metabolites

MDPI

Article

The *Staphylococcus epidermidis* RP62A metabolic network: Validation and intervention strategies

Francisco Guil * , Guillermo Sánchez-Cid and José M. García

Grupo de Arquitectura y Computación Paralela, Universidad de Murcia, 30080 Murcia, Spain; fguil@um.es (F.G.); jmgarcia@um.es (J.M.G.)

* Correspondence: fguil@um.es

Received: date; Accepted: date; Published: date

check for updates

Abstract: Increasingly, systems biology is gaining relevance in basic and applied research. The combination of computational biology with wet lab produces a synergy that results in an exponential increase in knowledge of biological systems. The study of microorganisms such as *Staphylococcus epidermidis* RP62A enables the researcher to understand better its metabolic network, which allows the design of effective strategies to treat infections caused by this species or others. *S. epidermidis* is the second cause of infection in patients with joint implants, so treating its proliferation seems vital for public health. There are different approaches to the analysis of metabolic networks. Flux Balance Analysis (FBA) is one of the most widespread streams. It allows the study of large metabolic networks, their structural properties, the optimization of metabolic flux, and the search for intervention strategies to modify the state of the metabolic network. This work presents the validation of the *Staphylococcus epidermidis* RP62A metabolic network model elaborated by Díaz-Calvo *et al.*. Then, we elaborate further on the network analysis's essential reactions, classifying them. Finally, we introduce some proposals to intervene in the network and design knock-outs.

Keywords: *Staphylococcus epidermidis*, metabolic network validation; minimal cut sets; knock-outs; systems biology

1. Introduction

During the last two decades, the study of biological systems from a holistic point of view, together with the application of improvements in laboratory techniques and computational resources, has allowed a significant development of knowledge of living organisms at the molecular level. Cellular metabolism can be studied as a set of reactions that occur in the cell and allow its development and activity. These reactions are coordinated to maintain the balance of the cellular metabolism, also known as cellular homeostasis. With the assistance of high-performance computing, the metabolic network can be examined better to understand the overall activity of a whole cell. The metabolic network enables the generation of large amounts of omic data that allow identifying and quantifying biological molecules to discover the connections between genotype and phenotype. Then, researchers can make predictions of hypothetical states of the metabolic network by generating metabolic fluxes at different levels of the biological system.

Due to a large amount of omic data continuously generated, the most successful approaches are those known as data-driven ones. These are based on combining omics data and computational methods to generate parts or complete metabolisms called Genome-Scale Metabolic Models (GSMMs) [1]. GSMMs are mathematical representations that describe a complete set of stoichiometry-based mass-balanced metabolic reactions in an organism using gene-protein rule (GPR) association. GSMMs simulate how metabolism regulates and responds to changes, endogenous or exogenous, as mutations or changes in environmental conditions. The use of GSMMs has been successful in many applications,

including the discovery of drug targets in pathogenic organisms, the prediction of enzyme functions, the analysis of multiple reactomes, and modeling cell-to-cell interactions in studying human diseases [2].

After the generation of GSMMs, the network has to be analyzed. To this end, the modeling approaches can be divided into two frameworks: constraint-based modeling (CBM) and kinetic modeling. Constraint-based modeling (CBM) offers a representation of the metabolic network based on structural and stoichiometric aspects of the network with the aim of simulations of different states of metabolic flux and thus predicting phenotypes. On the other hand, the kinetic model aims to study the variations in metabolite concentrations over time through parameters related to the kinetics of chemical reactions and the concentration of enzymes in the metabolic network. Due to the nature of both frameworks, the first is used to study large-scale metabolic networks in a steady state; the second is restricted to networks of reduced size or parts of more extensive networks. Both frameworks are compatible and can be used to obtain a complete view of the model and draw more well-founded conclusions [3].

Focusing on CBM, one of the main tools to develop the analysis is Flux Balance Analysis (FBA), a computational method to predict flux distributions while optimizing (maximizing or minimizing) a given cellular function or a combination of them [4]. Inside FBA there are plenty of tools, but in this study, we are mainly concerned with Minimal Cut Sets (MCS). This tool tries to determine minimal interventions necessary through minimum sets of deletions to be carried out in a metabolic network at the protein or gene level to generate phenotype knock-outs [5].

In this work, an analysis of the GSMM model of the *Staphylococcus epidermidis* RP62A pathogen proposed by Teresa Díaz Calvo *et al.* [6] has been performed. For the first time, they presented the biological network model fixing the equations in their paper by observing the real behavior of the pathogen in the wet laboratory. Their paper has attracted us due to the importance of the *Staphylococcus epidermidis* pathogen. Therefore, we thought we could check and validate the proposed model and extract from the network model some intervention strategies to produce knock-outs in the pathogen.

