage was rather high at $\text{eff}_{\%}^{(\text{orig})} = \frac{14209}{15000} \cdot 100\% = 94.73\%$. The new sampling produced significantly worse results in both cases. Simulation results for 15000 xmls of new IC sampling, but narrower std. showed an efficiency of $\text{eff}_{\%}^{(\text{new1})} = \frac{91}{15000} \cdot 100\% = 6.1\%$. When the standard deviation used was σ_2 , the results were more acceptable at $\text{eff}_{\%}^{(\text{new2})} = \frac{1835}{15000} \cdot 100\% = 12.23\%$.

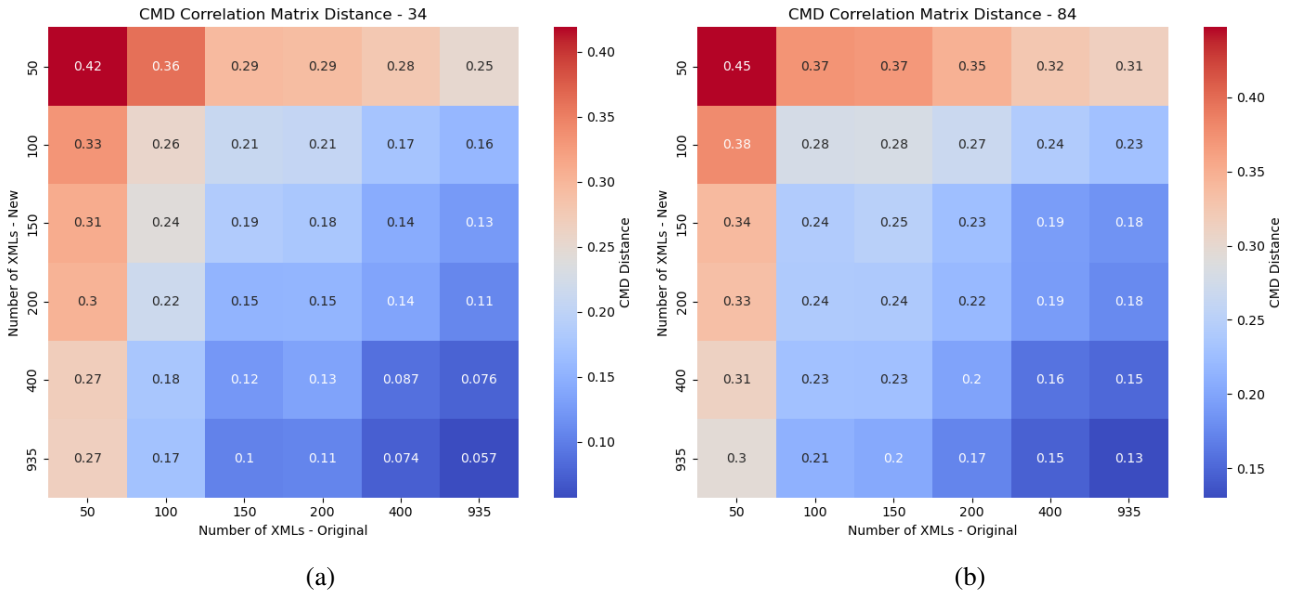


Figure 3: The two heatmaps depict the differences of the correlation matrices along the different xml counts as quantified by the CMD metric. The horizontal axis (x) stands for the narrower, original sampling strategy, while the vertical one (y) represents the new, wider approach. The xml counts (N_{cond} , or 'Number of xmls' in the figures) determine the number of simulation outputs used for the calculation of the correlation matrix. Each block, therefore, is a distance between correlation matrices estimated from a given number of outputs either from the new, or old sampling. The block at $x = 400$, $y = 935$, for example, shows the difference between $\text{corr}_{\text{orig}}^{(400)}$ and $\text{corr}_{\text{new}}^{(935)}$. In fig. 3a only the correlations of the 34 predictor variables are considered, whereas in fig. 3b, all 84 species of the model were used for the computations. In both heatmaps, it is visible that the distances significantly decrease as N_{cond} increases. From this, we can conclude that results of the two sampling procedures do in fact approach the same correlations as $N_{\text{cond}} \rightarrow \infty$.

This result may be a consequence of the fact that for multiple variables of the model, still no well-defined ranges exist. The simple solution for this thus, is to use the analysis results to move back to the first step of the workflow and perform a more extensive literature review (see fig. 2) or perhaps even experimentation. Another possibility, however, is that the theoretical boundaries set for the predictor variables represent physiological values accurately and the model performance is also satisfactory (i.e., it represents reality correctly). If this is true, then, when the ICs are sampled from the wider range, more outliers are chosen which drive model predictions astray more frequently. If the lower $\text{eff}_{\%}$ value is caused by the elevated number of outliers, then the distribution of the two outcomes produced by the different sampling strategies should be more and more similar (for the successful xmls) as the xml count N_{cond} increases. To assess this similarity as N_{cond} grows, we compared the correlation matrices - as these are the primary descriptors of the model's ability to capture biological connections - of the variables for sequentially increasing values of N_{cond} .

6 different N_{cond} values were considered. The number of compared conditions was kept at 1835, since there were no more successful xmls for the results of the wider sampling. The xml counts for the comparison are: 50, 100, 150, 200, 400, and 935. Our rationale is that if the correlations of the two groups of xmls get more and more similar as the xml count increases, then the two correlations should converge to the same matrix as $N_{\text{cond}} \rightarrow \infty$. This would also mean that - in terms of a statistical analysis - the two sampling strategies produce the same result.

