Research Statement

Rahul, PhD, BlackRock, Gurgaon, India

Contents

Introduction

I have worked in the area interdisciplinary research where I have worked on the mathematics and its applications to solving research problems[7, 4]. I developed core expertise in the high-performance scientific computing, matrix computations, optimization, inverse problems, and high precision numerical algorithm. I rigorously pursued the topics to perfect thinking and abstract key ideas to solve problems. The focus on the process of solving the problem allowed me to look at the problem with fresh perspective, recognize patterns, and connect ideas across discipline to solve problems[20]. Furthermore, I ensured to drill deep into problem statements and develop deep thinking about the research objective. I have the opportunity to solve complex industry problems in the areas of finance and catastrophe risk models. In my post-doctoral and doctoral work, I worked in the field of systems biology. Next, I believe in the synergy between teaching and research at the university. I had the opportunity to teach courses at university, where I applied various modern methods to promote active learning in students. The figure 1 provides the overview of a research organization.

First, I will introduce my doctoral and post-doctoral research work. Second, I will explain an ongoing research to understand the signaling of insulin secretion in the type 2 diabetes. Finally, I will describe the proposal for teaching and training at the University.

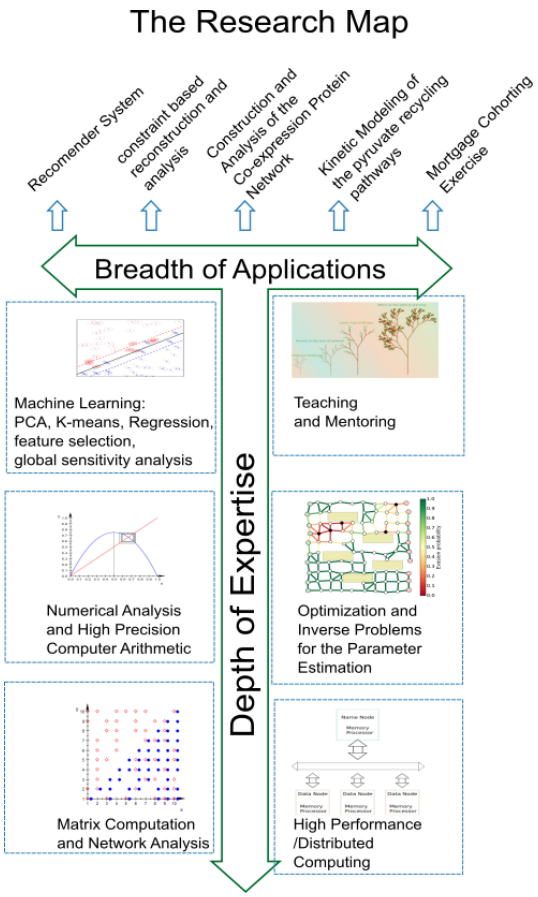

Figure 1: The figure provides the overview of research outline.

Previous Research Projects

Post-Doctoral Research Work

In postdoctoral work, I developed the rigorous procedure for data aggregation and built a model based on the constraint-based reconstruction and analysis (COBRA) method. The figure 2 provides the overview of the steps involved in the construction of a model through COBRA method.