S.epidermidis is a human opportunistic pathogen known to cause infections in medical implants (joints and heart). It is a facultative anaerobic bacterium with a gram-positive coccus phenotype that produces biofilm that, in the absence of disease, is located in collagen-rich regions such as the skin and mucous membranes. It belongs to the *Staphylococcus* genus, a widely extended group with well-known species such as *Staphylococcus aureus*, known for causing respiratory infections, among others. Concretely, prosthetic joint infections (PJI) are a recurring problem during arthroplasty or knee replacement. This type of infection occurs during the operation due to contamination of the material or the implant. PJI occurs when bacteria manage to adhere to the surface of the prosthetic implant and colonize it through the production of biofilm. Most PJIs are of monomicrobial origin, with the *Staphylococcus* genus being the first etiological agent (*S. aureus* and *S.epidermidis*, in order of incidence). Arthroplasty allows millions of patients to relieve their pain and recover or improve the mobility and functionality of the operated area. However, it is common for the diagnosis of PJI to be made when it is already chronic. At this point, surgical replacement of the implant is the only treatment with guarantees of success, although with quite a few possibilities of relapse [7]. The economic impact of PJIs is significant. Arthroplasty is a frequently performed procedure, and due to the cumulative effect of the aging population and increased life expectancy, it has a growing incidence. Therein lies the importance of studying the metabolic network of one of the main aetiological agents (*S. epidermidis*) to find solutions to prevent the growth of this bacterium (drugs, recombinant technologies, etc.). Finding feasible and safe solutions would improve patients' quality of life and reduce the economic impact on the health system.

There are several ways to intervene in the metabolic network. The first way of doing so is by knocking out one (or several) internal reactions. In this approach, the minimal interventions are provided by the essential reactions (those active in any non-trivial state of the network). Canceling one of these reactions, the cell must die (internal cut sets). A second way is removing (or limiting) some of

 doi:10.20944/preprints202208.0059.v1

the available nutrients. This limitation can be simulated by knocking-out exchange reactions associated with external metabolites (external cut sets). Finally, a third approach (a mix of the previously mentioned ones) is based on knocking out both internal and exchange reactions (hybrid cut sets).

The main goals we try to cover in our work are the following:

- Replicate the results obtained from Díaz Calvo *et al.* [6], validate the model, and discuss improvements to the calculations performed in the original work.
- Provide a better understanding of the *Staphylococcus epidermidis* RP62A metabolic network.
- Propose new minimal interventions to the metabolic network to feasibly eliminate *S.epidermidis*.

2. Results

2.1. Validation of the model

The original paper posed an optimization problem to obtain a minimum total flux state of the metabolic network using Equation 5 and replicating the conditions and restrictions given in the model. They obtained a unique solution consisting of 227 reactions (excluding transport pseudo-reactions), of which 127 were essential. The total flux value produced by this solution was not reported.

In our work, we have successfully imported the model using COBRApy. We have replicated the optimization problem with the same conditions and restrictions. As the optimal solution is not unique, we have not been able to find the exact solution reported in the original paper. Instead, we have calculated some optimal solutions. In this paper, we have worked with 108 optimal solutions, their support size from 295 to 344 reactions (from 218 to 268 reactions if transport pseudo-reactions are discarded). We have found 130 reactions (not counting the transport reactions) present in all the 108 obtained solutions and 553 different reactions that appear in the support of at least one solution. The Supplementary Material details all this information.

Regarding the essential reactions, we have found 128 ones (excluding the ATPase reaction). This value is nearly identical to the one obtained in the original paper (127 essential reactions). The new essential reaction we have found is the $DIACYLGLYKIN - RXN$ one.

Moreover, we have used two other techniques to solve the optimization problem: Mixed-integer linear programming (MILP) and a new optimization problem by creating a lumped biomass pseudo-reaction that gathers all its precursors. The solutions obtained in both methods are very similar and entirely consistent with those obtained with the previous LP method.

In their paper, Díaz Calvo and colleagues also calculated the impact of removing individual amino acids. The impact is obtained by Euclidean distance between the flux vectors obtained by solving the original FBA problem and the one obtained after removing each amino acid. We decided not to replicate this part of the research because, as previously stated, the solution associated with the optimal flux value is not unique, so it makes no sense to estimate the impact of the elimination of an amino acid by just comparing two of those solutions.

2.2. Minimal intervention strategies

2.2.1. Internal minimal cut sets

We have classified the essential reactions (internal cut sets of lenth 1) into 35 equivalence classes. 17 essential reactions have no equivalent ones, while the other 111 are located in 18 equivalence classes with more than one element.

After studying these equivalences, we can also analyze their implications. This study leads to distinguishing between different types of essential reactions giving interesting additional information about them.

 doi:10.20944/preprints202208.0059.v1

Remember that essential reactions can be viewed as minimal cut sets of length 1 for the biomass precursors. So, we have also explored other ways of achieving the intervention in the metabolic network for these biomass precursors by finding minimal cut sets of greater length.

2.2.2. External minimal cut sets

Focusing only on reactions associated with the importation of nutrients, we have found a new minimal cut set of length 2 (composed of cysteine and sulfate ion). Due to the importance of these two reactions, a deeper study of their impact on the biomass precursors has been carried out. The results are included in the Supplementary Material.

2.2.3. Hybrid minimal cut sets

There are no more minimal cut sets consisting only of exchange reactions. So the next natural step is to calculate cut sets in which culture medium components were intervened along with some internal reactions (hybrid CS). In this sense (using a method similar to the well-known Berge Algorithm [8]), 48 hybrid minimal cut sets of length 2 formed by a medium reaction and an internal one have been calculated. There is also a hybrid cut set of length 2 involving *sulfate* (which is not in the medium) and *RIB5PISOM-RXN*. Continuing with the idea of finding ways to intervene in the network and taking into account that cysteine and sulfate are essential for the functioning of the network, hybrid cut sets were calculated in which at least one of the two elements was present in the CS. In this sense, an additional minimal cut set of length 3 containing cysteine has also been found. The complete list of computed minimal cut sets can be downloaded from https://github.com/biogacop/SEpidermidis_Analysis.