Multiple distance metrics were considered to quantify correlation matrix similarities. To get a preliminary idea, we first looked at the Frobenius-norm of the difference matrices ($\mathbf{D} = \text{corr}_{\text{orig}} - \text{corr}_{\text{new}}$) across the given xml counts. This norm simply compresses an $(n \times n)$ matrix into a $(n \cdot n \times 1)$ vector and computes its euclidean norm: $\|\mathbf{D}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n |d_{ij}|^2}$. This metric is enough to provide an intuitive sense of the differences, but is insufficient if capturing the matrix structure is also desirable since the norm simply compresses the matrix into vectors. Therefore, we used the correlation matrix distance (CMD) metric [30] that is finely-tuned for correlation matrices of multiple input multiple output (MIMO) systems - i.e., systems that are fundamentally similar to our computational model. The CMD is defined as follows:

$$d_{12}^{\text{CMD}}(\mathbf{R}_1, \mathbf{R}_2) = 1 - \frac{\text{tr}\{\mathbf{R}_1 \mathbf{R}_2\}}{\|\mathbf{R}_1\|_F \|\mathbf{R}_2\|_F} \in [0, 1], \quad \text{where } \mathbf{R}_i \text{ is the correlation matrix.} \quad (2)$$

The CMD is zero if the two correlation matrices are identical (up to scaling), and one if they are completely orthogonal (since the trace - $\text{tr}\{\}$ - would be 0).

The pairwise distances indicate that as $N_{\text{cond}} \rightarrow \infty$, the output correlations converge to each other. This is depicted in fig. 3. Unfortunately, comparison of covariances was not possible, as the number of successful xmls from the new procedure was too little (see in more detail in section VII). For further analysis, I used the results of the narrower sampling strategy, as the much bigger number here simply allowed for a more thorough statistical analysis, and the correlations of the two sampling processes are the same as $N_{\text{cond}} \rightarrow \infty$.

VI. CORRELATION MATRIX CONVERGENCE

Two central questions that this analysis is meant to answer are whether the correlations of the model's output approach a given correlation matrix, and if they do, does this convergent matrix represent the biological reality - i.e., the known connections - correctly. To assess all this, first the convergence of the correlation matrices was examined by looking at the CMD metric between correlation matrices computed from a sequentially increasing number of xmls. The idea here, therefore, is the same as it was in the case of the previous convergence analysis in section V.

The capped xml counts are as follows: 50, 250, 700, 1500, 4500, and $14209 - 7000 = 7209$. The distances are again plotted in the form of heatmaps in fig. 4. Correlation matrices for all 84 basal state variables (fig. 4b) show convergence, and therefore, we may conclude that the correlation matrix computed from all 14209 outputs will estimate the true correlation matrix of the model. We thus calculated the correlation matrix of the system using the results from all 14209 initial conditions.

Figure 6 shows the correlation matrix for every species of the model. These correlations largely agree with the known molecular connections and signaling networks, and consequently, we can argue that the model represents real interactions accurately. For the detailed analysis of these connections please refer to the description of fig. 6.

