IEEE Open Journal of the Systems Engineering

Received 13 April 2024; revised 11 July 2024; accepted 16 July 2024. Date of publication 22 July 2024; date of current version 27 August 2024.

*Digital Object Identifier 10.1109/OJSE.2024.3431868*

# A Model-Based Reverse System Engineering Methodology for Analyzing Complex Biological Systems With a Case Study in Glycolysis

**GERALD L. FUDGE [1] (Senior Member, IEEE) AND EMILY BROWN REEVES [2]**

[1]Department of Engineering and Technology, Texas A&M University-Commerce, Commerce, TX 75428 USA
[2]Nutriplexity, Bremerton, WA 98312 USA

CORRESPONDING AUTHOR: GERALD L. FUDGE (e-mail: gerald.fudge@tamuc.edu.)

(Gerald L. Fudge and Emily Brown Reeves are co-first authors and contributed equally to this work.)

**ABSTRACT** We propose a model-based reverse systems engineering (MBRSE) methodology for biological systems that relies on requirements analysis in conjunction with model-based systems engineering (MBSE). The goal of this methodology is to better understand complex multiscale biological systems, discover knowledge gaps, and make testable predictions. The similarities between human-engineered and biological systems motivate this approach. Furthermore, traditional reductionist paradigms in biology have proven insufficient for understanding and accurately predicting complex biological systems, as opposed to systems engineering approaches that have proven effective in supporting the design and analysis of complex engineered systems spanning multiple spatiotemporal scales. We employ our MBRSE methodology to analyze glycolysis in a biological case study using object process methodology as the primary MBSE language for conceptual qualitative modeling, in conjunction with SysML use case modeling. Using the MBRSE methodology, we derive twenty-two requirements, uncover five gaps in knowledge, and generate six predictions for the core metabolic pathway of glycolysis. One significant prediction is that the Warburg effect associated with cancer is the result of a natural response to tissue injury that has become unstable due to a failure in the feedback mechanism of the tissue injury control system.

**INDEX TERMS** Biological systems, glycolysis, model-based systems engineering (MBSE), object process methodology (OPM), reverse systems engineering.

## I. INTRODUCTION

Systems engineering is a methodical, multidisciplinary, top-down approach to efficiently design and develop an engineered system meeting a set of requirements covering the desired functionality [1]. Reverse engineering, by contrast, is the process of analyzing an existing system in order to understand what the system does (i.e., functionality) and how it works [2]. Here, we focus on biological applications, which typically involve the reverse engineering of an already existing system (excepting synthetic biology and bioengineering).[1] Additionally, we are interested in reverse *systems* engineering, which can be thought of as applying systems engineering concepts, including requirements analysis, to the reverse engineering process [3]. Finally, we also use model-based systems engineering (MBSE), which emphasizes the role of modeling and MBSE languages and corresponding tools [4].

Although elements of systems engineering, including the use of MBSE, have been applied to the reverse engineering of biological systems [5], very little attention has been paid to requirements analysis. One notable exception is Somekh et al. [6], who analyze requirements while developing their mRNA transcription process model. Influenced by their work, as well as our own systems engineering background,

[1]Although our focus is on biological systems, our methodology applies to the reverse engineering of any "black-box" top-down engineered system in which we lack a comprehensive predefined set of requirements and design documents.

we propose to formally unite these elements in a reverse systems engineering methodology for biological systems that relies on requirements analysis in conjunction with MBSE. Our reverse systems engineering approach is motivated by the fact that biological systems exhibit features that are traditionally associated with good top-down requirements-driven system engineering practices, such as modularity, optimality, robustness, common protocols, and design reuse [7], [8]. Furthermore, reductionism, which asserts that biological systems can best be understood from the low-level components, has been the dominant approach in biology for decades [9], [10], [11] despite having been found inadequate with notably inaccurate predictions [11], [12]. Finally, top-down approaches to understand biological systems, even if not requirements-driven, have been successfully employed, illustrating the promise of top-down systems thinking in biology [13], [14], [15].

### A. REVERSE ENGINEERING

Although there are many disciplines that employ reverse engineering, we will discuss following application areas in order to highlight the differences as well as the similarities with biological reverse engineering.

1) *Software:* Software reverse engineering goals include upgrading legacy code, analyzing malware, decrypting file formats, recovering lost data, improving interoperability, identifying software vulnerabilities, and even stealing intellectual property [16]. This often involves model-driven reverse engineering (MDRE) [17], which is similar to MBSE, but applied to the reverse engineering of software. Furthermore, MDRE can also employ systems engineering requirements analysis, including the use of forward engineering to derive the requirements specifications after reverse engineering yields the desired models [18]. However, MDRE has a much stronger focus on modeling than our proposed methodology for biological systems.
2) *Legacy complex systems:* Another application area is recovery of the top-level requirements specification of an existing complex system given legacy design documentation in order to reengineer and improve the system as illustrated by Park et al. [19] and Han et al. [20], both of which employ reverse systems engineering with requirements analysis in railway systems.
3) *Biology:* Biology has long relied upon reverse engineering in order to understand the function and operation of observed biological systems and make useful predictions [7]. Reverse engineering in biology has similar goals as with human-engineered systems; however, there are some significant differences in goals and methodology: 1) Except in synthetic biology and bioengineering, the goal of biology typically focuses on understanding and predicting rather than preparing for a forward engineering phase. 2) The design documentation is never available a priori for biological systems. 3) Reductionism has long been the traditional approach in biology as noted earlier. Although some still argue for reductionist approaches [9], [10], biology is gradually moving away from reductionism toward a systems approach. Note that biological reductionism is analogous to bottom-up engineering, which, while still used, has been gradually giving way toward top-down systems engineering. 4) Surprisingly little has been written from a systems engineering perspective, including the role of requirements analysis, in systems biology.

### B. REQUIREMENTS ANALYSIS

While both forward and reverse engineering can benefit from requirements analysis, the role of requirements analysis is different for the following two contexts:

1) *Forward engineering:* In a forward engineering (i.e., design) context, the systems engineer starts with the high-level requirements, then begins deriving requirements in order to achieve a design that is compliant with the high-level requirements, while making requirement trades at all levels to balance competing goals. At the lowest level, the requirements specification details how the design is to be accomplished. Thus, high-level requirements focus on what to achieve, while low-level requirements focus on how to achieve it. Although forward systems engineering is often not 100% top-down in practice, the emphasis is nonetheless top-down, especially for large complex systems. However, we should note that middle-out engineering is also common, and whether top-down or middle-out, the process is often iterative.
2) *Reverse engineering:* In a reverse engineering context, the design is complete, but the requirements and design are typically poorly understood. Thus, the systems engineer starts with observations of the existing system behavior and infers the requirements and their connections to observed structure and function. Since the observations may be at any level of the system hierarchy, we may need to infer higher level requirements from lower level observations. As the reverse engineering progresses, we may identify gaps in the requirements hierarchy due to limited ability to observe all aspects of the system. We can then use a forward engineering process to derive predicted requirements (i.e., requirements flow-down) to fill in the gaps. We can then update the model based on the predicted requirements, and we can compare the revised model behavior against the real-world system to test the predicted requirements and validate the revised model. In summary, reverse engineering a system that exhibits top-down systems engineering characteristics starts at some arbitrary detail level and involves an iterative combination of top-down and bottom-up analysis cycles where knowledge is gradually accumulated and refined, resulting in a deeper understanding of the system. In the case of both Park et al. [19] and Han et al. [20], they place requirements analysis either at or near the end of the reverse

engineering process, while in our proposed methodology, requirements derivation and analysis starts much earlier. Note that the necessary level of detail of the requirements analysis will depend on the goals. In some cases, only partial requirements analysis may bear fruit in guiding research to better understand the biological system, while in other cases high fidelity modeling may entail more in-depth requirements analysis.

### C. MODEL-BASED SYSTEMS ENGINEERING

This section discusses why MBSE is useful when reverse engineering biological systems, and why we selected object process methodology (OPM), supplemented by SysML, to support this research.

Although the use of MBSE has grown rapidly in recent years, especially in the aerospace and defense industry, with a number of reported benefits [21], [22], some researchers have questioned whether the utility of MBSE has been conclusively demonstrated. As per Campo et al. [23], the top three reported benefits are improved communication, analysis capability, and system understanding. However, these benefits are partially offset by reported increased time, cost, and effort, which appear to derive from the complexity of the MBSE approach. While complexity (with associated schedule and labor costs) is a significant consideration based on the authors' experience, we argue that for biological applications, the benefits can outweigh the drawbacks.

Given the dispersed nature of biological research and the scientific enterprise across many sites globally, the MBSE benefit of improved communication (and associated information sharing) is even more important than in a typical industry setting. On the other hand, the problem with MBSE complexity in biological applications will likely be even more severe given the lack of systems engineering experience or training among biologists. This drawback can be mitigated by collaboration with systems engineers and by selecting the least complex MBSE tool that meets the modeling needs.

These factors suggest that the benefits of using MBSE in the reverse systems engineering of biological systems can outweigh the disadvantages if done properly. This conclusion is supported by recent successful cases, including the work by Dori and Choder [24] and by Somekh et al. [6] and [25] in discovering knowledge gaps, making predictions, and validating predictions within the mRNA lifecycle. In a more recent example, Johansen [26] used MBSE to model and better understand the process of chemotaxis that flagella-propelled bacteria use to search for nutrients.

Given the potential benefits of using MBSE, the next question is: which language and tool? Due to space constraints, we focus on some of the more prominent MBSE and MBSE-derivative languages used in biology, including SysML, Modelica, BioUML, system biology markup language (SBML), and OPM. Although Simulink, Matlab, and Python are all extensively used in biology, they are primarily used as computational tools rather than to support system modeling, although they all provide significant external support of system modeling languages, such as the Python–Modelica modeling framework [27]. Factors to consider in the use of a language include standardization, support, widespread use, ease of use, and capability. All of these MBSE languages are standardized to at least some extent and have support communities. Therefore, the primary factors we will consider here are widespread use, ease of use, and capability.