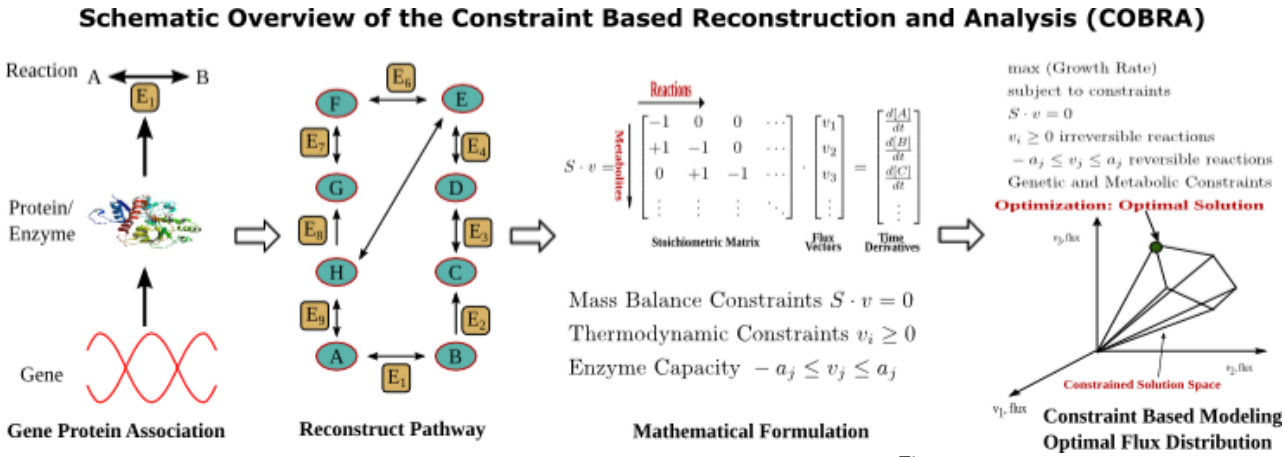

Figure 2: The figure provides the outline of steps involved in building and analyzing Genome-Scale Metabolic Models using COBRA method. The first step is to identify the gene-protein-reaction association. Second, is to construct the metabolic pathway based on the available knowledge base. Third, translate the model into the mathematical framework and add constraints to the flux. Finally, a growth objective is defined and optimization is performed to find the optimal flux distribution for the desired growth objective.

There is growing evidence that the cancer cells metabolism is reprogrammed in many different ways compared to the healthy cell to meet the metabolic needs of cancerous cells. The constraint-based reconstruction and analysis (COBRA) methods integrate the biochemical, genetic, and metabolic knowledge into a mathematical framework that enables the systematic study of the metabolic phenotype of the cells[9].

The crucial step in studying the cancer metabolism through COBRA methods is to obtain a generic GSM model, which can be fine-tuned to integrate gene expression and metabolomics data pertinent to cancer cells[18].

We used modelBorgifier to integrate the three generic mouse Genome-scale model (GSM) into one aggregated mouse GSM model[13]. Next, we computationally integrate the cancer cell-specific gene expression and metabolomics data into the model through iMAT algorithm[22].

After, constraining the GSM model to tissue-specific expression and metabolomics data, we performed computer simulations to compare simulated model results with the NCI 60 cell lines such that the model reproduces the common metabolic dysregulation found across the cancer cell lines. Next, we performed the flux variability analysis, which showed interesting result about increased activity of Lactate Dehydrogenase enzyme. The computer simulation showed that the Pyruvate Dehydrogenase, which is the entry point for TCA cycle

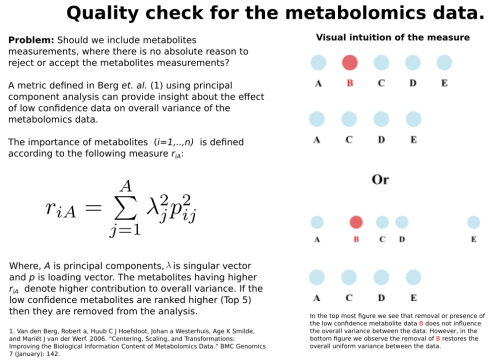

Figure 3: The figure provides an exemplar of data quality check used to maintain the high quality data.

The iMAT algorithm restricts the reaction fluxes according to the expression profile of genes, that is enzymes corresponding to high expressed genes will carry more flux compared to low expressed genes.

from glycolysis cannot carry all the flux from increased activity of glycolytic enzymes, which leads to flux redirecting towards pyruvate dehydrogenase.

Doctoral Research Work

In my doctoral research work, I developed a model for the pyruvate recycling metabolic pathways to identify key regulatory components in the pathway, which influence pyruvate recycling and NADPH[10]. Both, pyruvate recycling and NADPH has been shown to play a critical role in the insulin secretion. The malfunction of the key metabolic pathways such as pyruvate recycling pathway is shown to be correlated with the onset of Type 2 diabetes. Therefore, a better understanding of this pathway can suggest better targets for performing experiments and therapeutic innervation.

The model, which I developed for the pyruvate recycling pathways, describes the TCA cycle, the pyruvate/malate shuttle, the pyruvate/citrate shuttle, and the pyruvate/isocitrate shuttle (Figure). The model consists of 24 states, 31 reaction fluxes, and 129 parameters. The majority of the parameters are pulled from literature, and a subset of 34 parameters was optimized to validate the model against the experimental data of Ronnebaum *et al.*[11, 14, 21]. After testing the model against various results related to properties of pyruvate recycling pathways, I analyzed the model using global sensitivity analysis methods of Partial Rank Correlation Coefficient and Extended Fourier Amplitude Sensitivity Test and local sensitivity analysis[12]. The objective of sensitivity analysis was to identify the important control points in the pyruvate recycling pathways. The model predicts that the dicarboxylate carrier (DIC) and pyruvate transporter (PYC) are the most important regulators of pyruvate recycling and NADPH production. Our analysis showed that variation in the pyruvate carboxylase (PC) flux was compensated for by a response in the activity of mitochondrial isocitrate dehydrogenase (ICDm) resulting in the minimal changes in overall pyruvate recycling flux. The model predictions suggest points for further experimental investigation, as well as potential drug targets for the treatment of type 2 diabetes.