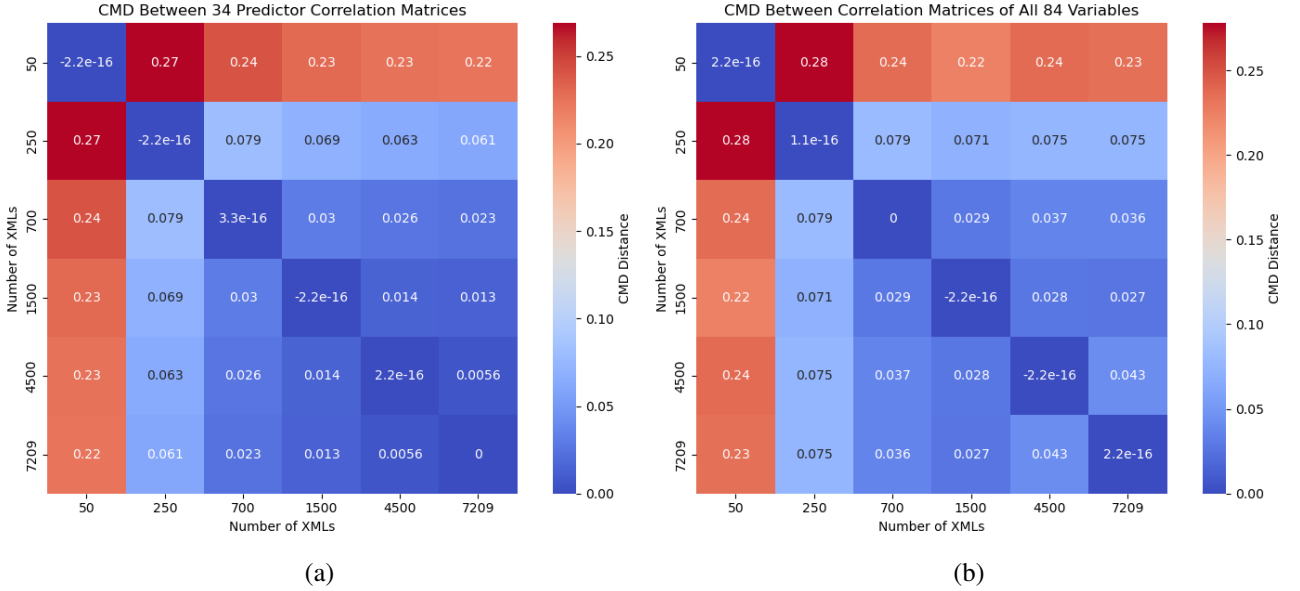


Figure 4: The two figures depict the correlation matrix differences across correlation matrices computed from different numbers of xmls. Figure 4a in the left is computed from matrices that refer only to the 34 followed variables, whereas fig. 4b uses all 84 variables. Like in figure 3, the value of a given block refers to the CMD between the two corresponding matrices. An important difference here is that this heatmap is symmetric, because the same sequence of correlation matrices are found on the two axes. The consequence of this is that the diagonal values are extremely low, since these values represent the distance of the given matrix from itself. The conclusion that is to be drawn from these figures is that the distance (i.e. the difference) decreases as N_{cond} increases. In fact, the decrease is apparent not only for the predictors (fig. 4a), but also for all other variables (fig. 4b), even though the correlation for the whole system converges more slowly - possibly due to the bigger uncertainty in the non-predictor variables. The convergence justifies the inference of model characteristics from the correlation matrix that is calculated from all 14209 variables.

VII. COVARIANCE MATRIX AND ITS CONVERGENCE

The last question we tried to answer was whether the covariance matrix of the model output was convergent. As mentioned previously in section V, the covariance matrix was not well-conditioned for analysis in case of the low sample size of 1835 xmls. However, for the 14209 different ICs of the original sampling, we could derive stable enough covariance matrices to perform a statistical examination of convergence, again, via heatmaps of covariance distances across different N_{cond} values.

The different N_{cond} values are again 50, 250, 700, 1500, 4500, and 7209. Covariance distances primarily describe distribution distances, and thus metrics used for correlations may not be sufficient for covariances. Therefore, we considered two different metrics for the quantification of covariance matrix distances. Firstly, the Affine-Invariant Riemannian Metric (AIRM) (3), which is a distance metric with applications e.g., in image processing in the field of Neuroinformatics [39]. Riemannian distances are defined as curve lengths on Riemannian manifolds [39]. This metric was chosen because it provides a general statistical framework, and is an invariant metric [40] - a property that could be useful for our ill-conditioned covariance with several diverging entries. The definition of AIRM is given below in eq. (3).

$$d_{12}^{\text{AIRM}}(\Sigma_1, \Sigma_2) = \left\| \text{Log} \left(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} \right) \right\|_F = \left(\sum_{c=1}^C \log^2 \lambda_c \right)^{1/2}, \quad (3)$$

where, Log is the matrix logarithm, and the λ_c eigenvalues correspond to the argument of $\text{Log}()$. Secondly, we applied the Kullback-Leibler (KL) divergence metric (4), which specifically quantifies differences between probability distributions [41]. The KL metric is derived from the Shannon entropy formula: $f(x) = x \log(x)$, and is thus also known as the relative entropy [39]. The KL formula:

$$d_{12}^{\text{KL}}(\mathcal{N}_1(\mu_1, \Sigma_1) \parallel \mathcal{N}_2(\mu_2, \Sigma_2)) = \frac{1}{2} \left(\log \frac{|\Sigma_2|}{|\Sigma_1|} + \text{tr}\{\Sigma_2^{-1} \Sigma_1\} + (\mu_2 - \mu_1)^\top \Sigma_2^{-1} (\mu_2 - \mu_1) - n \right) \quad (4)$$

Here, $\mathcal{N}_1(\mu_1, \Sigma_1)$ $\mathcal{N}_2(\mu_2, \Sigma_2)$ are both n dimensional normal distributions, with $\mu_i = \vec{\mu}_i$ as the mean vector, and Σ_i as the covariance matrix. This formulation of the metric thus assumes normal distribution, which could cause instability in matrices if the true distribution is not normal. Quantifying this metric may, therefore provide some insight into how well the data is estimated by a normal distribution.

Both previously mentioned covariance distances require symmetric positive definite (SPD) input matrices. This condition is not fulfilled for the covariances that involve all variables (the covariance of only the predictor variables is in SPD space). An elegant solution to this issue is to use a linear combination of the original matrix and the identity matrix to compute the metric. This idea originates from the 2001 paper of Leodit & Wolf [42]. The transformation is defined in eq. (5)

$$\hat{\Sigma}_n^* = (1 - \gamma_n)\hat{\Sigma}_n + \gamma_n\hat{\mu}_n\mathbf{I}_n, \quad \text{with} \quad \hat{\mu}_n = \langle \hat{\Sigma}_n, \mathbf{I}_n \rangle = \frac{1}{s} \sum_{i=1}^s \sigma_{ii}^{(n)} \quad (5)$$

$\hat{\Sigma}_n$ is the sample covariance matrix - estimated from a sample of size n - with dimensions $(s \times s)$. $\hat{\Sigma}_n^*$ then, is the shrunk sample covariance. \mathbf{I}_n is the identity matrix, and s is the number of variables. The critical parameter is the *shrinkage* coefficient γ_n . The choice of γ_n is such that it minimizes the squared Frobenius-norm of the error between the true covariance Σ_n and the estimated one $\hat{\Sigma}_n$ [43].

Using the shrunk covariance indeed yielded SPD matrices for all xml counts, and thus we could compute the differences. The heatmaps depicting this are shown in fig. 5. Again, we can conclude from the heatmaps that the covariance matrix converges.