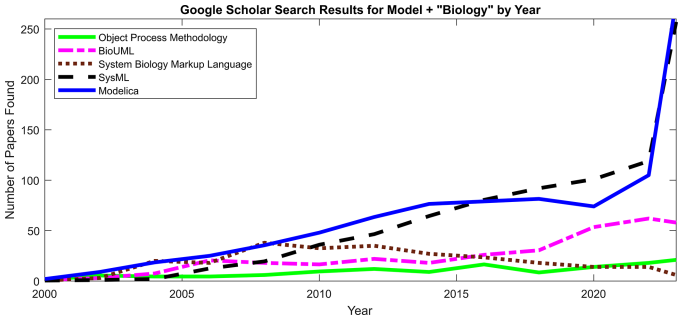


**FIGURE 1.** Trend in use of selected MBSE languages in biology, 2000–2022.

#### 1) WIDESPREAD USE

Fig. 1 illustrates the trends in tool use as measured by the number of search results from Google Scholar for the time period 2000–2022 (not counting citations) when searching on the model's name and biology (for example, one search term was <"object process methodology" + biology>). Although this method is not a definitive measure of how widespread the use of a tool is in biological applications, it does highlight the overall trends and relative popularity. As can be seen from Fig. 1, SysML, while it had a slow start, rapidly caught up with the other languages considered here, and now appears to be a leading MBSE tool in biological applications, along with Modelica. The third most popular is BioUML, which appears to be gradually gaining in use. SBML, on the other hand, was a contender with Modelica until about 2008, and since then has declined in use. Although OPM appears to have now passed SBML, it is still not widely used, especially compared to SysML and Modelica. Although limited use is a negative factor for OPM, there are some positive factors that we will discuss next.

#### 2) EASE OF USE

The ease of use (or relatively low complexity and learning curve) is a significant benefit of OPM. In particular, unlike SysML, OPM has a single diagram representation, which makes the model much easier to visualize, and which makes OPM easier to use since different diagram types do not need to be kept consistent, and since there is only one diagram type to learn. In addition, as noted by Dori and Choder, OPM is process oriented, making it suitable for biological applications, in contrast with languages based on the object-oriented (OO)

paradigm that makes processes subordinate to objects [24]. All of the languages considered in Fig. 1 are rooted in the OO paradigm except for OPM. Finally, as indicated by its name, OPM is not only an MBSE language, but is also an MBSE methodology that naturally supports top-down analysis and design.

#### 3) CAPABILITY

All of the modeling languages considered have qualitative and quantitative modeling capabilities, including OPM, and all have graphical representations available. In addition, SysML and OPM have explicit features to support requirements analysis. OPM features include the following:

1) A focus on processes, objects, and their associated relationships.
2) A single common diagram.
3) A text equivalent to the graphical representation.
4) The ability to support requirements analysis.
5) The ability to support qualitative conceptual modeling.
6) The ability to support quantitative modeling (including via embedded Python).
7) It is both a methodology and language as noted earlier.
8) The planned ability to export OPM models to SysML.[2]

The ease of use, combined with OPM having all of the necessary features, led us to focus on OPM even though it is not as widely used as SysML. However, OPM and SysML are not mutually exclusive [28], and we found it useful to create an SysML use case model of cellular metabolism with glycolysis because: 1) Use case models show how external entities (actors) interact with the system to achieve the desired objectives, aiding in the visualization and comprehension of key requirements and relationships. Thus, use case models offer a high-level systems engineering perspective. 2) Many systems engineers are more familiar with SysML than OPM, including use case modeling, and thus use case modeling is a good way to introduce glycolysis and OPM to systems engineers. 3) This example helps to show that our methodology can be implemented with different or mixed MBSE languages. Note that this is not an argument against OPM, although it might be useful to supplement OPM with other tools.

Having selected the language of OPM, we considered the two available software tools, OPCAT and OPCloud. OPCloud, as the newer tool with a cloud-based development environment, became our tool of choice [29], [30].

## II. MODEL-BASED REVERSE SYSTEMS ENGINEERING METHODOLOGY

### A. OVERVIEW

Fig. 2 shows an idealized flow diagram overview of our proposed model-based reverse systems engineering (MBRSE) approach for understanding biological systems. We start with system observations at any level in the system, from the high-level ecosystem to the lower level cellular or subcellular systems, such as glycolysis. Observation in biology often starts with mining the literature to understand the scope of present knowledge. However, whether doing a literature review or new experiments, useful systems questions in guiding the observational process include the following:

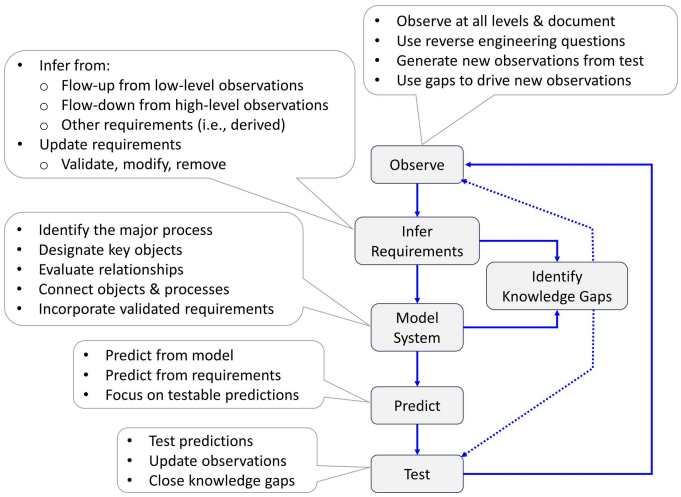


**FIGURE 2. Flow diagram for model-based reverse systems engineering of biological systems.**

1) Function: What does the system accomplish?
2) Context: In what context is the function accomplished?
3) Architecture: How is the system architected to perform its function, and what are its subsystems?
4) Performance: How well does the system perform its function?
5) Constraints: What are the system's constraints and external dependencies?
6) Interfaces: What are the system interfaces—external as well as internal subsystems?

Based on the observations, we infer corresponding requirements and begin modeling when we have sufficient information. We then make predictions, test, and update the observations based on observed test results. We also identify and work to close knowledge gaps. If necessary, we also update the requirements. Note that Fig. 2 represents an iterative process involving a combination of both top-down and bottom-up system modeling and requirements analysis.

#### 1) EXAMPLE MBRSE PROCESS

As an example of this process, we will consider passenger vehicles. If we compare a Toyota Highlander with a Toyota Corolla at the "organism" level, we can observe that the Highlander seats more passengers, has a significantly greater interior volume, has larger tires, and is reconfigurable to trade passengers against cargo volume. From these observations, we can infer that the suspension needs to be heavier duty to support greater weight as a flow-down requirement. We can develop mechanical models of the suspension and chassis system (after making additional observations) and compare

[2]As per email with Dr. Hanan Kohen, 4 July 2023; described here with permission of Dr. Kohen.

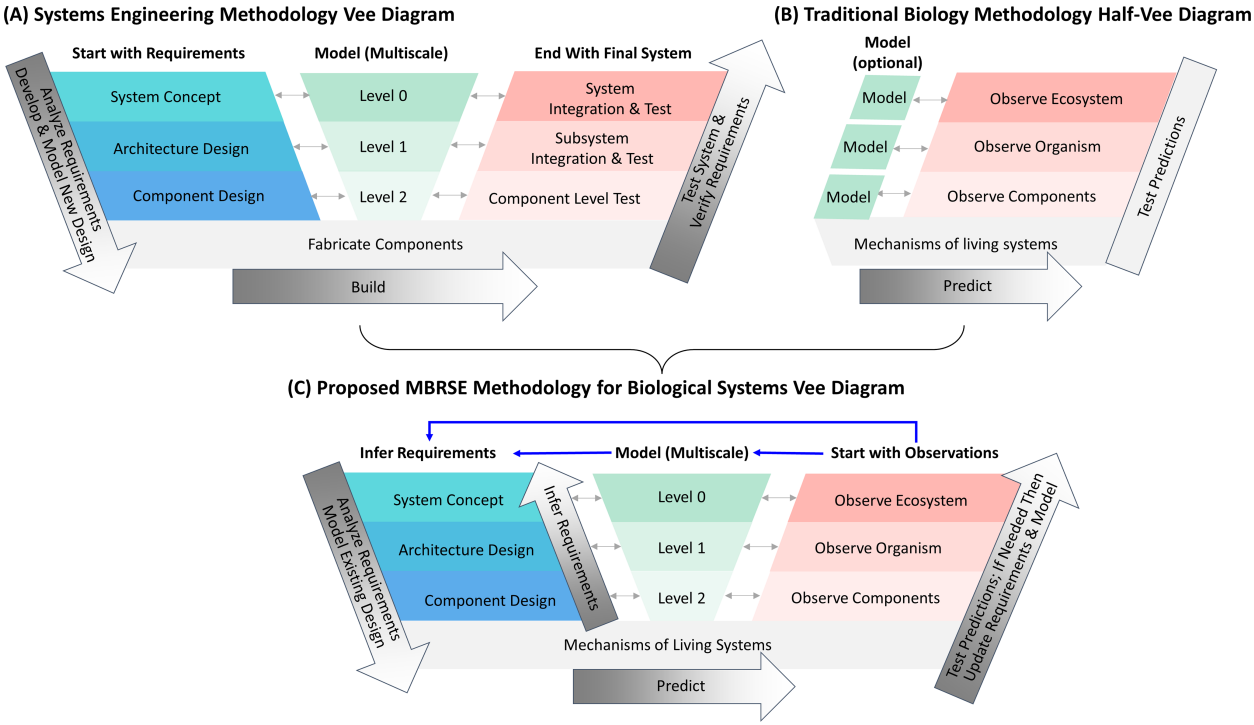


**FIGURE 3.** Vee diagram adapted for reverse systems engineering of biological systems.

the model behavior with vehicle behavior under test loading conditions to validate the suspension design specifications. As an example of requirements flow-up (and knowledge gap), we may also observe that both the Highlander and Corolla, along with other types of vehicles, have the same fuel interface, and hypothesize that there is an "ecosystem" requirement for a common fuel delivery system. If we then observe at the "ecosystem" level, we may find that there are two primary common systems at gas stations—gasoline and diesel—and we would then need to revise our "ecosystem" requirement to include both fuel systems. Furthermore, we may observe electric charging stations and natural gas stations. We can then infer that there are two primary types of vehicle "metabolism" systems—one based on gasoline and one based on diesel, in addition to a few other less common variants, such as all-electric and natural gas. We may also take sample measurements at a large used car dealership, where we can find a significant variety of different vehicle "species" from various geographic regions. We can then observe that the vehicle width for all "species" falls within a narrow range. From this, we can infer an "ecosystem" requirement on road width, from which we can predict that the typical road width in different regions of the country will be similar.

### *B. VEE DIAGRAM FOR MODEL-BASED REVERSE SYSTEMS ENGINEERING OF BIOLOGICAL SYSTEMS*

Panel C of Fig. 3 shows the proposed MBRSE methodology using a Vee diagram representation in comparison with a standard forward systems engineering Vee diagram showing the developmental life cycle of a multiscale system (panel A) and a half-Vee diagram representing the traditional reductionist biological method (panel B).