3. Discussion

3.1. Model validation and best practices

It has been verified that the model created by Díaz Calvo *et al.* is interesting and has quite a good quality to be so recent. The model is in its first version, and it would be very positive to improve aspects such as the standardization of the *ids* and *names* of both metabolites and reactions.

Even though the solution to the optimization problem posed is not unique, it has been possible to obtain solutions that are very similar to that of Díaz Calvo *et al.*. However, it has not been possible to check if we have obtained precisely the same solution as in the reference due to the lack of the total flux value. Even if the comparison between solutions cannot be completely accurate, we consider that the model has been validated by reproducing these results.

The minimization problem has also been studied from a more technical point of view. The same metabolic network flux minimization problem has been solved using MILP. We have observed that the calculated solutions are similar to those obtained using LP methods. Once again, getting multiple solutions shows that the network has different states for the same conditions. Regarding the inclusion of a lumped biomass pseudo-reaction, the results obtained fall within the range of support sizes and the optimal obtained total flux.

Regarding the concentrations of the components of the culture medium, the flux of the reactions that incorporate them into the metabolic network has been analyzed. We have looked for the optimal total flux conditions subject to the minimization objective function. We have observed that the maximum value consumed for some components (isoleucine, leucine, lysine, methionine, tyrosine and phenylalanine, and niacin (vitamin B3)) in minimal flux condition is lesser that the maximum amount available in the culture medium (a table describing it can be consulted in Supplementary Material). After all, the models represent living beings, and these do not always incorporate all the nutrients available in the environment, only those that are required for the optimal functioning of the metabolism.

We have considered that the maximum limit of nutrients available in the medium that has been set is the concentration of nutrients that *S.epidermidis* incorporates into its metabolism under steady-state

conditions and continuous culture. The one that has taken less concentration of certain nutrients can be due to the conditions in which the state of the metabolic network is simulated. In short, the biological system is subjected to optimal minimum flux conditions (like a basal state). Therefore, it is logical to think that the body in basal conditions does not have the same energy and nutritional needs as when it is in ideal conditions. Even so, it is striking that these components of the medium are not fully incorporated, and yet the rest of the components are.

Finally, after our experience replicating the network model, we have considered proposing a few points that should be taken into account when building a model (best practices):

- Homogenize the *id* and *name* naming system. Although, indeed, the notation of the model's reactions, metabolites, and genes is usually automatized, when curating the model, a *naming* system must be taken into account to maintain the consistency of the model. Otherwise, it can lead the observer to misinterpret the model, making it difficult to interact with the metabolic network and make calculation mistakes.
- There are several possible reasons for a metabolite to be a *dead-end*: missing complete annotation, missing/absent exchange reactions, or simply that the reaction cannot carry flux at steady-state. In any case, this model is expected to reduce this ratio in future updates. So it is important to declare the external and dead-end metabolites of the model explicitly. If there are differences between external metabolites that represent the limit of the metabolic network represented by the model and metabolites related to the representation of biomass and by-products, they should be mentioned.
- Detail on which databases (version) and organisms (assembly accession) the notation has been based on to build the model.
- Declare if there are pseudo-reactions or pseudo-metabolites and what function they fulfill in the model.
- If a solution to an FBA optimization problem is given, explicitly declare the total flux obtained. The solution support should be included in the Supplementary Material if the solution is unique. This statement is for reproducible purposes.

In conclusion, we have observed that the solution indicated in the original paper fits perfectly in the ranges of values we have found. Therefore, we claim that the original model is correct, and this can be validated.

3.2. Minimal intervention strategies

After the metabolic network model, some studies can be done to give insights and properties of the *Staphylococcus epidermidis* behavior.

Regarding the essential reactions, we have found $DIACYLGLYKIN - RXN$, a previously unobserved essential reaction. We have found that this reaction is different from the other essential ones. All the other essential reactions are cut sets for at least a biomass precursor in the original model (leaving aside the additional restrictions given by the values of the biomass precursors). That is, for each essential reactions, r_i, there is at least a biomass precursor, r_j, such that for any mode $v \in C$, $v_i = 0$ implies $v_j = 0$. This implication is not true for $r = DIACYLGLYKIN - RXN$, since in this case, each biomass precursor can be active even if r is inactive. However, if r is inactive, it is not possible for all the biomass precursors to simultaneously achieve their correspondent values in the biomass composition. Let us remark that this reaction is catalyzed by the enzyme diacylglycerol kinase (ATP), a transferase that catalyzes the ATP-dependent phosphorylation of phospholipids for the cell membrane. Among the different phospholipids that this reaction can synthesize is diacylglycerol (DAG), one of the biomass precursors present in the *S.epidermidis* RP62A model. Therefore, it is a reaction that participates indirectly in synthesizing this precursor.