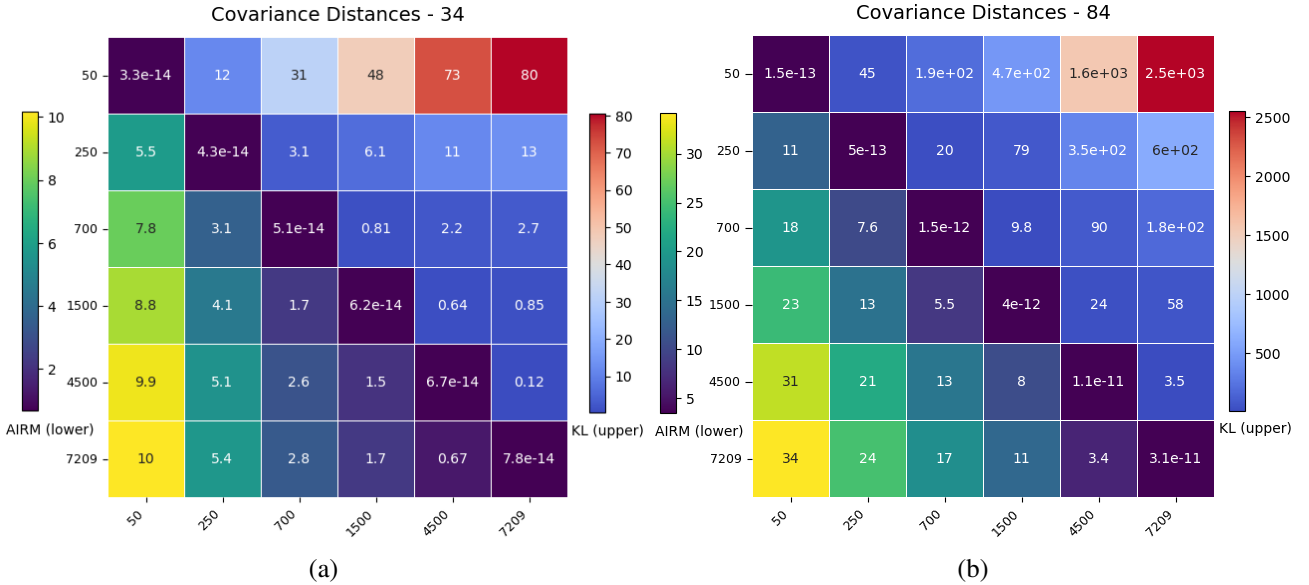


Figure 5: The figure depicts the distance heatmaps for the covariance matrices computed for sequentially increasing values of N_{cond} . Differences are quantified using the AIRM metric (lower triangle and diagonal) as well as the KL divergence measure (upper triangle). Both axes show N_{cond} i.e., the number of xmls used for the calculation of the covariance matrices. The figure in the left is computed only from the 34 followed species, which results in more stable results (even after shrinkage), and consequently, the distance values do not explode with small N_{cond} values. In contrast, in the right, we can see that the heatmap created from all 84 variables has a slightly divergent upper corner. This could indicate that the distribution is non-normal. However, as N_{cond} increases, the divergence disappears, possibly signaling a normal distribution as $N_{\text{cond}} \rightarrow \infty$. The key result is that (in both cases) the covariance differences decrease as N_{cond} increases.

The primary consequence of the convergent covariance matrix is that it (along with the KL decrease) justifies the definition of a multivariate normal distribution (mvnd) of the form given below, in eq. (6).

$$f(\mathbf{x}, \hat{\boldsymbol{\mu}}, \hat{\Sigma}_n^*) = \frac{1}{\sqrt{|\hat{\Sigma}_n^*|(2\pi)^s}} \exp\left(-\frac{1}{2}(\mathbf{x} - \hat{\boldsymbol{\mu}})^\top (\hat{\Sigma}_n^*)^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}})\right) \quad (6)$$

Here, \mathbf{x} and $\boldsymbol{\mu}$ are $(1 \times s)$ vectors and $\hat{\Sigma}_n^*$ is the Leodit-Wolf shrunk symmetric, positive definite $(s \times s)$ matrix calculated from all 14209 ICs. Future research will focus on elaborating the utility of this mvnd.