#### 1) SYSTEMS ENGINEERING METHODOLOGY VEE DIAGRAM (PANEL A)

The Vee diagram in panel A is adapted from Estefan [4] as a graphical representation of the developmental life cycle of a multiscale system. This systematic top-down process starts with requirements and relies on requirements flow down to ensure that the system concept, architecture design, and components are all fully compliant with the requirements. The process then transitions to system build and test, starting with the components, and integrates components and subsystems up to the top-level system. By testing at each level before further integration, problems are identified and corrected earlier in the process. Modeling supports the design phase at all levels as well as the testing phase. Levi-Soskin et al. [31] recommend that MBSE begins with qualitative conceptual modeling, followed by computational modeling, and then executable modeling in order to uncover errors at early stages in the process. In actual practice, the flow includes some amount of iteration, but the Vee shape captures the primary flow for the design, build, integrate, and test phases.

#### 2) TRADITIONAL BIOLOGY METHODOLOGY HALF-VEE DIAGRAM (PANEL B)

This panel expresses the traditional reductionist biology methodology as a half-Vee diagram. This approach starts with observations and relies on low-level experimentation, such as knock-out methods, to uncover how lower level objects affect

function, but does not connect them to system requirements [11]. Although systems biology, which focuses on the integrated system, is moving away from reductionist approaches [32], even recent systems biology approaches seldom involve requirements analysis. Furthermore, many biologists are still not sure how to effectively use systems principles, and tractable methodologies for them remain an important gap in the systems biology revolution.

#### 3) PROPOSED MBRSE METHODOLOGY FOR BIOLOGICAL SYSTEMS VEE DIAGRAM (PANEL C)

MBRSE, similar to the traditional biology methodology of panel B, starts with observations. However, unlike panel B, MBRSE moves quickly toward a systems perspective by inferring requirements and developing system models that tie together observations from different system levels. MBRSE, similar to the traditional systems engineering methodology in panel A, has a strong requirements and modeling focus, but unlike traditional systems engineering, MBRSE is focused on reverse engineering with a goal of developing testable predictions and understanding the observed system. Furthermore, in reverse engineering applications, including biology, the observational data may be at any level, so the MBRSE methodology needs to incorporate both top-down and bottom-up analysis.

This methodology is similar to how engineers can adapt a top-down systems engineering approach in reverse engineering situations in order to methodically elucidate the design. As with modern complex multiscale human engineered systems, top-level requirements constrain lower level subsystems in biological systems. Thus, linking mechanistic details to higher level requirements is important for a complete explanation of the architecture of systems. This methodology can identify knowledge gaps in current understanding and avoid incorrect predictions due to missed requirements.

### C. INFERRING REQUIREMENTS IN A REVERSE ENGINEERING CONTEXT

In many reverse engineering situations, including biological systems, the requirements are not given. Therefore, we must infer the requirements from observations. While this can involve a significant amount of work, we argue that requirements analysis is worth it because:

1) Systems engineering methodology is a proven approach for handling highly complex multiscale systems, and the first step is to understand the system requirements.
2) Requirements help us connect structure to function and better understand system behavior.
3) The process of inferring requirements can expose knowledge gaps at relatively early stages.
4) Understanding requirements can lead to more accurate predictions.
5) Often it is possible to make significant progress once a relatively small number of key requirements are understood. Thus, we do not necessarily need to derive a full set of detailed top-down systems requirements before starting modeling and making predictions, so that the time required for requirements analysis can be significantly reduced in many situations.

The requirements can be inferred directly from observations (whether from new or previously reported observations), or from other requirements via requirements flow-down, or from modeling results (which should be then compared with observational data to verify). For example, Somekh et al. [6] discuss how modeling can be used 1) to help understand and identify knowledge gaps in the temporal sequencing of processes, and then 2) to fill in the knowledge gaps by hypothesizing solutions, revising and running the model, and verifying with experimental results. In particular, they show how to use modeling to identify and infer unknown requirements on temporal dependencies in a multistep process within the mRNA life cycle. We will now illustrate our proposed MBRSE methodology with glycolysis as a case study at the conceptual qualitative modeling stage [31].

## III. RESULTS: MBRSE CASE STUDY WITH GLYCOLYSIS

We selected glycolysis as a case study to illustrate the MBRSE methodology because 1) glycolysis is a complex biological system that is reasonably well-understood after years of study, and 2) glycolysis still has open questions which a full MBRSE project could help address, and even a limited case study such as ours could make contributions.

We consider three different conceptual qualitative models: Fig. 4 in subsection A shows a use case model focusing on cellular level metabolism while including the ecosystem and organism context in order to help organize observations into key glycolysis requirements; subsection B presents a cellular level metabolism OPM model in order to provide the system context of glycolysis; subsection C presents an OPM model that focuses on glycolysis specifically. For each one of the models, we will walk through the process illustrated in Figs. 2 and 3. Although we have organized the discussion in a linear sequence of [Observations → Infer requirements → Model → Identify knowledge gaps → Predict → Test], in reality the process is iterative rather than linear: requirements are inferred from observations and modeling, modeling is based on observations as well as requirements, and knowledge gaps can be linked to modeling or requirements analysis, etc. Subsection D illustrates an OPM requirements example.

### A. USE CASE MODEL FOR CELLULAR LEVEL METABOLISM WITH ECOSYSTEM AND ORGANISM CONTEXT

#### 1) OBSERVE

Based on decades of observational data, glycolysis is a 10-step process that can operate under either aerobic or anaerobic conditions to convert glucose into pyruvate while producing energy in the form of ATP or while providing precursor molecules to support biosynthesis [33]. Scientists have also observed that virtually all organisms use glycolysis to meet their specific metabolic needs [34].

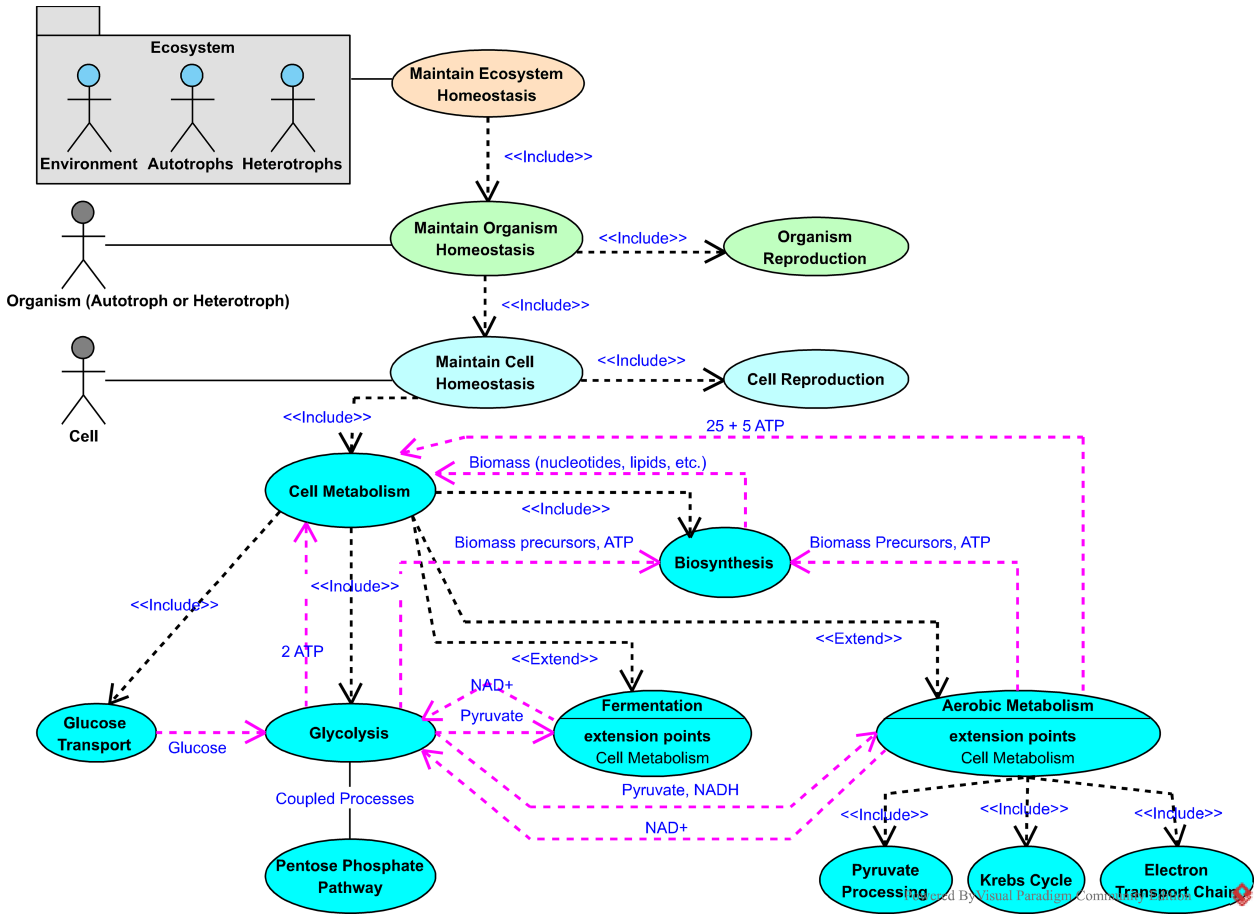


**FIGURE 4.** Cellular level metabolism use case model (simplified).

### 2) INFER REQUIREMENTS

Based on these observations, inferred glycolysis requirements include:

1) *Glycolysis shall provide energy within the cell in the form of ATP under either aerobic or anaerobic conditions*.
2) *Glycolysis shall provide precursor molecules to support biosynthesis of complex molecules if not immediately needed for energy*. These molecules can be provided either directly or indirectly via the pentose phosphate pathway and aerobic metabolism pathway.
3) *Glycolysis shall provide pyruvate and NADH to aerobic metabolism in aerobic conditions*. This requirement supports increased ATP delivery and regeneration of NAD+ to allow glycolysis to continue.
4) *Glycolysis shall provide pyruvate to fermentation metabolism under anaerobic conditions*. This requirement provides for the regeneration of NAD+ to allow glycolysis to continue through lactate synthesis in muscles or ethanol synthesis in yeast.

Although there are additional requirements that glycolysis must meet, these are sufficient to develop a use case system model.