Additionally, we have studied some intervention methods in the metabolic network to find ways to design possible knock-outs for the network. In this sense, we have examined the set of essential reactions, including the previously unobserved one.

All possible implications among essential reactions have been calculated and used to obtain equivalences. It is worth noting that there appear trivial equivalences (those given by a coupling of reactions in a metabolite like $r_0 \to m \to r_1$) and non-trivial ones (the causal relationship is not known at first glance).

After obtaining these equivalences, we have also studied their implications to distinguish between direct and indirect implications. Let us start by observing that if r_i is a cut set for a target set of reactions T and r_j is a cut set for r_i, then r_j is also a cut set for T. So we can think of the set of essential reactions (cut sets of length 1 for the biomass precursors) as chains formed by cut sets of those precursors, cut sets for these (primary) cut sets, and so on. This distinction allows us to differentiate between reactions whose deletion directly blocks a biomass precursor (direct implication) and those whose deletion blocks a reaction that directly implies a biomass precursor (indirect implication). This classification leads to a better explanation of the essential reaction behavior.

Two subsets of essential reactions are significant here: the set of direct implications and final ones (that is, cut sets for the biomass precursors that have no other cut sets of length 1). Regarding the direct implications, we have found 20 of them (the complete list is provided in Supplementary Materials). On the other hand, we have seen the following 12 final essential reactions:

1. 5 evident implications: the ATPase reaction, the NAD cofactor activation reaction, the hydrolysis of pyrophosphate (which has many secondary implications), RXN66-532 (Alpha-D-phosphohexomutase, catalyzes the interconversion between glucose-6-phosphate and alpha-glucose-1-phosphate), and the phosphorylation of diacylglycerol, which participates in the glycolytic pathway.
2. 7 other implications: NICONUCADENYLYLTRAN-RXN (nicotinate-nucleotide adenylyltransferase, adenylation of nicotinate mononucleotide to nicotinic acid adenine dinucleotide), RXN-12002 (UMP/CMP kinase, phosphorylates UMP to UDP), SHIKIMATE-5-DEHYDROGENASE-RXN (shikimate dehydrogenase (NADP+), catalyzes the reversible NADPH linked reduction of 3-dehydroshikimate to shikimate and involved in the biosynthesis of aromatic amino acids), HYDROXYMETHYLGLUTARYL-COA-SYNTHASE-RXN (hydroxymethylglutaryl-CoA synthase, participates in the ergosterol biosynthesis by condensing acetyl-CoA with acetoacetyl-CoA to yield hydroxymethylglutaryl-CoA), IPPISOM-RXN (isopentenyl-diphosphate *delta*-isomerase, catalyzes the isomerization of isopentenyl pyrophosphate to isopentenyl pyrophosphate taking part in the biosynthesis of isoprenoids), PRPPSYN-RXN (ribose-phosphate diphosphokinase, involved in the chorismate synthesis pathway, which is part of the synthesis of aromatic-type amino acids), and Palmitate_synth (palmitate synthase, yielding palmitate a saturated fatty acid which is a component of the cell membrane) .

A detailed study of the structure of the essential reactions can be found in the Supplementary Material.

We have also explored other ways of achieving the intervention in the metabolic network for the biomass precursors. The results are included in the Supplementary Material. Focusing only on reactions associated with the importation of nutrients, we have found a new minimal cut set of length 2 (composed of cysteine and sulfate ion). Due to the importance of these two reactions, a deeper study of their impact on the biomass precursors has been carried out. This minimal cut set (MCS) is interesting as the *S.epidermidis* RP62A metabolic network is capable of synthesizing the amino acid cysteine but requires sulfate. However, a sulfate MCS alone would not be viable since the culture medium also provides exogenous cysteine to the microorganism so that it can be incorporated into the metabolic network through transporters. Consequently, even if the entry of sulfate into the metabolic network

 doi:10.20944/preprints202208.0059.v1

or even enzymes of the cysteine synthesis pathway were interrupted, the metabolic network would continue to carry the flux through another subset of reactions; in other words, the microorganism would not die. This MCS corroborates that there is no auxotrophy for cysteine.

This MCS provides minimal intervention through the deprivation of cysteine and sulfate from the medium or by blocking the transports that facilitate the entry of these components into the interior of the microorganism. This finding can lead to clinical research on treatments that try to reduce the proliferation of *S.epidermidis* in infections suffered by patients. Drugs such as antibiotics (experimented related to tetracycline water-soluble formulations) [9] and new antimicrobial biomaterials such as nisin/polyanion layer-by-layer films [10] have been investigated but none related to the intervention of cysteine and sulfate at a medium level or metabolic pathways involving these two biomolecules.

Continuing with the study of the MCS of the network, we have checked for which reactions sulfate was essential and for which cysteine. As expected, the reactions that specifically required cysteine were those that participate in pathways related to cysteine biosynthesis. The only one not associated with sulfate was the reaction involving ribose-5-phosphate isomerase, belonging to the pentose phosphate pathway.