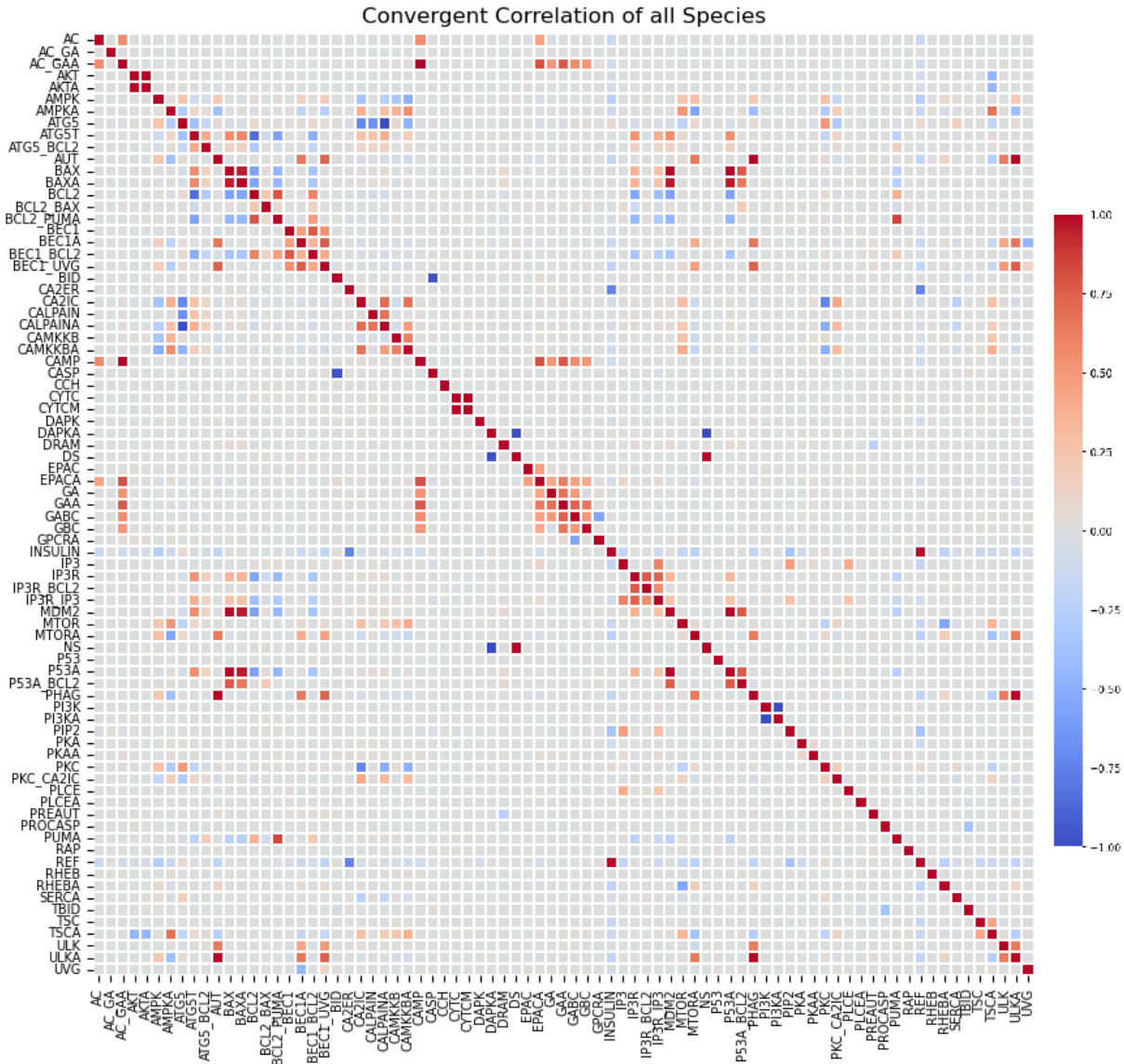


Figure 6: The figure presents the correlations in the model output for the whole system as computed from 14209 simulations. The full names of the species that are listed on the two axes are found in the [Supplementary Material](#). From this matrix, SERCA, SERCA*, thapsigargin (TG), and the TG-SERCA complex were removed, as their correlations were zero with all variables. As expected, several correlations can be observed in the matrix. A notable connection for example is the negative correlation between the ATG5 proteins, CA2IC (i.e., intracellular Ca^{2+}), and Calpain. This reinforces the known relationship in which Ca_{ic}^{2+} activates calpain, and calpain inactivates ATG5t, turning it into ATG5t (thus the positive correlation between the truncated ATG5t, Ca_{ic}^{2+} , and calpain) [31][32]. Another nice example from the calcium module of fig. 1 is the correlation between the concentrations of cAMP and the various G-protein constituents (e.g., GA which stands for the α subunit) [33]. The relationship between AMPK and CaMKKb, Ca_{ic}^{2+} , and calpain is also captured by the model, connecting the calcium and mTOR units from fig. 1 [34]. The activating link between BAX and P53 is present in the correlation matrix as well. P53 activates Bax in a simple transducer circuit, as well as through disinhibition, as it down-regulates Bcl2 - hence the negative correlation between these two in the figure [35]. An important regulatory link from the autophagy module is reassured by the ULK(A)-PHAG-AUT positive correlation. PHAG stands for phagophore, while AUT represents autophagosomes which are signals of active autophagy [36]. Phagophores are an intermediary step that lead to the formation of autophagosomes, ultimately resulting in autophagy [37]. ULK1 is present in the mTOR unit of fig. 1, it is inhibited by mTOR, but activated by AMPK* [28]. As an interesting last example, an established role of Mdm2 is to inhibit p53, which should lead to the decrease of Bax amounts. In the model, however, the correlation between Mdm2, and Bax is positive. This may arise from the fact that p53 activates both Bax, and Mdm2 (this also, is captured by the model, since Mdm2 positively correlates with p53). Thus, elevated levels of p53 might lead to increases in both species' concentrations [38].

VIII. CONCLUSIONS

As more and more information surfaces about the key roles of autophagy in numerous physiological and pathological conditions [44][45] it is of utmost importance to increase our understanding of this absolutely vital cellular process. The role of autophagy in the rapidly emerging field of aging biology for example, is becoming more and more crucial [46][47]. Mounting evidence appears to suggest that the suppression of mTOR or the up-regulation of AMPK and thus of autophagy and the resulting recycling of old and damaged cellular constituents are all linked to biological aging [47][48][49]. The consequence of this is that all pathologies that tend to be more frequent in elderly patients (e.g., Huntington's or cancer) will have some intricate connection to autophagy and its dysregulation [50][51][52]. For all these reasons, in aging western societies primary emphasis should be placed on the enhancement of our understanding of autophagy. Our research, therefore, tries to integrate the historical reductionist approaches of molecular biology with the relatively new field of systems biology [29], aiming to provide a holistic model that includes multiple signaling pathways that are normally treated separately, all in order to further inform our understanding of autophagy.