### 3) MODEL SYSTEM

Use case models provide a high-level overview of system functionality, behavior, and relationships [35]. The cellular level metabolism use case shown in Fig. 4 is a hierarchical model starting at the ecosystem level to provide a top-level system context that helps connect glycolysis with higher level system requirements. As shown in Fig. 4, *glycolysis* is included within *cell metabolism*, which is included within *maintain cell homeostasis* to keep the cell alive and allow reproduction. *Maintain cell homeostasis* is included in *maintain organism homeostasis*, which in turn is included from the top-level system use case to *maintain ecosystem homeostasis*. The *ecosystem* actors consist of the *environment*, *autotrophs* (organisms that can get their energy and biomass without consuming other organisms), and *heterotrophs* (organisms that consume other organisms to get energy and biomass), all interacting together. These actors are interdependent since the oxygen in the *environment* required by *heterotrophs* is primarily produced by *autotrophs*, while *heterotrophs* consume other organisms and provide environmental carbon dioxide needed by *autotrophs*.

As shown in Figs. 4 and 5, *glycolysis* produces two ATP energy molecules for each glucose molecule in addition

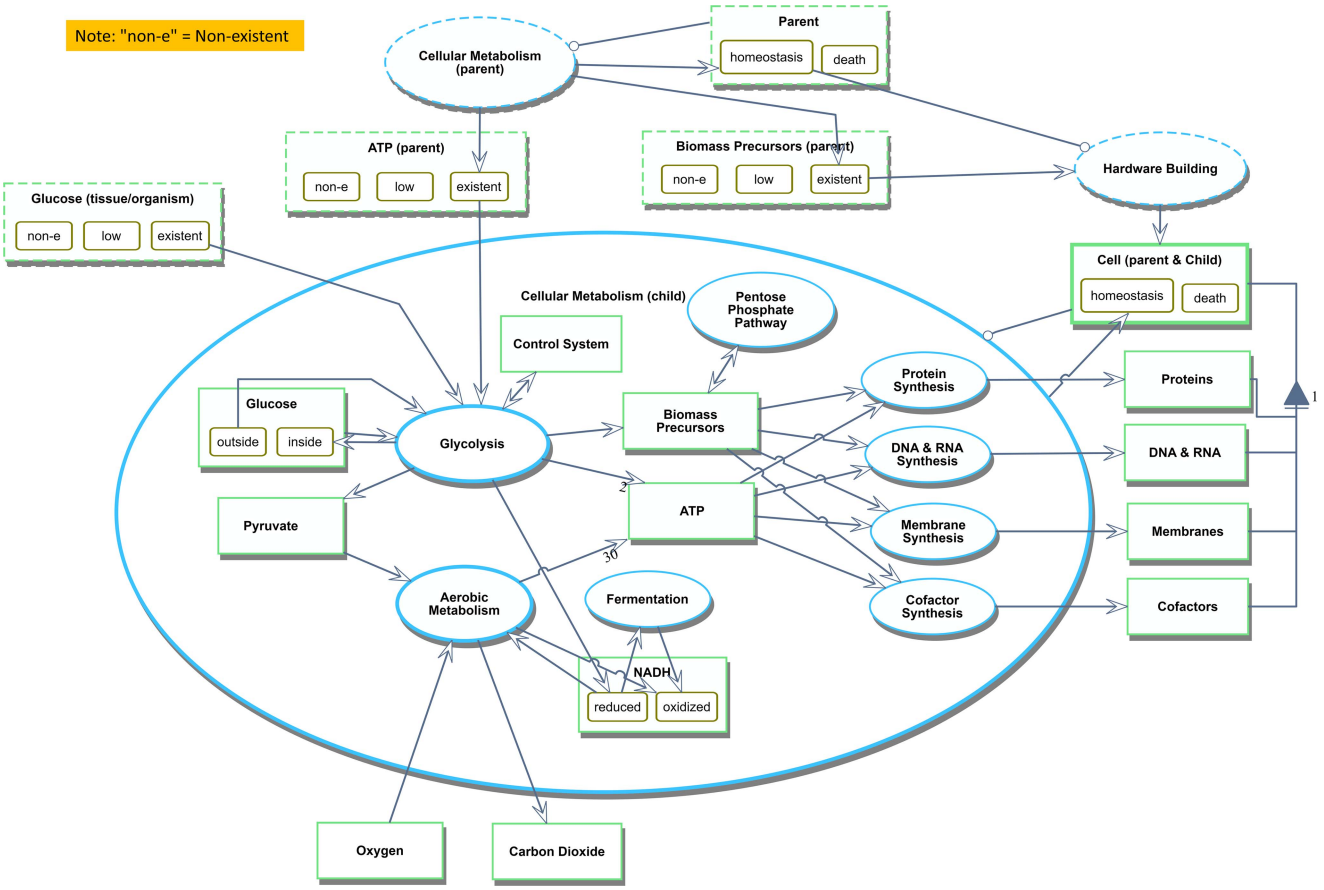


**FIGURE 5.** **Cellular level metabolism OPM model.**

to pyruvate to support *aerobic metabolism* or *fermentation*, while *aerobic metabolism* provides approximately 25 ATP directly plus 5 more from NADH provided by *glycolysis* under aerobic conditions. Note that *glycolysis* provides pyruvate to support either *fermentation* or *aerobic metabolism*, depending on oxygen availability, while these processes provide the NAD+ required by *glycolysis*. *Aerobic metabolism* includes *pyruvate processing* to convert pyruvate to acetyl CoA plus high energy molecules, the *Krebs cycle* to generate ATP plus high energy molecules from acetyl CoA, and the *electron transport chain* to generate significantly more ATP using the high energy molecules from *pyruvate processing*, *Krebs cycle*, and *glycolysis*. Both *glycolysis* and *aerobic metabolism* can provide biomass precursor molecules to support *biosynthesis* if the cell needs to focus on building complex molecules for various activities, including repair and maintenance (not shown) or reproduction at the cellular level or at the organism level in multicellular organisms.

In addition to *glycolysis*, *cell metabolism* includes *glucose transport* into the cell, *fermentation* to support continued *glycolysis* in anaerobic conditions, *aerobic metabolism* to provide increased energy production under aerobic conditions, and *biosynthesis* of various critical molecules. *Glycolysis* is coupled with the *pentose phosphate pathway* because a number of intermediate biomass precursor products in *glycolysis* can go to the *pentose phosphate pathway* to support *biosynthesis*, and vice-versa to support energy production.

### 4) IDENTIFY KNOWLEDGE GAPS

The near-uniformity of central metabolism across life has traditionally been attributed solely to universal common descent. However, from a systems engineering perspective, there are two requirement-based reasons why uniformity might be expected. First, we would expect some level of biomass commonality to maximize thermodynamic efficiency in reusing complex molecules by minimizing the amount of required biomolecule break-down and rebuild. Second, waste buildup can be reduced via commonly used recyclable waste products, and thus simplify maintenance of ecosystem homeostasis. Therefore, a significant gap in knowledge is: to what extent do the ecosystem requirements for compatibility and sustainability account for uniformity in the glycolytic topology? As can be seen from Fig. 4, a second significant gap is: How did the system shown in Fig. 4 originate in the first place? *Glycolysis* is included in *cell metabolism*, which in turn is included within *maintain cell homeostasis* in order to *maintain organism homeostasis*, but this assumes that organisms exist in a state of homeostasis,

which raises the question of how life got started in the first place. At the glycolysis level, there are two major origin models. The "genetics first" model argues that enzyme evolution gave rise to the current optimal glycolytic topology. The "metabolism first" model argues that a nonenzymatic prebiotic reaction sequence provided metabolites to early enzymes and acted as a template for Darwinian evolution to exploit. This is still an unresolved knowledge gap [36].

#### 5) PREDICT AND TEST

The requirements for any major subsystem of a complex system generally involve competing tradeoffs, and we would expect glycolysis to be no different. Thus, treating glycolysis as the result of a top-down systems engineering process, we would hypothesize that any apparent inefficiencies are really the result of competing trades. Also, glycolysis requires configuration variants to maintain efficiency given that diverse organisms have vastly different metabolic requirements. We can thus predict the following requirements, which can be confirmed from the research literature:

5) *Glycolysis shall support universal use across all three domains of living organisms*. As noted above, empirical evidence reveals a near uniformity of central metabolic pathways across a remarkable diversity of organisms (Archaea, Bacteria, and Eukarya). We hypothesize that this uniformity derives from constraints that flow from compatibility requirements between the actors of Fig. 4 that in turn flow from the requirement to maintain ecosystem homeostasis. Following this line of thought, exceptions to uniformity may best be thought of as meeting different interface requirements in the ecosystem rather than the mere result of genetic lineage.
6) *Glycolysis shall be efficient*. Although this prediction results naturally from a systems engineering view of glycolysis, this was not expected based on traditional reductionist approaches. As Bar-Even et al. [37] and Raven et al. [38] noted, biologists did not understand why so many steps were necessary and pondered why the pathway occurs at all in modern metabolism given that the energy production was so low compared to aerobic metabolism. As per Cornish-Bowden and Cárdenas [39], a representative view from papers given at a NATO Advanced Research Workshop in 1989 was that the evolutionary development of metabolism certainly "did not lead to a 'global optimum state,' but rather to a 'local optimum' instead." However, Ebenhöh and Heinrich [40] showed that the glycolysis architecture with a preparatory phase followed by a payoff phase is highly efficient based on kinetic and thermodynamic analysis. Similarly, Court et al. [41] discovered that the payoff phase has a maximally efficient throughput rate. In 2010, Noor et al. [42] demonstrated that the 10 step glycolytic pathway is minimally complex, with all glycolytic intermediates essential either for building biomass or for ATP production. In fact, it turns out that glycolysis is Pareto-optimized to maximize efficiency while serving multiple, often competing, purposes. Ng et al. [43] published their analysis in 2019 by analyzing over 10 000 possible routes between glucose and pyruvate to show that the two primary glycolysis variant pathways are Pareto-optimized to balance ATP production against biomass production while simultaneously minimizing protein synthesis cost.
7) *Glycolysis shall provide configuration options (i.e., pathway variants) to support universal use across all three domains of living organisms*. There are two major variants of the glycolytic pathway, the Emden–Meyerhof–Parnas (EMP) and the Entner–Doudoroff (ED) glycolytic pathway, which handle in different ways the tradeoff between energy production and biomass production. The ED variant produces one ATP, while the EMP pathway produces two ATP, but at a high cost for the cell because some of the EMP pathway reactions need to be very close to equilibrium to produce the extra ATP. Being near equilibrium means that a high concentration of enzymes is required for EMP glycolysis to move forward (EMP enzymes are some of the most common proteins in these cells, comprising 4%–6% of the *E. coli* proteome [44]). Because the extra protein synthesis for EMP is very costly, organisms that do not need the second glycolysis ATP, such as bacteria that can use aerobic metabolism to get extra ATP downstream of glycolysis, tend to use the ED pathway [44]. Similarly, marine bacteria in reduced oxygen environments preferentially use the ED pathway because they need more NADPH to support biosynthesis and oxidative stress reduction rather than ATP [45].