On the other hand, the reactions of biomass precursors blocked in the absence of cysteine and sulfate have also been explored. The results have been as expected since the blocked reactions were related to cysteine, methionine, and compounds necessary for synthesizing both amino acids and cofactors acetyl-CoA and CoA. The utilization of cysteine to synthesize methionine and vice versa is a process that occurs continuously through the transsulfuration pathway. Both are sulfur-containing amino acids and are of vital synthesis importance for a wide variety of enzymes (e.g., cysteine can form disulfide bonds with other cysteine residues, which plays a crucial role in the folding and protein structure). Another aspect of interest is the reactions in which cysteine and sulfate are involved.

An interesting aspect for future works would be to carry out the gene and functional annotation of the genes that code for the enzymes that carry out the reactions of interest.

4. Material and Methods

4.1. Constraint-Based Modeling

A metabolic model is defined by metabolites, the reactions that interconvert them (described by a fixed stoichiometry), and the fluxes that quantify the extent to which each reaction intervenes. The different components that make it up should be identified to establish a metabolic model.

Let R and M be the sets of reactions and metabolites of the system and denote by m, n the number of metabolites and reactions in the network. Each reaction would convert specific amounts of some metabolites into amounts of other metabolites. This information is provided by the stoichiometric coefficients of the reactions that can be summarized in a matrix $S \in M_{m\times n}(\mathbb{R})$.

Each possible state of the network can be represented by a vector $v \in \mathbb{R}^n$ where the components r_i indicates the amount of flux through reaction r_i and a vector $x \in \mathbb{R}^m$ to represent the concentration of each metabolite $m_j \in M$. The variation of concentrations of the metabolites is summarized in Equation 1

$$\frac{dx}{dt} = S \cdot v \tag{1}$$

The state of equilibrium in which the concentrations of internal metabolites of the metabolic network do not change over time (production and consumption of metabolites by r biochemical reactions are equal) is called the steady-state constraint [11]

$$S \cdot v = 0 \tag{2}$$

Observe that, in this formulation, only internal metabolites must be included as rows in the stoichiometric matrix S.

Considering the thermodynamic directionality, the reactions of a metabolic network can be classified into irreversible and reversible. A reaction $r_i \in R$ is said to be irreversible if it can only carry flux in one possible direction. Irreversible reactions from a subset $Irr \subset R$. Reversible reactions are those that can carry flux in both directions.

The restrictions on the sign of the flux of the irreversible reactions are called the thermodynamic constraint (it will be supposed that all irreversible reactions always occur in the positive direction [12])

$$r_i \geq 0 \; \forall r_i \in R \tag{3}$$

A mode (or state) of the network is a flux vector satisfying Equations 2 and 3. Two states of the network are considered equivalent if one is a non-negative multiple of the other. With a slight abuse of notation, they are always identified as the same mode. The set of all possible network modes (the flux cone of the metabolic network) is denoted as C.

$$C = \{v \in \mathbb{R}^n \mid S \cdot v = 0, v_i \geq 0 \; \forall r_i \in Irr\}$$

Given a mode $v \in C$, its support, $supp(v)$, is defined as the set of reactions that appear with non-zero flux in v. That is

$$supp(v) = \{r_i \in R \mid v_i \neq 0\}$$

A reaction r is said to be essential if $r \in supp(v) \; \forall \; 0 \neq v \in C$. That is, r is active in any non-trivial state of the metabolic network. Having the essential reactions implies knowing which reactions are necessary for the organism's life. On the other hand, a reaction is said to be blocked if it is always inactive.

Given a set of target reactions, $T \subset R$, a cut set for T is a set of reactions whose inactivation induces the inactivation of all the reactions in T. Formally, if $T \subset R$ is a set of target reactions, a subset $S \in R$ is called a cut set for T if $S \cap T = \emptyset$ and $\forall v \in C$, $supp(v) \cap S = \emptyset \Rightarrow supp(v) \cap T = \emptyset$. A cut set S is minimal (MCS) if $\nexists \; S' \subset S$ such that S' is also a cutset for T (see [13]).

Another important concept is that of implication between reactions. Given two unblocked reactions r_i and r_j, it is said that:

- r_i (partially) implies r_j if whenever r_i is active, r_j must also be active.
- r_i is equivalent to r_j if $r_i \rightarrow r_j$ and $r_j \rightarrow r_i$.
- r_i (totally) implies r_j if $r_i \rightarrow r_j$ and there is a constant $0 \neq c$ such that in any state in which r_i is active, $v_j = c \cdot v_i$ is satisfied.

Flux-balance analysis. Flux Balance Analysis (FBA) performs the network analysis of a model. In this framework, adding a function transforms our steady-state and thermodynamic constraints into a linear programming (LP) optimization problem. Concretely, if $\{a_i\} \in \mathbb{R}^n$ is any sequence of constants, the linear function on the fluxes $f(v) = \sum_{i=1}^{n} a_i \cdot v_i$ can be considered to pose the associated LP problem

$$\begin{aligned} \text{Minimize} \quad & f(v) \\ \text{subject to} \quad & S \cdot v = 0 \\ & v_i \geq 0 \qquad \forall r_i \in Irr \end{aligned} \tag{4}$$

The obtained solution is a network mode. Using different additional constraints and functions, different scenarios in the metabolic network can be emulated (see [12]).

 doi:10.20944/preprints202208.0059.v1

4.2. Staphylococcus epidermidis RP62A model

The original model of *Staphylococcus epidermidis* RP62A was implemented by Díaz Calvo *et al.* [6] in the Systems Biology Markup Language (SBML[1]), a free and open data format for computational systems biology used by thousands of people worldwide. The authors used the ScrumPy metabolic modeling package, a widely used framework for metabolic network analysis.