In this work specifically, I analyzed the properties of an existing, basal state computational model of autophagy [5]. The assessment of different sampling strategies suggested that a narrower input range better serves the accuracy of the model predictions. Analysis of the *convergent* output correlation matrix indicated that the biologically known interactions are preserved in the model. Finally, an mvnd was defined resulting in a new sampling distribution that could serve as a better basis either for the basal or the stressed conditions.

Our future research will focus on deriving a mathematically more rigorous convergence analysis and ultimately on the successful integration of various stresses into the model in order to enable us to draw valid conclusions that may provide insights that can inform and accelerate experimental approaches that address the etiology and progression of the aforementioned, devastating pathologies.

REFERENCES

- [1] Jiang, P., & Mizushima, N. (2014). *Autophagy and human diseases. Cell research*, 24(1), 69–79. <https://doi.org/10.1038/cr.2013.161>
- [2] Shao, B. Z., Yao, Y., Zhai, J. S., Zhu, J. H., Li, J. P., & Wu, K. (2021). *The Role of Autophagy in Inflammatory Bowel Disease. Frontiers in physiology*, 12, 621132. <https://doi.org/10.3389/fphys.2021.621132>
- [3] Pasquier B. (2016). *Autophagy inhibitors. Cellular and molecular life sciences : CMLS*, 73(5), 985–1001. <https://doi.org/10.1007/s00018-015-2104-y>
- [4] Chen, J. L., Wu, X., Yin, D., Jia, X. H., Chen, X., Gu, Z. Y., & Zhu, X. M. (2023). *Autophagy inhibitors for cancer therapy: Small molecules and nanomedicines. Pharmacology & therapeutics*, 249, 108485. <https://doi.org/10.1016/j.pharmthera.2023.108485>
- [5] Hajdú, B., Kapuy, O., & Nagy, T. (2024). *Basal State Calibration of a Chemical Reaction Network Model for Autophagy. International journal of molecular sciences*, 25(20), 11316. <https://doi.org/10.3390/ijms252011316>
- [6] Yorimitsu, T., & Klionsky, D. J. (2005). *Autophagy: molecular machinery for self-eating. Cell death and differentiation*, 12 Suppl 2(Suppl 2), 1542–1552. <https://doi.org/10.1038/sj.cdd.4401765>
- [7] Meléndez, A., & Neufeld, T. P. (2008). *The cell biology of autophagy in metazoans: a developing story. Development (Cambridge, England)*, 135(14), 2347–2360. <https://doi.org/10.1242/dev.016105>
- [8] Ryter, S. W., Cloonan, S. M., & Choi, A. M. (2013). *Autophagy: a critical regulator of cellular metabolism and homeostasis. Molecules and cells*, 36(1), 7–16. <https://doi.org/10.1007/s10059-013-0140-8>
- [9] Kaur, J., & Debnath, J. (2015). *Autophagy at the crossroads of catabolism and anabolism. Nature reviews. Molecular cell biology*, 16(8), 461–472. <https://doi.org/10.1038/nrm4024>
- [10] Zhu, Y., Liu, F., Jian, F., & Rong, Y. (2024). *Recent progresses in the late stages of autophagy. Cell insight*, 3(2), 100152. <https://doi.org/10.1016/j.cellin.2024.100152>
- [11] Yang, X., Yu, D. D., Yan, F., Jing, Y. Y., Han, Z. P., Sun, K., Liang, L., Hou, J., & Wei, L. X. (2015). *The role of autophagy induced by tumor microenvironment in different cells and stages of cancer. Cell & bioscience*, 5, 14. <https://doi.org/10.1186/s13578-015-0005-2>
- [12] Mizushima, N., & Komatsu, M. (2011). *Autophagy: renovation of cells and tissues. Cell*, 147(4), 728–741. <https://doi.org/10.1016/j.cell.2011.10.026>
- [13] Medina, D. L., Di Paola, S., Peluso, I., Armani, A., De Stefani, D., Venditti, R., Montefusco, S., Scotto-Rosato, A., Prezioso, C., Forrester, A., Settembre, C., Wang, W., Gao, Q., Xu, H., Sandri, M., Rizzuto, R., De Matteis, M. A., & Ballabio, A. (2015). *Lysosomal calcium signalling regulates autophagy through calcineurin and TFEB. Nature cell biology*, 17(3), 288–299. <https://doi.org/10.1038/ncb3114>
- [14] Mariño, G., Niso-Santano, M., Baehrecke, E. H., & Kroemer, G. (2014). *Self-consumption: the interplay of autophagy and apoptosis. Nature reviews. Molecular cell biology*, 15(2), 81–94. <https://doi.org/10.1038/nrm3735>
- [15] Obeng E. (2021). *Apoptosis (programmed cell death) and its signals - A review. Brazilian journal of biology = Revista brasileira de biologia*, 81(4), 1133–1143. <https://doi.org/10.1590/1519-6984.228437>
- [16] Gump, J. M., & Thorburn, A. (2011). *Autophagy and apoptosis: what is the connection?. Trends in cell biology*, 21(7), 387–392. <https://doi.org/10.1016/j.tcb.2011.03.007>
- [17] Liu, B., Oltvai, Z. N., Bayır, H., Silverman, G. A., Pak, S. C., Perlmutter, D. H., & Bahar, I. (2017). *Quantitative assessment of cell fate decision between autophagy and apoptosis. Scientific reports*, 7(1), 17605. <https://doi.org/10.1038/s41598-017-18001-w>