The Model Components of the Pyruvate Recycling Pathways

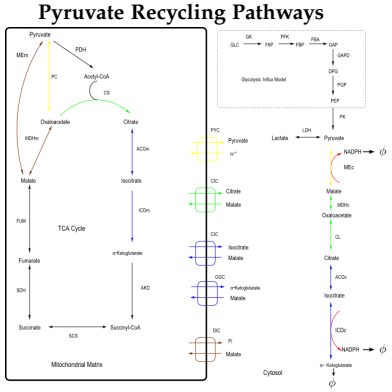

Figure 4: Pyruvate recycling pathways in β-cells: Brown color represents Pyruvate/malate shuttle. Green color highlights the Pyruvate/citrate shuttle is shown in green. Blue color describes the Pyruvate/isocitrate shuttle. Red reactions are NADPH-producing steps. Yellow color shows the Shared reactions. The reactions inside the box shows the influx glycolysis model.

Global vs. Local Sensitivity Analysis

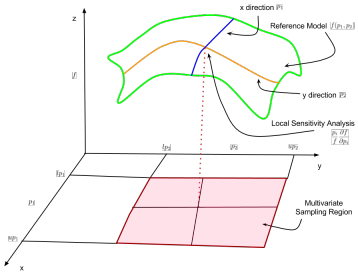

Figure 5: The global sensitivity analysis methods scan the entire parameter space in contrast local sensitivity analysis is restricted to individual perturbations.

Example: Global Sensitivity analysis Result

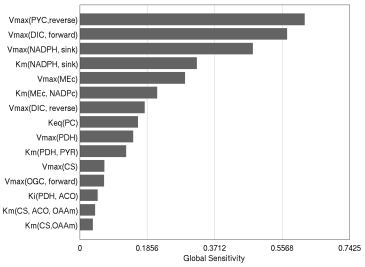

Figure 6: The figure describes the eFAST Total Effect Ranking of Parameters for NADPH. We observe the transport enzymes PYC and DIC are ranked higher. The sensitivity ranking is consistent with hypothesis we put together that transport enzymes PYC and DIC have critical role to play in the regulation pyruvate recycling ratio and NADPH.

Research Proposals

Spatial-Temporal Model

Next step in understanding the Insulin secretion is developing in the understanding of the signal integration through different pathways at the plasma membrane of the β-cells. Next, I describe my research plan to use a mathematical model to understand the signal integration at the plasma membrane of the β-cells.

The plasma membrane of β-cells is a hub for both signal integration and generation. The complex interplay of different time scales and spatial constraints plays a crucial role in integrating and regulating the cellular signals generated through the different biochemical pathways at the plasma membrane. Early mathematical models of β-cell signaling through the plasma membrane focused on the temporal dynamics of the interactions between the proteins and membrane channel proteins, whereas recent studies have revealed that spatial heterogeneity, such as clustering, plays a central role in regulating signals at the plasma membrane[5, 8].

Recently, the availability of the three-dimensional simulation software tools and new experimental data have enabled researchers to investigate membrane protein interaction mechanisms through realistic three-dimensional simulations. Therefore, I propose to develop a spatiotemporal model of the interaction between the Kv2.1 channel and the SNARE (soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor) complex[5]. Below, I provide an example procedure of model building and analysis by modeling Kv2.1 channel inhibition through the SUMO1 (small ubiquitin-like modifier) protein.

Research design and methods: The Monte-Carlo simulator for cell micro-physiology (MCELL) allows simulations to be performed in any complex cellular geometry[15, 16]. Furthermore, it has a rich class of volume and surface reactions, which can be used to model complex membrane interactions in greater detail. Here, I present a simulation of approximate reaction scheme based on the proposed model of McDonald in which Kv2.1 channel is inhibited by SUMO1 protein. The wiring diagram and reaction schemes are described in the Figures and respectively.