The GSMM of *S.epidermidis* RP62A consists of 893 reactions plus 97 transport reactions and 864 internal metabolites, and 74 external metabolites. The curated model has mass and conservation energy balanced. Biomass generation is represented by pseudo-transporters (exchange reactions) to simulate the export of biomass precursors. There is no gene information in the SBML file provided by the creators.

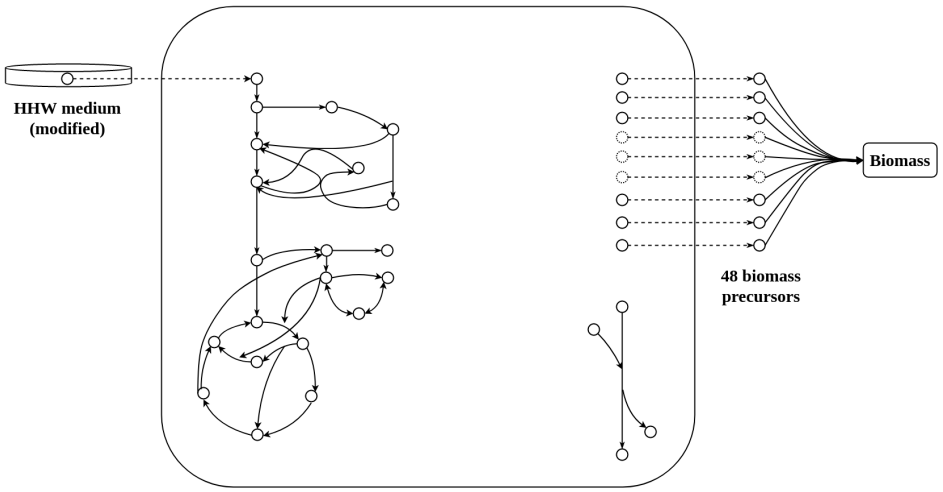

Figure 1. Model Scheme
Simplification of the model for *Staphylococcus epidermidis* RP62A. On the left, the modified HHW culture medium (Hussain, Hastings, and White) [14] which represents the uptake of nutrients incorporated by the metabolic network via pseudo-transporters (spotted arrow); on the center, the internal metabolic network is represented by a big square with the principal pathways on the left and the ATPase reaction on the right (represents ATP maintenance cost), on the right, the representation of the 48 biomass precursors declared as external metabolites exiting the metabolic network via pseudo-transporters. In this image, all biomass precursors have been gathered in a biomass lumped reaction.

Biomass composition. The model simulates biomass production ($mmolgDCW^{-1}$) through 48 precursors that are declared as external metabolites by using transport pseudo-reactions. Biomass consists of biofilm-forming precursors and planktonic cell components such as amino acid residues, deoxynucleotides, nucleotide triphosphate, cell wall/membrane components, and soluble metabolites. The fluxes of the reactions associated with these precursors were taken experimentally in continuous culture at a steady state. A complete explanation of the biomass composition can be found in Supplementary Material from [6].

Denote by $R_B \subset R$ the subset formed by those transport pseudo-reactions. For each reaction, $r_{i_j} \in R_B$, a constant b_{i_j} has been experimentally obtained so that any vector flux $v \in C$ must fulfill

$$v_{i_j} = b_{i_j}, \forall r_{i_j} \in R_B$$

Culture medium. A set of nutrients represents the medium made up of ions, amino acids, and secondary metabolites that are incorporated into the metabolic network through transport reactions and is measured in $gDCW^{-1}$ units [15]. This is the HHW medium described by Hussain *et al.* [14]

[1] https://sbml.org/

which has been modified by Diaz Calvo *et al.* [6] adding asparagine and eliminating unnecessary components of the medium. This medium contains all amino acids except glutamine [6].

Denoting by R_M the subset of these transport reactions and by m_{k_l} the maximum value of the transport reaction m_{k_l} (provided in the paper mentioned above), the following additional constraint for the model is obtained.

$$0 \leq v_{k_l} \leq m_{k_l}, \ \forall r_{k_l} \in R_M$$

ATP maintenance cost and Specific Growth Rate. Measures of growth (Y_{ATP}) and non-growth ATP (m_{ATP}) maintenance costs growth-associated ATP maintenance cost (60 $mmolgDCW^{-1}$) and non-growth-associated ATP maintenance cost (8 $mmolgDCW^{-1}h^{-1}$), respectively were taken by experimental procedures.

In this model, an ATPase reaction, r_{ATPAse}, is introduced, and the flux through this reaction must fulfill

$$v_{ATPAse} = A = Y_{ATP} \cdot \mu + m_{ATP}$$

The value μ has been experimentally calculated under different medium compositions. This parameter is used to describe the dynamic behavior of microorganisms, and it is measured in $gDCW^{-1}h^{-1}$. The specific growth rate (μ) must be measured in cell growth kinetics experiments under continuous-medium conditions. When growth rates are obtained experimentally, it may be provided in h^{-1} units [15].

The parameter values were taken experimentally in wet lab conditions by Díaz-Calvo and collaborators.