- [18] Moreau, K., Luo, S., & Rubinsztein, D. C. (2010). Cytoprotective roles for autophagy. *Current opinion in cell biology*, 22(2), 206–211. <https://doi.org/10.1016/j.ceb.2009.12.002>
- [19] Zhang, L., Zhu, Y., Zhang, J., Zhang, L., & Chen, L. (2022). Inhibiting Cytoprotective Autophagy in Cancer Therapy: An Update on Pharmacological Small-Molecule Compounds. *Frontiers in pharmacology*, 13, 966012. <https://doi.org/10.3389/fphar.2022.966012>
- [20] Tavassoly, I., Parmar, J., Shajahan-Haq, A. N., Clarke, R., Baumann, W. T., & Tyson, J. J. (2015). Dynamic Modeling of the Interaction Between Autophagy and Apoptosis in Mammalian Cells. *CPT: pharmacometrics & systems pharmacology*, 4(4), 263–272. <https://doi.org/10.1002/psp4.29>
- [21] Goodwin, D.G.; Speth, R.L.; Moffat, H.K.; Weber, B.W. *Cantera: An Object-Oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes. Version 2.5.1.* [Hyperlink](#)
- [22] Papp, M.; Varga, T.; Busai, Á.; Zsély, I.G.; Nagy, T.; Turányi, T. *Optima++ v2. 5: A General C++ Framework for Performing Combustion Simulations and Mechanism Optimization.* [Hyperlink](#)
- [23] Goitom, S. K., Papp, M., Kovács, M., Nagy, T., Zsély, I. Gy., Turányi, T., & Pál, L. (2022). Efficient numerical methods for the optimisation of large kinetic reaction mechanisms. *Combustion Theory and Modelling*, 26(6), 1071–1097. <https://doi.org/10.1080/13647830.2022.2110945>
- [24] Kim, J., Guan, K.L. (2019). mTOR as a central hub of nutrient signalling and cell growth. *Nat Cell Biol* 21, 63–71 <https://doi.org/10.1038/s41556-018-0205-1>
- [25] Laplante, M., & Sabatini, D. M. (2012). mTOR signaling in growth control and disease. *Cell*, 149(2), 274–293. <https://doi.org/10.1016/j.cell.2012.03.017>
- [26] Fingar, D. C., Richardson, C. J., Tee, A. R., Cheatham, L., Tsou, C., & Blenis, J. (2004). mTOR controls cell cycle progression through its cell growth effectors S6K1 and 4E-BP1/eukaryotic translation initiation factor 4E. *Molecular and cellular biology*, 24(1), 200–216. <https://doi.org/10.1128/MCB.24.1.200-216.2004>
- [27] Zou, Z., Tao, T., Li, H. (2020). mTOR signaling pathway and mTOR inhibitors in cancer: progress and challenges. *Cell Biosci* 10, 31 <https://doi.org/10.1186/s13578-020-00396-1>
- [28] Holczer, M., Hajdú, B., Lőrincz, T., Szarka, A., Bánhegyi, G., & Kapuy, O. (2020). Fine-tuning of AMPK-ULK1-mTORC1 regulatory triangle is crucial for autophagy oscillation. *Scientific reports*, 10(1), 17803. <https://doi.org/10.1038/s41598-020-75030-8>
- [29] Christopher Wanjek (2011). *Systems Biology as Defined by NIH An Intellectual Resource for Integrative Biology*. National Institutes of Health • Office of the Director | Volume 19 Issue 6 • November–December 2011 <https://doi.org/10.1128/MCB.06159-11>
- [30] M. Herdin, N. Czink, H. Ozcelik & E. Bonek. (2005). Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels. 2005 IEEE 61st Vehicular Technology Conference, Stockholm, Sweden, pp. 136-140 Vol. 1. <https://doi.org/10.1109/VETECS.2005.1543265>
- [31] Yousefi, S., Perozzo, R., Schmid, I., Ziemiecki, A., Schaffner, T., Scapozza, L., Brunner, T., & Simon, H. U. (2006). Calpain-mediated cleavage of Atg5 switches autophagy to apoptosis. *Nature cell biology*, 8(10), 1124–1132. <https://doi.org/10.1038/ncb1482>
- [32] Enomoto, A., & Fukasawa, T. (2022). The role of calcium-calpain pathway in hyperthermia. *Frontiers in molecular medicine*, 2, 1005258. <https://doi.org/10.3389/fmmed.2022.1005258>
- [33] Pidoux, G., & Taskén, K. (2010). Specificity and spatial dynamics of protein kinase A signaling organized by A-kinase-anchoring proteins. *Journal of molecular endocrinology*, 44(5), 271–284. <https://doi.org/10.1677/JME-10-0010>
- [34] Tokumitsu, H., & Sakagami, H. (2022). Molecular Mechanisms Underlying Ca²⁺/Calmodulin-Dependent Protein Kinase Kinase Signal Transduction. *International journal of molecular sciences*, 23(19), 11025. <https://doi.org/10.3390/ijms231911025>
- [35] Rangkuti, I., Hasibuan, P., Widyawati, T., & Siregar, Y. (2019). Metformin Inhibits Growth of Breast Cancer Cell T47D through Decreasing Expression of Protein P53, BCL2 and Cyclin D1. *Journal of Medical Research and Innovation*, 3(2), e000164. <https://doi.org/10.32892/jmri.164>
- [36] He, C., & Klionsky, D. J. (2009). Regulation mechanisms and signaling pathways of autophagy. *Annual review of genetics*, 43, 67–93. <https://doi.org/10.1146/annurev-genet-102808-114910>
- [37] HPuri, C., Vicinanza, M., & Rubinsztein, D. C. (2018). Phagophores evolve from recycling endosomes. *Autophagy*, 14(8), 1475–1477. <https://doi.org/10.1080/15548627.2018.1482148>
- [38] Wang, H., Guo, M., & Wei, H. (2023). Targeting p53 pathways: mechanisms, structures and advances in therapy. *Sig Transduct Target Ther* 8, 92 <https://doi.org/10.1038/s41392-023-01347-1>
- [39] Chevallier, S., Kalunga, E.K., Barthélemy, Q. (2021). Review of Riemannian Distances and Divergences, Applied to SSVEP-based BCI. *Neuroinform* 19, 93–106 <https://doi.org/10.1007/s12021-020-09473-9>
- [40] Fletcher, P.T., Joshi, S. (2004). Principal Geodesic Analysis on Symmetric Spaces: Statistics of Diffusion Tensors. In: Sonka, M., Kakadiaris, I.A., Kybic, J. (eds) *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis. MMBIA CVAMIA 2004* 2004. Lecture Notes in Computer Science, vol 3117. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-27816-0_8
- [41] Zhang, Y., Pan, J., Li, K., Liu, W., Chen, Z., Liu, X., & Wang, J. (2023). On the properties of kullback-leibler divergence between multivariate Gaussian distributions. In *Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS '23)*. Curran Associates Inc., Red Hook, NY, USA, Article 2535, 58152–58165. [arXiv](#)
- [42] Ledoit, O., Wolf, M., (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*. Volume 88, Issue 2. Pages 365-411. ISSN 0047-259X. [https://doi.org/10.1016/S0047-259X\(03\)00096-4](https://doi.org/10.1016/S0047-259X(03)00096-4)
- [43] Ledoit, O., Wolf, M., (2022). The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation, *Journal of Financial Econometrics*, Volume 20, Issue 1, Winter 2022, Pages 187–218 <https://doi.org/10.1093/jfinec/nbaa007>
- [44] Park, J.M., Lee, D.H. & Kim, D.H. (2023). Redefining the role of AMPK in autophagy and the energy stress response. *Nat Commun* 14, 2994 <https://doi.org/10.1038/s41467-023-38401-z>
- [45] Uddin, M. S., Stachowiak, A., Mamun, A. A., Tzvetkov, N. T., Takeda, S., Atanasov, A. G., Bergantin, L. B., Abdel-Daim, M. M., & Stankiewicz, A. M. (2018). Autophagy and Alzheimer's Disease: From Molecular Mechanisms to Therapeutic Implications. *Frontiers in aging neuroscience*, 10, 04. <https://doi.org/10.3389/fnagi.2018.00004>