Preliminary Studies and Progress Report: The model was simulated for 2000 iterations with scaled parameters to save simulation time, but keeping the simulation duration close to the average lifetime of insulin secretion. The results are shown in Figure 2. The simulation can reproduce the effect of inhibition on the channel activity. The source codes used for the simulations are freely available at `https://github.com/r2rahul/KvChannel` (accessed August 30, 2016).

This example shows the stochastic effect on the inhibition of Kv2.1 through the SUMO1 protein. We observe a reduction in the number of active channels for the inhibited case at the end of simulation time course, although the initial time course looks similar to the non-inhibited case. Significantly, we observe a sharp shift in the number of active channels when SUMO molecules are clustered at the membrane. This model can be further extended by adding other molecular interactions to understand the principles of signal integration

The Model Components of Plasma Membrane Model

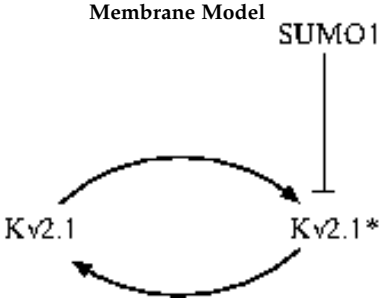

Figure 7: Figure describes the wiring diagram of the interactions. Kv2.1 is activated to Kv2.1*, which is inhibited by SUMO1.

The Reaction Diagram

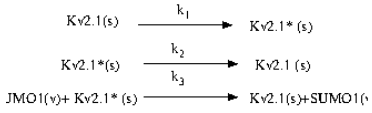

Figure 8: Figure shows the corresponding reaction schemes and rate constants for the MCELL simulation. Labels (v) and (s) denote membrane and volume molecules respectively. The diffusion constants are 10-6 cm^2 sec^{-1} and 10^{-8} cm^2 sec^{-1} for the cytoplasm and membrane respectively. Rate constants are k1=200 s^{-1}, k2=100 s^{-1}, and k3=1.73 M^{-1} s^{-1}, where parameters are pulled from the literature, based on the similarity of the biochemical processes[3, 17].

Simulation Result

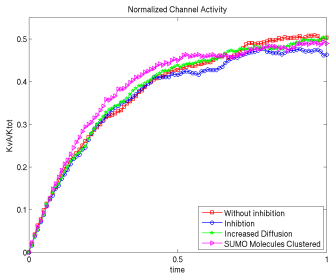

Figure 9: Figure shows the results of the simulations with and without inhibition by SUMO1.

at the plasma membrane of the β-cells. A well corroborated mathematical model will allow us to answer important questions related to signaling in a quantitative manner, for example, "how much does the time-scale diffusion matter at the plasma membrane?", "what is the approximate time-scale that is necessary for the signals to be processed by the channel proteins?", "what are the effects of rebinding and clustering?". The current simulation setup example can be further modified to investigate the mechanistic principles of signal integration on the β-cells plasma membrane.

The Architecture of Distributed Computing Center

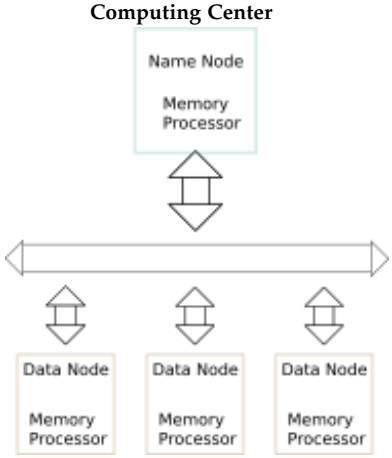

Figure 10: The figure illustrates the architecture of the distributed computing. The data node carries out the computation assigned through the program. The name node functions as an aggregator of the computation performed at the each data node.

High-Performance Computing

The three research projects outlined here in the research statement require high computational resources. Given the computational requirements, my research interest also extends in adapting the numerical methods of computation on the distributed computing environment[19, 1]. To adapt numerical methods the necessary condition is that for distributing computation the method must follow associative, distributive, and commutative rules (Figure 10). To reformulate the numerical algorithms to satisfy these conditions require careful reformulation and unit testing of the algorithms[2]. Building on my industry experience, I would like to develop numerical methods, which will harness the power of distributive computation.

University Teaching

During Ph.D., I developed a very keen interest into the pedagogy of the university teaching. To get trained into the field I completed the *certificate in the university teaching (CUT)* offered by Center of Teaching Excellence at the University of Waterloo. While teaching and designing tutorials as a teaching assistant I developed many interactive demonstrations and open-ended exploratory exercises. Also, I like to use technology in the classroom, and I had used document projectors, clickers, Wiki, and open-sankore learning aids in the classroom.