An LP program is formulated to obtain the optimal total flux as a minimization problem. Due to the presence of reversible reactions, to obtain this minimal flux state, the function $f(v) = \sum_{i=1}^{n} |v_i|$ is minimized taking into account the additional constraints previously indicated.

So, the optimization problem can be stated as follows:

$$\begin{aligned} \text{Minimize} \quad & f(v) = \sum_{i=1}^{n} |v_i| \\ \text{subject to} \quad & S \cdot v = 0 \\ & 0 \leq v_i \ , \ \forall r_i \in Irr \\ & v_{i_j} = b_{i_j} \ , \ \forall r_{i_j} \in R_B \\ & 0 \leq v_{k_l} \leq m_{k_l} \ , \forall r_{k_l} \in R_M \\ & v_{ATPase} = A \end{aligned} \tag{5}$$

Observe that, in this problem, the function f to be minimized is not linear. But, if all the reactions are irreversible, then f can be rewritten as $f(v) = \sum_{i=1}^{n} v_i$. Again, and due to the first two constraints, each solution to this FBA problem is particularly a network mode.

4.3. Our experimental approach

In our approach, the analysis of the model has been performed using the COBRApy (COnstraint-Based Reconstruction and Analysis in Python) package [16] working within Jupyter-Notebook [17] and Python3.6 as the kernel [18]. For solving the associated LP problems, we have used Cplex version 12.10 [19].

All the Python computer programs have been run in the Gacop's Cluster formed by several x86-64 computing nodes connected with an internal Gigabit Ethernet Network running Centos GNU/Linux

8.2. Cluster support has been provided by the Research Group of the High-Performance Computer Architecture (GACOP) of the University of Murcia (Spain). After finishing the importation of the model by COBRApy, an analysis of the structural properties was performed.

The provided *S.epidermidis* model has been successfully imported in COBRApy after solving a few issues (a detailed explanation of the importing process is provided in the Supplementary Material). It should be noted that the advice of the former research group was of great help when the first experiments with the model began. However, a few difficulties have been found while dealing with the calculations performed during that work.

The GSMM of *S.epidermidis* RP62A consists of 893 reactions plus 97 transport reactions and 864 internal metabolites, and 74 external metabolites. After finishing the importation of the model by COBRApy, an analysis of the structural properties has been performed.

The model contains 74 metabolites that are declared as external. There also are another 277 metabolites (dead-end metabolites) that behave as external ones in the sense of having no reaction producing or consuming them. These metabolites can be considered 'blocked,' and their presence in any reaction implies that these reactions are blocked. The average percentage of blocked metabolites in other models have been computed by checking curated models of a similar size taken from the BIGG database [20]. The average amount of blocked metabolites in these models is around 40-50%, which fits with the proportion of our case model ($\approx$ 30%). An analysis of the number of blocked metabolites in this model compared with several other models from BIGGs is available in the Supplementary Material.

Regarding the directionality of reactions, there are 405 reversible reactions and 585 irreversible ones. An unfolding of the reversible reactions of the model (substituting each reversible reaction with two irreversible ones that represent the two possible directions) was done to avoid the appearance of absolute values while calculating modes of minimal total flux. This unfolding step turns our optimization problem posed in Equation 5 into a linear one so we can use standard LP methods. After unfolding the metabolic network, the new model has 864 metabolites and 759 reactions (all irreversible). The original and decoupled SBML models can be downloaded from https://github.com/biogacop/SEpidermidis_Analysis.

We have also explored two additional ways to solve the original optimization problem (see Supplementary Material for details):

- Mixed-integer linear programming (MILP). Many solutions are obtained depending on the parameter that controls the optimization procedure.
- Formulating a new optimization problem by creating a lumped biomass pseudo-reaction that gathers all its precursors in a simplified way. The modified model can also be downloaded from https://github.com/biogacop/SEpidermidis_Analysis

Finally, we have examined some structural properties of the model. We have obtained the model's percentage ratio of the dead-end metabolites (29.53%). This value can be considered quite reasonable compared to models of similar size from the BIGG database [20], which is known for having carefully crafted and curated models. This comparison can be found in Supplementary Material.

After fixing the minimal total flux as an additional constraint, we have also examined the possible oscillations of the concentration values of the culture medium's component under the minimal total flux condition. These variations are as expected: they range between 0 and a value corresponding to the maximum values for the corresponding exchange reactions. The exception is found in a few amino acids (isoleucine, leucine, lysine, methionine, tyrosine, and phenylalanine) and niacin (vitamin B3), with significant differences between the minimum and maximum values that are incorporated into the metabolic network and those obtained in our analysis. Isoleucine, leucine, tyrosine, and phenylalanine are halved, lysine is doubled, methionine is increased by one order of magnitude, and niacin is increased by two orders of magnitude (A table with all those oscillation values can be found in Supplementary Material).

Author Contributions: G.S-C. has participated in the idea conception, code implementation, evaluation of the results, and the article writing. F.G. has participated in the idea conception, code implementation, mathematical development, and the article revision. J.M.G. has participated in the idea conception, discussion of results, and the article revision. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partially funded by the AEI (State Research Agency, Spain) and the ERDF (European Regional Development Fund, EU) under the Contract RTI2018-098156-B-C53, (MCI/AEI/FEDER, UE).