The utility of this method is seen in that our use case model and inferences of requirements allowed us to identify two knowledge gaps and make three predictions, with one of these predictions not expected on the basis of traditional reductionist approaches.

### B. OPM CELLULAR METABOLISM MODEL

#### 1) OBSERVE

The OPM cellular metabolism model is a conceptual model that corresponds approximately to *cell metabolism* and the related use case in Fig. 4, but with some additional features to reflect the well-known fact that cellular metabolism provides energy and biomass to support organismal homeostasis, which involves all the processes recognized as central to life, including information preservation, repair, reproduction, separation from the environment, and adaptability. Regarding adaptability, we can observe that since the energy production needs of an organism can suddenly increase (for example, when an animal flees from danger), there must be features in place to support the rapid increase in energy production at the cellular level. Or, after an injury, an organism may need to quickly ramp up local biomass production to support healing. In addition, the metabolic requirements of many organisms, such as

mammals, change as the organism matures. Thus, we anticipate that glycolysis will be required to support varying energy and biomass needs over longer durations of time to support optimal metabolism. This also implies that glycolysis needs tuning knobs to maintain efficiency as metabolic requirements vary.

### 2) INFER REQUIREMENTS

We can now organize these additional observations into lower level requirements to support an OPM conceptual architecture model of cellular metabolism as follows:

8) *Glycolysis shall provide an adaptable rate of energy production over time for any given tissue type.*
9) *Glycolysis shall provide an adaptable rate of biosynthesis production over time for any given tissue type.*
10) *Glycolysis shall be tunable to maintain optimized performance under varying conditions of energy and biosynthesis production.*
11) *Glycolysis shall be controllable to dynamically balance energy production versus supporting biosynthesis.*
12) *Glycolysis shall support cellular metabolism by providing nucleotides, RNA, DNA, fatty acids and triglycerides, ascorbic acid (plants and most animals except humans), steroids, and cholesterol or other structurally similar lipids*. This requirement expands on requirement #2 by specifying the biomass precursors.

### 3) MODEL SYSTEM

As discussed in the OPM introduction by Somekh et al. [25], OPM focuses on processes (ovals), objects (rectangles), and the relationships between them. In the conceptual OPM model shown in Fig. 5, the central process is *cellular metabolism (child)*. The object *cell (parent and child)* represents the physical architecture required by the *cellular metabolism* process (indicated by the open circle instrument link) to maintain the state of homeostasis. The *cell* in turn is produced by the *hardware building* process, which relies on *biomass precursors*. The *cell* parts (*proteins*, *DNA and RNA*, *membranes,* and *cofactors*) and their assemblies are necessary for subprocesses of the *cell*. *Hardware building* is controlled by the *reproduction and development* process (relationship not shown here) of the parent cell. The model includes environmental *oxygen* required by *cellular metabolism*, as well as *biomass precursors, glucose, and ATP*, which are assumed to be existent in sufficient quantities, but could be low, or even nonexistent ("non-e" in the diagram). There are also several circular dependencies in this system. From an energy perspective, the first step of *cellular metabolism* and *glycolysis* produces *ATP*, but first requires *ATP*. From a hardware perspective, *cellular metabolism* requires various complex molecules, such as *proteins* that are produced by *cellular metabolism*. These circular dependencies are resolved during the process of cell division, in which a new cell, rather than being produced from scratch, is built using the energy and hardware from an already existing cell. Of course, one must ultimately have a first cell to get this whole process started, but how this happened (i.e., abiogenesis) is still an unresolved knowledge gap. The *control system* for the *glycolysis* process reflects requirements #11 and related requirements #8, #9, and #10.

### 4) IDENTIFY KNOWLEDGE GAPS

In addition to the question of abiogenesis mentioned above, another question raised by the cellular level metabolism OPM model is: *How is the child cell weaned from the parent cell's provision of energy and biomass?* It appears to be the case for cells that the initial lineage homeostasis from boot-up of the first cell is never actually lost because reproduction maintains homeostasis of the lineage by externalizing it for child cells. So glycolysis remains active during mitosis allowing the parent cell's metabolism to provide biomass and energy to the child cell. But at what point is the child cell's DNA used for making glycolytic enzymes? What are the changes in metabolic flux that occur throughout mitosis? These are important knowledge gaps of particular interest in biology that can yield insight to the theory of self-replicating systems. Another knowledge gap is the function and purpose of the "Warburg effect," discussed below.

### 5) PREDICT AND TEST

Cancer cells often have an abnormal metabolic condition called the Warburg effect in which 1) the rate of glycolysis metabolism is much higher than normal, and 2) anaerobic fermentation is preferentially used over the aerobic electron chain transport, even in the presence of oxygen [46]. Thus, since glycolysis alone produces less energy per glucose molecule than when combined with the electron chain transport cycle, and since fermentation results in lactate, the Warburg effect is characterized by increased glucose uptake and increased production of lactate. The Warburg effect appears to be required to support cancerous tumor growth, however, Liberti and Locasale [46] remark that the function of the Warburg effect and how exactly it benefits cancerous growth is still not well understood despite thousands of publications in the research literature. As per Liberti and Locasale, the primary proposed functions under debate are as follows:

1) Rapid ATP production: Although glycolysis does not produce much ATP per glucose molecule, the rate of glycolysis can be accelerated to produce as much or more ATP than under normal aerobic metabolism, which may provide cells an advantage when competing for energy resources.
2) Increased biosynthesis: The Warburg effect may be an adaptive mechanism to provide increased biosynthesis as well as increased NADPH production (produced primarily by the pentose phosphate pathway shown in Fig. 5) in support of increased cell proliferation.
3) Tumor microenvironment: The Warburg effect may provide an advantage for cell growth and proliferation via a

localized environment effect, such as decreased pH due to excess lactate.

4) Cell signaling: The Warburg effect may improve tumor cell signaling by generating reactive oxygen species (ROS), since insufficient ROS can hinder cell signaling required for cell proliferation.

Based on the research literature, we predict the following additional glycolysis system requirements:

13) *Glycolysis shall be adaptable to support optimum response to local organism injury*. We hypothesize that the Warburg effect is a normal adaptive system response to local organism injury or other temporary situations that require rapid tissue growth, such as during certain early developmental stages. Under certain (currently unknown) conditions, the feedback control loop for injury response can be broken, resulting in an under-controlled or completely uncontrolled response. In other words, we hypothesize a cellular level failure in the control system that upregulates cellular processes for division including glycolysis such that the rate of glycolysis is unconstrained at the cellular level. Note that all four proposed functions of the Warburg effect, plus its ability to support cellular metabolism if the oxygen supply is interrupted due to local loss of normal blood flow, are beneficial for tissue repair after an injury where 1) there might be reduced oxygen, 2) faster cell division and local ATP energy supply is needed, and 3) more biomass is required. A similar situation can occur during early organism development when tissue growth is more rapid than in the adult stage, and in which the blood supply is developing simultaneously. As an initial check of our hypothesis about the Warburg effect, we conducted a literature search which revealed that few peer-reviewed papers focus on the Warburg effect in the context of injury or tissue repair. One notable exception is Heiden et al. [47], who suggest that the increased cellular multiplication rate associated with the Warburg effect can be beneficial in tissue repair as well as immune responses. Additional research is necessary to test our prediction that the Warburg effect is a natural response to the context of tissue repair and normal biological development during the growth phase for multicellular organisms. In particular, research that focuses on feedback mechanisms in the control system responsible for upregulating the rate of glycolysis should be able to verify or falsify our hypothesis.

14) *The glucose transport step shall be tunable to support time-varying and tissue-varying glycolysis needs*. Although not traditionally viewed as a dynamic tuning parameter for glycolysis, from a systems engineering perspective, we would expect the system input rate to be tunable, and it turns out that this is the case. As discussed in [48], there are a variety of glucose transporters (GLUTs), each with different properties to support the metabolic needs of different cell types. Furthermore, the expression level of these GLUTs can adapt over time in response to factors, such as available glucose supply, or developmental status from embryo through adult [49]. For example, vital tissues, such as red blood cells and the brain, primarily rely on GLUT1, a transporter characterized by a high affinity for glucose in order to prioritize the uptake of glucose even under fasting conditions. In contrast, control tissues, such as the pancreas and liver express primarily GLUT2, which is characterized by a high glucose saturation point to support sensing and regulating blood glucose levels. Absorptive tissues like the muscle express GLUT4, an insulin-sensitive GLUT characterized by a high glucose affinity and high maximum transfer rate under high glucose concentrations, but is sequestered from the membrane until the pancreas secretes insulin. Similar to how throttle valves convert high-pressure fluids into low-pressure fluids without changing enthalpy or performing mechanical work, GLUTs are an energy efficient mechanism to facilitate the movement of glucose across a concentration gradient. However, unlike a traditional throttle valve, the GLUT response is not adjustable over a continuous range of values because the adjustable parameters are the quantity and type of GLUTs in the cell membrane. This is tied to the constraint that GLUTs only transport glucose. Thus, it is not feasible for the transporter to widen its opening in a manner akin to a throttle valve, as this would result in the indiscriminate passage of various molecules without any form of selective filtration. Furthermore, in contrast to ordinary unidirectional throttle valves, GLUTs are bidirectional. Thus, GLUTs can be viewed as throttle valves that are reversible, gradient-sensitive, and molecule-specific. This analogy, which is one of the contributions of this article, helps explain why glucose transport is quantized, and may provide insight as to how the glucose transport regulation mechanism can fail due to extra GLUTs, leading to an out-of-control Warburg effect.