Clicker is an audience response tool used in the classroom to increase the active learning of the students. Sample Clicker Question

CUT project: Wiki as a Course Management Tool for the Systems Biology Courses. Research Project submitted for the CUT program[**?**]

Open-Sankore is open source interactive whiteboard tool.

I had a keen interest in designing and developing demonstrations for the lecture[6]. The purpose of the demonstration will be to promote active learning during the lectures and help illustrate complex ideas. Furthermore, I would like to further research on what are the best practices to create and integrate the demonstrations during the lecture because the timing of the demonstration in the lecture is also going to play to retain student interest.

Next, I would like to develop a wiki-based collaborative environment for the development of exemplary question banks for the lecture. The question words and pattern play a critical role in the learning of students. However, developing exemplary problems for the lectures is a non-trivial task, which requires experience and col-

Example Teaching Demonstration

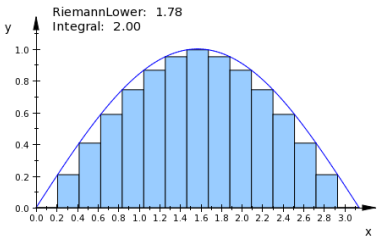

Figure 11: The illustration of Riemann Sum. The complete animation can be accessed through URL `http://tinyurl.com/pfxg8lp` (Accessed 05 May 2015)

laborative effort. Therefore, I propose to develop a Wiki based peer reviewed question bank, which will serve as an example questions to develop student's critical thinking. The design pattern of questions will focus on the process of learning, persistence, and space for a mistake for deep learning of the concepts.

Conclusion

In my research plan, I described my post-doctoral and doctoral research work. Next, I outlined my research design for understanding the spatiotemporal model of the plasma membrane of β-cells, where signals from different pathways are integrated to carry out regulated insulin secretion. Finally, I provided a research plan for contribution to university teaching. Since mathematical models are a simplified representation of physical processes, so by nature, it has assumptions and simplification (12). So, quite often models need to be refined to find a reasonable answer to problems. Therefore, by carefully identifying the questions, formulating the model accurately, and analyzing the model thoughtfully, I will continue to advance my research goals.

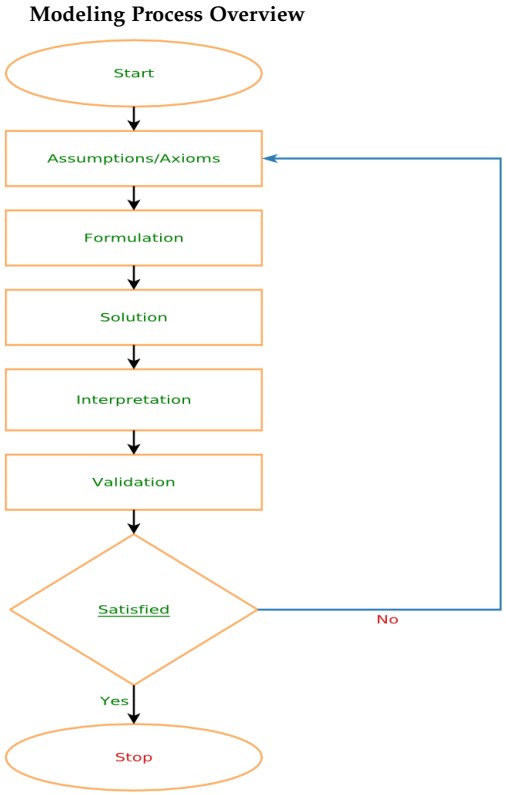

Figure 12: The figure provides the overview of the model building process. Figure adapted from National Council of Education Research and Technology (NCERT). `http://ncert.nic.in/` (accessed August 30, 2016).

References

[1] Volkan Cevher, Stephen Becker, and Mark Schmidt. Convex Optimization for Big Data: Scalable, randomized, and parallel algorithms for big data analytics. *IEEE Signal Processing Magazine*, 31(5):32–43, 9 2014.

[2] Jeremy Freeman, Nikita Vladimirov, Takashi Kawashima, Yu Mu, Nicholas J Sofroniew, Davis V Bennett, Joshua Rosen, Chao-Tsung Yang, Loren L Looger, and Misha B Ahrens. Mapping brain activity at scale with cluster computing. *Nature Methods*, 11(9):941–950, 7 2014.

[3] James P. Keener and James Sneyd. *Mathematical Physiology*. Springer, 1 1998.