Conflicts of Interest: The authors declare no conflicts of interest.

Acknowledgments.: The authors are indebted to Dr. Dipali Singh for her continuous support of this work and for her early replies to the doubts we had trying to reproduce her results.

Availability of Data Materials: Software is freely available at https://github.com/biogacop/SEpidermidis_Analysis.

1. Volkova, S.; Matos, M.R.A.; Mattanovich, M.; de Mas, I.M. Metabolic Modelling as a Framework for Metabolomics Data Integration and Analysis. *Metabolites* **2020**, *10*, 303. Place: Basel Publisher: Mdpi WOS:000564100700001, doi:10.3390/metabo10080303.
2. Gu, C.; Kim, G.B.; Kim, W.J.; Kim, H.U.; Lee, S.Y. Current status and applications of genome-scale metabolic models. *Genome Biology* **2019**, *20*, 121. doi:10.1186/s13059-019-1730-3.
3. Moulin, C.; Tournier, L.; Peres, S. Combining Kinetic and Constraint-Based Modelling to Better Understand Metabolism Dynamics. *Processes* **2021**, *9*, 1701.
4. Nielsen, J. Systems Biology of Metabolism, 2017. Archive Location: world Publisher: Annual Reviews, doi:10.1146/annurev-biochem-061516-044757.
5. Schneider, P.; von Kamp, A.; Klamt, S. An extended and generalized framework for the calculation of metabolic intervention strategies based on minimal cut sets. *Plos Computational Biology* **2020**, *16*, e1008110. Place: San Francisco Publisher: Public Library Science WOS:000558078100061, doi:10.1371/journal.pcbi.1008110.
6. Díaz Calvo, T.; Tejera, N.; McNamara, I.; Langridge, G.C.; Wain, J.; Poolman, M.; Singh, D. Genome-Scale Metabolic Modelling Approach to Understand the Metabolism of the Opportunistic Human Pathogen Staphylococcus epidermidis RP62A. *Metabolites* **2022**, *12*, 136. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, doi:10.3390/metabo12020136.
7. Diaz Calvo, T. Investigating the metabolism of non-aureus staphylococci relevant to prosthetic joint infection. PhD thesis, University of East Anglia, 2020.
8. Berge, C. *Hypergraphs: combinatorics of finite sets*; Vol. 45, Elsevier, 1984.
9. Meretoudi, A.; Banti, C.N.; Siafarika, P.; Kalampounias, A.G.; Hadjikakou, S.K. Tetracycline Water Soluble Formulations with Enhanced Antimicrobial Activity. *Antibiotics* **2020**, *9*, 845. Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, doi:10.3390/antibiotics9120845.
10. Fael, H.; Demirel, A.L. Nisin/polyanion layer-by-layer films exhibiting different mechanisms in antimicrobial efficacy. *RSC Advances* **2020**, *10*, 10329–10337. Publisher: The Royal Society of Chemistry, doi:10.1039/C9RA10135G.
11. Hädicke, O.; Klamt, S. Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metabolic Engineering* **2011**, *13*, 204–213. doi:10.1016/j.ymben.2010.12.004.
12. Guil, F.; Hidalgo, J.F.; García, J.M. Flux Coupling and the Objective Functions' Length in EFMs. *Metabolites* **2020**, *10*. doi:10.3390/metabo10120489.
13. Klamt, S.; Gilles, E.D. Minimal cut sets in biochemical reaction networks. *Bioinformatics* **2004**, *20*, 226–234.
14. Hussain, M.; Hastings, J.G.; White, P.J. A chemically defined medium for slime production by coagulase-negative staphylococci. *Journal of Medical Microbiology* **1991**, *34*, 143–147. doi:10.1099/00222615-34-3-143.
15. Feierabend, M.; Renz, A.; Zelle, E.; Nöh, K.; Wiechert, W.; Dräger, A. High-Quality Genome-Scale Reconstruction of Corynebacterium glutamicum ATCC 13032. *Frontiers in Microbiology* **2021**, *12*, 750206. doi:10.3389/fmicb.2021.750206.
16. Ebrahim, A.; Lerman, J.A.; Palsson, B.O.; Hyduke, D.R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology* **2013**, *7*, 74. doi:10.1186/1752-0509-7-74.

17. Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; Ivanov, P.; Avila, D.; Abdalla, S.; Willing, C. Jupyter Notebooks – a publishing format for reproducible computational workflows. Positioning and Power in Academic Publishing: Players, Agents and Agendas; Loizides, F.; Schmidt, B., Eds. IOS Press, 2016, pp. 87 – 90.
18. Van Rossum, G.; Drake, F.L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, 2009.
19. Cplex, I.I. V12. 1: User's Manual for CPLEX. *International Business Machines Corporation* **2009**, *46*, 157.
20. King, Z.A.; Lu, J.; Dräger, A.; Miller, P.; Federowicz, S.; Lerman, J.A.; Ebrahim, A.; Palsson, B.O.; Lewis, N.E. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* **2015**, *44*, D515–D522, [https://academic.oup.com/nar/article-pdf/44/D1/D515/16661243/gkv1049.pdf]. doi:10.1093/nar/gkv1049.