### C. OPM GLYCOLYSIS MODEL

#### 1) OBSERVE

We now consider additional observations that will help us to infer additional requirements and develop an OPM model of glycolysis. The 10-step glycolysis pathway (Fig. 6) can be split into two phases: a preparatory phase that consumes energy and a pay-off phase that releases net energy [33]. The preparatory phase consists of five reactions which perform molecular rearrangements for the production of the energy-rich molecules of NADH and ATP in the pay-off phase. During this preparatory phase of glycolysis, the glucose backbone is phosphorylated twice in two ATP consuming steps. These reactions (steps 1 and 3) give the molecule symmetry for being split in half by aldolase (step 4) so that the resulting trioses can

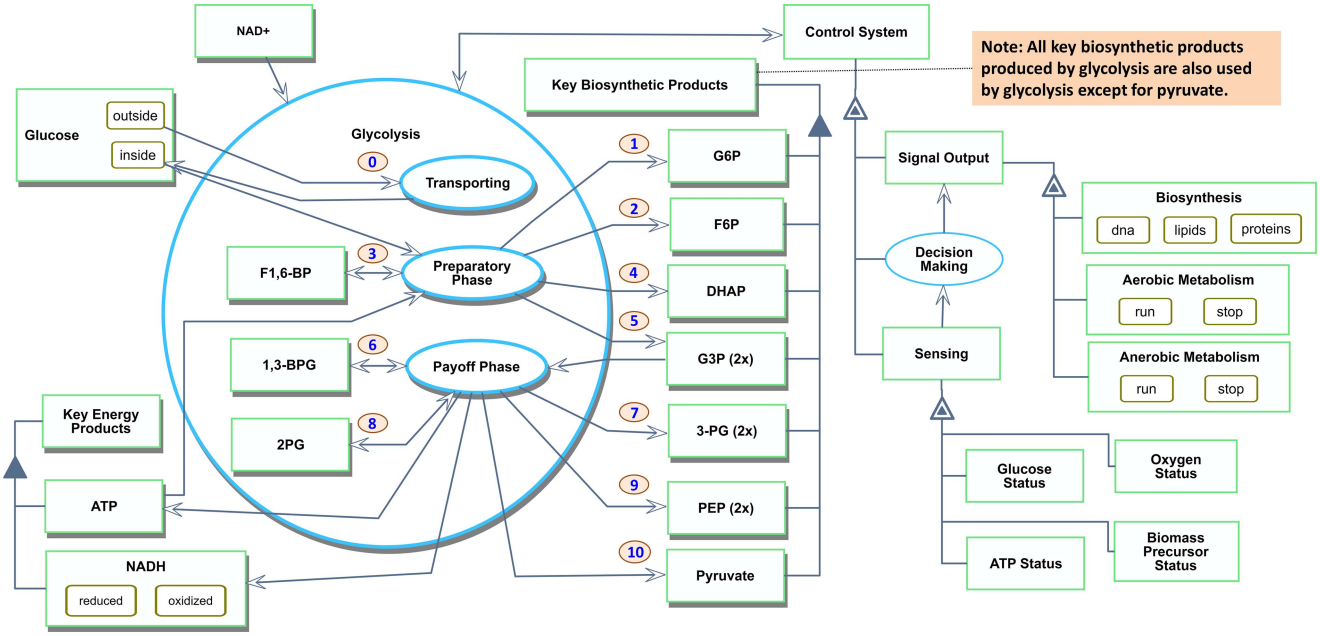


**FIGURE 6.** Glycolysis OPM model.

be shunted through the pay-off phase of the pathway. Nearly all the products of the enzymatic reactions are used in other parts of metabolism, including G6P, F6P, DHAP, G3P, etc.

### 2) INFER REQUIREMENTS

While an exhaustive catalog of glycolysis requirements is well beyond the scope of this article, some of the important additional inferred requirements are listed as follows:

15) *Glycolysis shall split the glucose molecule to extract the maximum energy*. See [37].
16) *Glycolysis shall not produce toxic metabolites in the pathway*. See [37] to see how the glycolysis pathway meets this requirement.
17) *Glycolysis metabolites shall not degrade too quickly*. This is because rebuilding metabolites is costly, and thus a half-life of greater than one minute is desired for key metabolites [37].
18) *Glycolysis metabolites shall not degrade too slowly*. This is to avoid excessive build-up of unused metabolites. Quantifying this requirement appears to be a knowledge gap.
19) *Glucose and derivative metabolites shall not leak out of the cell unless necessary for cellular signaling*. This requirement is met by adding a phosphate molecule to glucose after it is transported into the cell, and also by adding a second phosphate molecule prior to splitting. Adding phosphate molecules also reduces the activation energy needed to split the glucose for maximum energy yield.
20) *Glycolysis shall require energy as an input*. Although this follows directly from observational data, we suggest that this can be predicted since all energy production systems, such as power plants, require some form of energy input to get the system started, and glycolysis is fundamentally an energy production system (albeit one that is reconfigurable and which can also produce useful products at the expense of reduced energy output).
21) *Glycolysis metabolites shall have a high affinity for enzymes in order to maximize efficiency*. Adding phosphate molecules helps with affinity in addition to preventing leakage out of the cell [37]. This requirement could also be viewed as an enzyme requirement to have a high affinity for glycolysis metabolites.
22) *Glycolysis shall provide the minimal pathway to meet energy and biosynthesis requirements*. As per [37], the 10-step glycolysis process meets this requirement.

### 3) MODEL SYSTEM

Fig. 6 shows an OPM model corresponding to the zoomed-in *glycolysis* process of Fig. 5 that includes the standard steps 1–10. This model includes more detail than the higher level model of Fig. 5 while still providing a global view that allows the relationships and importance of different parts of the system to be better identified. This diagram also highlights the interfaces, which allows their interconnections and importance to be better recognized. In addition, we show an extra step 0 for the glucose *transporting* process incorporated within the glycolysis model. Although biologists traditionally have viewed glucose transport as separate from the 10-step glycolysis process, we have modeled glucose transport as part of glycolysis in the OPM model because glucose transport is a key system interface. Furthermore, we anticipate that including glucose transport as an integral subsystem of glycolysis will be necessary to develop a more complete understanding of phenomena, such as the Warburg effect. As discussed

earlier, *glycolysis* includes a *preparatory phase* that requires two ATP, followed by a *payoff phase* that produces ATP for a net gain of two ATP. In addition, *glycolysis* produces two NADH per glucose molecule, which under aerobic conditions can be converted to about five ATP by the *electron transport chain* shown in Fig. 4. Finally, *glycolysis* also produces 2 *pyruvate* molecules that can be converted to about 25 ATP by the *aerobic metabolism* process shown in Fig. 4 that includes *pyruvate processing*, *Krebs cycle*, and the *electron transport chain*. Thus, under aerobic conditions, *glycolysis*, in conjunction with additional processing, produces a net gain of about 32 ATP per glucose molecule. Under anaerobic conditions *glycolysis* produces a net gain of two ATP per glucose molecule, and relies on anaerobic *fermentation* to oxidize the *NADH* back to the *NAD+* required by *glycolysis*. The *control system* in Fig. 6 expands on the *control system* shown in Fig. 5 to support requirements #11 and related requirements #8, #9, and #10 in more detail than Fig. 5. The *control system* includes attributes for *sensing*, *decision-making*, and *signal output* to control *biosynthesis*, *aerobic metabolism*, and *anaerobic metabolism*.

#### 4) IDENTIFY KNOWLEDGE GAPS

We identified two knowledge gaps: 1) The first was found while inferring additional requirements from past research directly related to glycolysis: the optimal rate at which glycolysis metabolites should degrade. While some research places lower bounds on the rate of breakdown to avoid having to expend too much energy rebuilding metabolites (see requirement #17 and [37]), we did not find any similar results to quantify the requirement #18 upper bound to avoid too much metabolite buildup in the system. 2) We also identified another possible knowledge gap while developing the OPM model of glycolysis shown in Fig. 6. Note that nearly all of the intermediate metabolites also support building of *key biosynthetic products* (*G6P*, *F6P*, *DHAP*, *G3P*, *3-PG*, *PEP*, and *pyruvate*), and thus represent products in their own right. Furthermore, *F1,6-BP* has a known regulatory function in the glycolysis *control system*, and is thus also a product. Although *1,3-BPG* is too unstable to have any likely functionality, we hypothesize that the remaining metabolite, *2PG*, is also a product in its own right with additional functionality that has not yet been discovered. From a systems engineering perspective, it makes sense to have as many dual-use intermediates as possible to minimize the total number of different types of intermediates the system needs to be able to generate and accommodate.

#### 5) PREDICT AND TEST

We developed two predictions in conjunction with the glycolysis OPM model: 1) The first is that, as discussed above, the metabolite *2PG* associated with *glycolysis* step 8 is a product in its own right with additional functionality beyond the step 8 intermediate metabolite. 2) The second is that one possible approach to inhibiting cancer is to dampen the Warburg effect by inhibiting the GLUTs associated with the *transporting* process of Fig. 6. Although GLUT inhibitors have been utilized in combination chemotherapeutics [50], there is a need to better understand how the different types of GLUTs interact with inhibitors. We suggest that examining the GLUT response in the context of a normal response to tissue injury and repair may provide insights (assuming our hypothesis that the Warburg effect is really a normal tissue injury and repair response gone out of control as discussed earlier). It may be possible, for example, to inhibit the GLUT response more specifically by identifying and modifying tissue repair signaling associated with GLUT response to injury. Further research is needed in this area.

### D. OPM REQUIREMENTS EXAMPLE

In this section, we will explore the connection of requirements to structure and function within OPM, which is supported by the OPCloud development environment. In OPM, a requirement is represented as an informational object and is connected with an attribute link. Requirements may be satisfied by an object, process, or relationship and can have many component connections. In our methodology, we maintain requirements in a separate document and incorporate them into the relevant parts of the model as they are fulfilled. Mordecai and Dori [51] describe a requirement engineering process that is continuous and iterative. This process includes a step to assess whether or not a system requirement is already covered by the current design solution. To adopt their process for MBRSE in biology, an inferred requirement should be falsified if no connection to downstream structure and function can be established. If the entire architecture is satisfied by existing requirements, then there is no need to invoke additional requirements. In trying to connect requirements one might also find they are proposed for the wrong layer of the hierarchy. For instance, a requirement for a specific enzyme should not be put at the pathway level as it will not likely pertain to all versions of the glycolytic pathway. If there are exceptions to a requirement, then typically the requirement will need to be moved to a more specific context. Alternately, when requirements are not sufficient to explain all the structure and function there are likely unidentified upstream requirements.