[4] Paul B. Lowry, Aaron Curtis, and Michelle Renè Lowry. Building a Taxonomy and Nomenclature of Collaborative Writing to Improve Interdisciplinary Research and Practice. *Journal of Business Communication*, 41(1):66–99, 1 2004.

[5] Patrick E MacDonald. Signal integration at the level of ion channel and exocytotic function in pancreatic β-cells. *Am J Physiol Endocrinol Metab*, 301(6):1065–9, 12 2011.

[6] E Mazur. The Problem with Problems. In *Optics and Photonics News*, volume 6, pages 59–60. 1996.

[7] N. Metzger. SCIENCE POLICY:Interdisciplinary Research: From Belief to Reality. *Science*, 283(5402):642–643, 1 1999.

[8] Andrew Mugler, Aimee Gotway Bailey, Koichi Takahashi, and Pieter Rein ten Wolde. Membrane clustering and the role of rebinding in biochemical signaling. *Biophysical journal*, 102(5):1069–78, 3 2012.

[9] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245–8, 3 2010.

[10] Rahul Rahul. *Kinetic Modeling of Pyruvate Recycling Pathways in the Pancreatic β-cells.* PhD thesis, University of Waterloo, Waterloo, 2008.

[11] Sarah M Ronnebaum, Mette V Jensen, Hans E Hohmeier, Shawn C Burgess, Yun-Ping Zhou, Su Qian, Douglas MacNeil, Andrew Howard, Nancy Thornberry, Olga Ilkayeva, Danhong Lu, a Dean Sherry, and Christopher B Newgard. Silencing of cytosolic or mitochondrial isoforms of malic enzyme has no effect on glucose-stimulated insulin secretion from rodent islets. *The Journal of biological chemistry*, 283(43):28909–17, 10 2008.

[12] Andrea Saltelli. *Global Sensitivity Analysis: The Primer*. John Wiley, 3 2008.

[13] John T Sauls and Joerg M Buescher. Assimilating genome-scale metabolic reconstructions with modelBorgifier. *Bioinformatics (Oxford, England)*, 30(7):1036–8, 4 2014.

[14] Maurice Scheer, Andreas Grote, Antje Chang, Ida Schomburg, Cornelia Munaretto, Michael Rother, Carola Söhngen, Michael Stelzer, Juliane Thiele, and Dietmar Schomburg. BRENDA, the enzyme information system in 2011. *Nucleic acids research*, 39(Database issue):D670–676, 1 2011.

[15] J R Stiles, D Van Helden, T M Jr Bartol, E E Salpeter, and M M Salpeter. Miniature endplate current rise times less than 100 microseconds from improved dual recordings can be modeled with passive acetylcholine diffusion from a synaptic vesicle. *Proc. Natl. Acad. Sci. U.S.A.*, 93(12):5747–5752, 6 1996.

[16] Joel Stiles and Thomas Bartol. Monte Carlo Methods for Simulating Realistic Synaptic Microphysiology Using MCell. In Erik Schutter, editor, *Computational neuroscience: realistic modeling for experimentalists*. CRC Press, 2001.

[17] Koichi Takahashi, Sorin Tănase-nicola, and Pieter Rein. Spatiotemporal correlations can drastically change the response of a MAPK pathway. *Proc. Natl. Acad. Sci. U.S.A.*, 107(6):2473–2478, 2010.

[18] Ines Thiele and Bernhard Ø Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, 5(1):93–121, 1 2010.

[19] Sotirios A. Tsaftaris. A Scientist's Guide to Cloud Computing. *Computing in Science & Engineering*, 16(1):70–76, 1 2014.

[20] R L Wilder. The role of intuition. *Science (New York, N.Y.)*, 156(3775):605–10, 5 1967.

[21] Ulrike Wittig, Renate Kania, Martin Golebiewski, Maja Rey, Lei Shi, Lenneke Jong, Enkhjargal Algaa, Andreas Weidemann, Heidrun Sauer-Danzwith, Saqib Mir, Olga Krebs, Meik Bittkowski, Elina Wetsch, Isabel Rojas, and Wolfgang Muller. SABIO-RK database for biochemical reaction kinetics. *Nucleic Acids Research*, 40(D1):D790–D796, 2012.

[22] Hadas Zur, Eytan Ruppin, and Tomer Shlomi. iMAT: an integrative metabolic analysis tool. *Bioinformatics (Oxford, England)*, 26(24):3140–2, 12 2010.