- [46] Barbosa, M. C., Grosso, R. A., & Fader, C. M. (2019). *Hallmarks of Aging: An Autophagic Perspective*. *Frontiers in endocrinology*, 9, 790. <https://doi.org/10.3389/fendo.2018.00790>
- [47] López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2023). *Hallmarks of aging: An expanding universe*. *Cell*, 186(2), 243–278. <https://doi.org/10.1016/j.cell.2022.11.001>
- [48] Salminen, A., & Kaarniranta, K. (2012). *AMP-activated protein kinase (AMPK) controls the aging process via an integrated signaling network*. *Ageing research reviews*, 11(2), 230–241. <https://doi.org/10.1016/j.arr.2011.12.005>
- [49] Mannick, J.B., Lamming, D.W. (2023). *Targeting the biology of aging with mTOR inhibitors*. *Nat Aging* 3, 642–660 <https://doi.org/10.1038/s43587-023-00416-y>
- [50] Metzger, S., Saukko, M., Van Che, H., Tong, L., Puder, Y., Riess, O., & Nguyen, H. P. (2010). *Age at onset in Huntington's disease is modified by the autophagy pathway: implication of the V471A polymorphism in Atg7*. *Human genetics*, 128(4), 453–459. <https://doi.org/10.1007/s00439-010-0873-9>
- [51] Khandia, R., Dadar, M., Munjal, A., Dhama, K., Karthik, K., Tiwari, R., Yattoo, M. I., Iqbal, H. M. N., Singh, K. P., Joshi, S. K., & Chaicumpa, W. (2019). *A Comprehensive Review of Autophagy and Its Various Roles in Infectious, Non-Infectious, and Lifestyle Diseases: Current Knowledge and Prospects for Disease Prevention, Novel Drug Design, and Therapy*. *Cells*, 8(7), 674. <https://doi.org/10.3390/cells8070674>
- [52] Aman, Y., Schmauck-Medina, T., & Hansen, M. (2021). *Autophagy in healthy aging and disease*. *Nat Aging* 1, 634–650 <https://doi.org/10.1038/s43587-021-00098-4>