To illustrate, consider the requirement that glycolysis be controllable. Fig. 7 shows the OPM glycolysis model zoomed into step 1 of glycolysis (*phosphorylation I*), which is facilitated by the *hexokinase catalytic output* being turned on. The state of the *hexokinase catalytic output* is determined by the *hexokinase decision-making* process, which depends on the physical *hexokinase control system*. The *hexokinase control system* includes the *hexokinase sensing* process which consists of four sensing subsystems: *substrate sensing* to sense glucose, *product sensing* to sense *G6P*, *cellular energy sensing* to sense *ATP* and *Pi*, and *organism energy sensing*. Each of these categories requires physical *innate enzyme design features*, such as *maximum catalysis rate* and *affinity for substrate* (referred to as $V_{max}$ and $K_m$ in the field of biochemistry). The crucial point is the need to validate requirements through

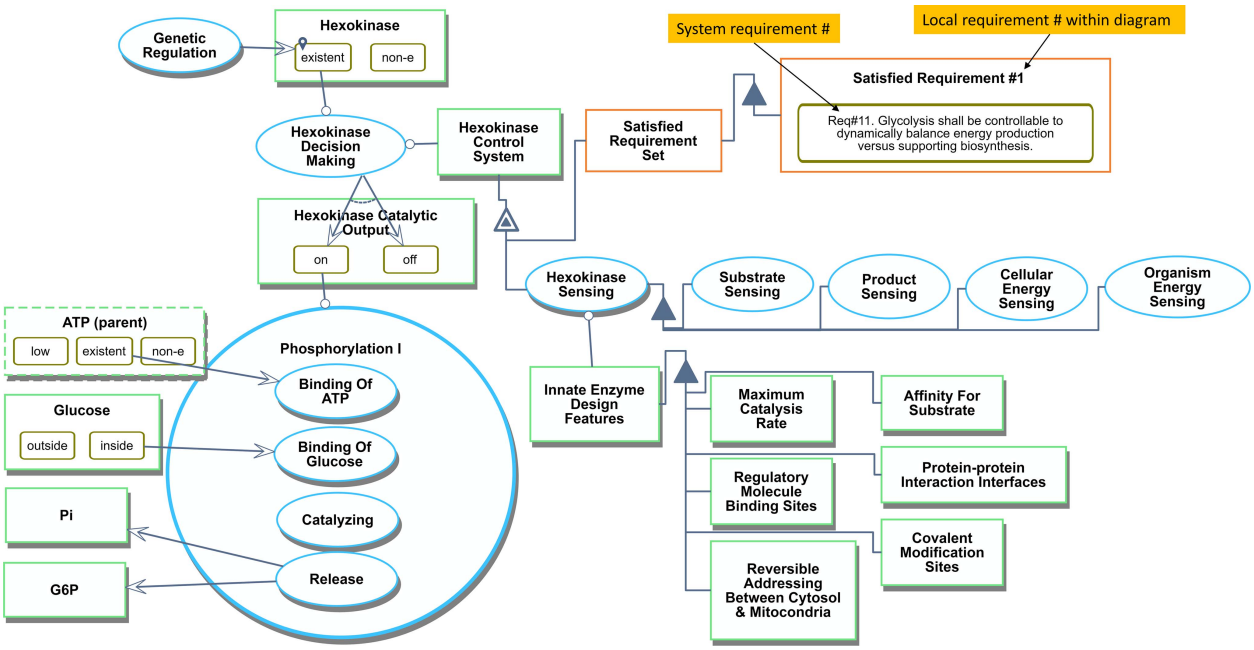


**FIGURE 7.** Glycolysis OPM model zoomed to first phosphorylation step with requirements.

observation of downstream structure and function. In our case, the observation of the detailed control system within step 1 suggests requirement #11 is met for this particular control subsystem. Accordingly, at the top of the diagram, we show that *hexokinase control system* satisfies requirement #11. We then continue this process for the entirety of glycolysis, linking all the components of control to this requirement to show that it is met globally for all derivative subsystems. Note that while linking this requirement did not create novelty outside the existing body of knowledge, it did allow for easy visualization and tracking of which components of glycolysis are related to regulation.

## IV. CONCLUSION

Our proposed MBRSE methodology for reverse engineering biological systems incorporates both top-down and bottom-up analysis with a goal of understanding a given biological system. The methodology starts with observations, models the system structure and function, and infers hypothetical requirements. As understanding of requirements develops, predictions can be made based on inferred requirements, allowing relationships in a system to be explained or hypothesized. This approach can identify knowledge gaps and avoid incorrect predictions due to missed requirements. Using glycolysis as a case study with SysML use case and OPM conceptual qualitative modeling in conjunction with requirements analysis, we were able to identify knowledge gaps (including the glycolysis hand-off from parent to child cell during mitosis not previously discussed in the literature) and make predictions (including that the Warburg effect commonly associated with cancer is caused by a normal response to tissue injury and repair that has become unstable due to a broken feedback mechanism in the glycolysis control system).

Our use case model generated knowledge gaps about the reason for the uniformity of glycolysis across life and the origin of the pathway. Traditionally, uniformity has been attributed as an artifact of common descent, meaning uniformity resulted from a historical relationship between all living organisms and does not have functional importance. However, in systems engineering, uniformity at a low level in a system design is often an optimized solution to upper-level requirements. We therefore propose that the striking similarity in the topology and metabolites of glycolysis across organisms is driven by a requirement for compatibility between organism energy interfaces, aiming to maximize efficiency at the ecosystem level. Furthermore, we predict that nonstandard interfaces between actors may necessitate deviations from the standard glycolytic pathway.

Our OPM cellular metabolism model highlights knowledge gaps for the transition between parent and child metabolism. Two predicted requirements stem from this model: 1) Glycolysis must accommodate varying energy and biomass demands as an organism progresses through development stages. We propose that the Warburg effect, associated with cancer, is related to this requirement. 2) The import of glucose into different tissues is a system level regulated step of glycolysis that is facilitated using the intrinsic design of unique GLUTs. By discerning the requirement-based rationale behind GLUTs operational characteristics, we provide a framework for understanding why they exhibit quantized behavior rather than possessing adjustable properties across a continuous spectrum.

With our glycolysis OPM modeling, we identified a gap in current knowledge for the optimality of glycolysis' metabolites in that while a lower bound has been established for the lifespan of glycolytic metabolites, an upper bound remains undefined. Next, we highlight that modeling glycolysis in

OPM revealed that most intermediates of the pathway are also key process input products, contrary to the traditional view that these are only intermediates. Given the observation of dual purpose for nearly all glycolytic intermediates, we predict that 2PG also possesses additional metabolic function beyond its established role as a glycolytic intermediate. We then propose that regulation of glucose uptake into cancerous tissues is an important area for future research since it is the first step of limiting glucose to a tumor's growth. This hypothesis underscores the translational implications of our research findings as it points toward novel strategies for combating cancer progression.

Finally, we show how to effectively integrate requirements into the structure and function of OPM models. This integration enables tracking the correspondence between specific system elements and the overarching requirements they fulfill. Such clarity is invaluable, as it enables anticipation of necessary modifications to downstream requirements in response to observed changes at upstream levels.

The glycolysis system requirements and corresponding models developed in this article serve as a foundation to catalyze further research within the field because our models, while retaining general applicability, can easily be tailored to specific organismic contexts.

We contend that a systems engineering approach with modeling and requirements analysis is the best solution for biological researchers to make sense of the enormous amounts of data that have been collected. While inferring requirements and building models are time-consuming front-end activities, MBSE languages, as illustrated by OPM, can be approachable for biologists, and with modest investment, could save a researcher countless hours trying to understand how to organize large amounts of raw data to elucidate understanding and make useful predictions. It can also help researchers identify the key questions to focus on, as well as provide a consistent model to reference, and improving communication within a large research team.

Potential future research suggested by this article include 1) closer examination of the glycolysis control system and its relationship to the Warburg effect and cancer; 2) computational modeling of glycolysis to supplement the conceptual modeling in this article; and 3) application to other biological systems.

## ACKNOWLEDGMENT

Gerald Fudge and Emily Brown Reeves thank Hanan Kohen for providing OPCloud support during our model development process. We also thank the reviewers of this article for their many useful recommendations.

## REFERENCES

[1] S. R. Hirshorn, L. D. Voss, and L. K. Bromley, "NASA systems engineering handbook (No. HQ-E-DAA-TN38707)," 2017.
[2] J. F. Garcia, "Un-building blocks: A model of reverse engineering and applicable heuristics," Ph.D. dissertation, Naval Postgraduate School, Monterey, CA, USA, 2015.
[3] B. W. Mar, "Back to basics again-a scientific definition of systems engineering," in *Proc. INCOSE Int. Symp.*, 1997, pp. 309–316.
[4] J. A. Estefan, "Survey of model-based systems engineering (MBSE) methodologies," *INCOSE MBSE Focus Group*, vol. 25, no. 8, pp. 1–2, 2007.
[5] A. F. Villaverde and J. R. Banga, "Reverse engineering and identification in systems biology: Strategies, perspectives and challenges," *J. Roy. Soc. Interface*, vol. 11, no. 91, Feb. 2014, Art. no. 20130505.
[6] J. Somekh, G. Haimovich, A. Guterman, D. Dori, and M. Choder, "Conceptual modeling of mRNA decay provokes new hypotheses," *PLoS One*, vol. 9, no. 9, Sep. 2014, Art. no. e107085.
[7] M. E. Csete and J. C. Doyle, "Reverse engineering of biological complexity," *Science*, vol. 295, no. 5560, pp. 1664–1669, Mar. 2002.
[8] U. Alon, "Biological networks: The tinkerer as an engineer," *Science*, vol. 301, no. 5641, pp. 1866–1867, Sep. 2003.
[9] S. Brenner, "Sequences and consequences," *Philos. Trans. Roy. Soc. B: Biol. Sci.*, vol. 365, no. 1537, pp. 207–212, Jan. 2010.
[10] S. A. Muñoz-Gómez, G. Bilolikar, J. G. Wideman, and K. Geiler-Samerotte, "Constructive neutral evolution 20 years later," *J. Mol. Evol.*, vol. 89, no. 3, pp. 172–182, Apr. 2021.
[11] Y. Lazebnik, "Can a biologist fix a radio?—Or, what I learned while studying apoptosis," *Cancer Cell*, vol. 2, no. 3, pp. 179–182, Sep. 2002.
[12] D. Noble, "It's time to admit that genes are not the blueprint for life," in *Books & Art, Nature*, vol. 626. Feb. 8, 2024.
[13] J. Yan, S. L. Risacher, L. Shen, and A. J. Saykin, "Network approaches to systems biology analysis of complex disease: Integrative methods for multi-omics data," *Brief. Bioinform.*, vol. 19, no. 6, pp. 1370–1381, Nov. 2018.
[14] M. Allarakhia and A. Wensley, "Systems biology: A disruptive biopharmaceutical research paradigm," *Technol. Forecasting Social Change*, vol. 74, no. 9, pp. 1643–1660, Nov. 2007.
[15] L. Chang, G. Zhou, O. Soufan, and J. Xia, "miRNet 2.0: Network-based visual analytics for miRNA functional analysis and systems biology," *Nucleic Acids Res.*, vol. 48, no. W1, pp. W244–W251, Jul. 2020.
[16] A. Nurgaliyev and H. Wang, "Analysis of reverse engineering," in *Proc. 15th Int. Conf. Adv. Comput. Intell.*, 2023, pp. 1–7.
[17] C. Raibulet, F. Arcelli Fontana, and M. Zanoni, "Model-driven reverse engineering approaches: A systematic literature review," *IEEE Access*, vol. 5, pp. 14516–14542, 2017.
[18] H. Bruneliere, J. Cabot, J. L. Canovas Izquierdo, L. O.-E. Arrieta, O. Strauss, and M. Wimmer, "Software modernization revisited: Challenges and prospects," *Computer*, vol. 48, no. 8, pp. 76–80, Aug. 2015.
[19] Y. Park, H. S. Won, and H. C. Song, "5.1. 3 traceability in a unified systems engineering framework for a high-speed railway system," in *Proc. INCOSE Int. Symp.*, 2000, pp. 39–43.
[20] S. Han, J. Kim, and J. G. Ohn, "The proposal for the application of reverse systems engineering to light rail transit project based on the outputs of ISO/IEC 15288 technical processes," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst.*, 2014, pp. 217–221.
[21] J. Topper and N. C. Stephen, "Model-based systems engineering in support of complex systems development," *Johns Hopkins APL Tech. Dig.*, vol. 32, pp. 419–432, 2013.
[22] P. J. Younse, J. E. Cameron, and T. H. Bradley, "Comparative analysis of a model-based systems engineering approach to a traditional systems engineering approach for architecting a robotic space system through knowledge categorization," *Syst. Eng.*, vol. 24, no. 3, pp. 177–199, May 2021.
[23] K. X. Campo, T. Teper, C. E. Eaton, A. M. Shipman, G. Bhatia, and B. Mesmer, "Model-based systems engineering: Evaluating perceived value, metrics, and evidence through literature," *Syst. Eng.*, vol. 26, no. 1, pp. 104–129, Jan. 2023.
[24] D. Dori and M. Choder, "Conceptual modeling in systems biology fosters empirical findings: The mRNA lifecycle," *PLoS One*, vol. 2, no. 9, Sep. 2007, Art. no. e872.
[25] J. Somekh, M. Choder, and D. Dori, "Conceptual model-based systems biology: Mapping knowledge and discovering gaps in the mRNA transcription cycle," *PLoS One*, vol. 7, no. 12, Dec. 2012, Art. no. e51430.
[26] J. D. Johansen, "Bacterial chemotaxis control process analysis with SysML," *Syst. Eng.*, pp. 1–13, 2024, doi: 10.1002/sys.21752.
[27] M. Leimeister, "Python-modelica framework for automated simulation and optimization," in *Proc. 13th Int. Modelica Conf.*, 2019, pp. 51–58.
[28] M. C. Hause and R. L. Day, "Frenemies: OPM and SysML together in an MBSE model," in *Proc. INCOSE Int. Symp.*, 2019, pp. 691–706.

[29] D. Dori, A. Jbara, N. Levi, and N. Wengrowicz, "Object-process methodology, OPM ISO 19450—OPCloud and the evolution of OPM modeling tools," *Syst. Eng. Newslett.—Project Perform. Int.*, vol. 61, pp. 6–17, Jan. 2018.
[30] D. Dori et al., "OPCloud: An OPM integrated conceptual-executable modeling environment for Industry 4.0," in *Systems Engineering in the Fourth Industrial Revolution*. Hoboken, NJ, USA: Wiley, 2019, pp. 243–271.
[31] N. Levi-Soskin, A. Jbara, and D. Dori, "The model fidelity hierarchy: From text to conceptual, computational, and executable model," *IEEE Syst. J.*, vol. 15, no. 1, pp. 1287–1298, Mar. 2021.
[32] C. Wanjek, "Systems biology as defined by NIH: An intellectual resource for integrative biology," *NIH Catalyst*, vol. 19, pp. 1–12, 2011.
[33] N. S. Chandel, "Glycolysis," *Cold Spring Harbor Perspectives in Biology*, vol. 13, no. 5, 2021, Art. no. a040535.
[34] M. Ralser, S. J. Varma, and R. A. Notebaart, "The evolution of the metabolic network over long timelines," *Curr. Opin. Syst. Biol.*, vol. 28, Dec. 2021, Art. no. 100402.
[35] G. Schneider and J. P. Winters, *Applying Use Cases: A Practical Guide*. Reading, MA, USA: Addison-Wesley, 2001.
[36] M. Ralser, "An appeal to magic? The discovery of a non-enzymatic metabolism and its role in the origins of life," *Biochem. J.*, vol. 475, no. 16, pp. 2577–2592, Aug. 2018.
[37] A. Bar-Even, A. Flamholz, E. Noor, and R. Milo, "Rethinking glycolysis: On the biochemical logic of metabolic pathways," *Nature Chem. Biol.*, vol. 8, no. 6, pp. 509–517, May 2012.
[38] P. H. Raven, S. Singer, K. A. Mason, G. B. Johnson, and J. Losos, *Raven, Biology © 2014, 10e*, AP student ed. New York, NY, USA: McGraw-Hill, 2013.
[39] A. Cornish-Bowden and M. L. Cárdenas, *Control of Metabolic Processes*, vol. 190. Berlin, Germany: Springer-Verlag, 2013.
[40] O. Ebenhöh and R. Heinrich, "Evolutionary optimization of metabolic pathways. Theoretical reconstruction of the stoichiometry of ATP and NADH producing systems," *Bull. Math. Biol.*, vol. 63, no. 1, pp. 21–55, Jan. 2001.
[41] S. J. Court, B. Waclaw, and R. J. Allen, "Lower glycolysis carries a higher flux than any biochemically possible alternative," *Nature Commun.*, vol. 6, no. 1, Sep. 2015, Art. no. 8427.
[42] E. Noor, E. Eden, R. Milo, and U. Alon, "Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy," *Mol. Cell*, vol. 39, no. 5, pp. 809–820, Sep. 2010.
[43] C. Y. Ng, L. Wang, A. Chowdhury, and C. D. Maranas, "Pareto optimality explanation of the glycolytic alternatives in nature," *Sci. Rep.*, vol. 9, no. 1, Feb. 2019, Art. no. 2633.
[44] A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, and R. Milo, "Glycolytic strategy as a tradeoff between energy yield and protein cost," *Proc. Nat. Acad. Sci.*, vol. 110, no. 24, pp. 10039–10044, Jun. 2013.
[45] A. Klingner et al., "Large-scale 13C flux profiling reveals conservation of the Entner–Doudoroff pathway as a glycolytic strategy among marine bacteria that use glucose," *Appl. Environ. Microbiol.*, vol. 81, no. 7, pp. 2408–2422, Apr. 2015.
[46] M. V. Liberti and J. W. Locasale, "The Warburg effect: How does it benefit cancer cells?," *Trends Biochem. Sci.*, vol. 41, no. 3, pp. 211–218, Mar. 2016.
[47] M. G. V. Heiden, L. C. Cantley, and C. B. Thompson, "Understanding the Warburg effect: The metabolic requirements of cell proliferation," *Science*, vol. 324, no. 5930, pp. 1029–1033, May 2009.
[48] B. Thorens and M. Mueckler, "Glucose transporters in the 21st century," *Amer. J. Physiol.-Endocrinol. Metab.*, vol. 298, no. 2, pp. E141–E145, Feb. 2010.
[49] O. Gómez, B. Ballester-Lurbe, E. Poch, J. E. Mesonero, and J. Terrado, "Developmental regulation of glucose transporters GLUT3, GLUT4 and GLUT8 in the mouse cerebellar cortex," *J. Anatomy*, vol. 217, no. 5, pp. 616–623, Nov. 2010.
[50] K. Tilekar, N. Upadhyay, C. V. Iancu, V. Pokrovsky, J.-Y. Choe, and C. S. Ramaa, "Power of two: Combination of therapeutic approaches involving glucose transporter (GLUT) inhibitors to combat cancer," *Biochimica Biophysica Acta-Rev. Cancer*, vol. 1874, no. 2, Dec. 2020, Art. no. 188457.
[51] Y. Mordecai and D. Dori, "Model-based requirements engineering: Architecting for system requirements with stakeholders in mind," in *Proc. IEEE Int. Syst. Eng. Symp.*, 2017, pp. 1–8.

**GERALD L. FUDGE** (Senior Member, IEEE) received the B.A. degree in mathematics from the Rice University, Houston, TX, USA, in 1983, the master's degree in applied mathematical sciences from the Rice University, Houston, TX, USA, in 1984, and the Ph.D. degree in electrical engineering from the University of Texas at Dallas, Richardson, TX, USA, in 1994.

He has been an Assistant Professor with the Engineering and Technology Department, Texas A&M University—Commerce, Commerce, TX, USA, since retiring from the L3Harris in 2021 as a Senior Fellow. While with the L3Harris, he led the research and development of advanced electronic intelligence systems, including analog and digital hardware with associated detection and measurement algorithms. He also performed systems engineering for technology transition to operational deployment. Prior to L3Harris, he was a Nuclear Engineer in the U.S. Naval submarine force. He is the inventor or coinventor on 24 U.S. patents and 3 trade secrets, including the Nyquist Folding Receiver, and has more than 20 peer-reviewed publications. His research interests include the intersection of systems engineering with biology; applied machine learning in engineering, agriculture, wildlife sciences, and human performance; engineering education; and innovation in engineering.

**EMILY BROWN REEVES** received the B.S. degree in chemistry with a minor in biology from the Northern Arizona University, Flagstaff, AZ, USA, in 2012, and the Ph.D. degree in biochemistry and biophysics from the Texas A&M University, College Station, TX, USA, in 2018.

She currently collaborates on several research projects at the intersection of biology and engineering. Alongside her research pursuits, she serves as an Active Clinician specializing in nutritional genomics with the Nutriplexity, Bremerton, WA, USA, and contributes as a Volunteer Curator for the American College of Medical Genetics Inborn Errors of Metabolism Disease Panel. Since the summer of 2012, when she interned with the Department of Biological Engineering, MIT, Cambridge, MA, USA, she has been interested in how engineering methodologies can deepen understanding of biological systems. Her research interests include efforts to adapt systems engineering methods for use in biology, namely object-process methodology, so that better models of metabolism and biological network communication can be built. Ultimately, she aims to discover the general underlying design principles of organisms for the purpose of remediating human disease.