UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA

E(xtraterrestris). Coli: Adapting genome-scale metabolic models to non-standard thermodynamical constraints

Lucas da Costa Monteiro

Mestrado em Bioinformática e Biologia Computacional

Versão Provisória

Dissertação orientada por:
Professor Doutor Francisco Rodrigues Pinto
Professor Doutor André Moitinho de Almeida

2024

Be fruitful and increase in number; fill the earth and subdue it

Genesis 1:28

Agradecimentos

بِسْمِ اللهِ الرَّحْمَنِ الرَّحِيمِ
الْحَمْدُ لِلهِ رَبِّ الْعَالَمِينَ

Em nome da Una-divindade, A eterna-compaixão, O todo-misericordioso,
Todos os louvores e graças devidos a Ti, soberano dos mundos.

Muitas são as oportunidades para Te agradecer,
mas raras as que tomamos devidamente.

Agradeço-Te a família,
Mãe, Pai, Avó, Irmã,
Aqueles que nutrem o reino.

Agradeço-Te a amizade,
Massas, Nuno, Bernardo, Rahim,
Aqueles que tornam os anos breves.

Agradeço-Te o encontro,
Pais, Pedro, Daniela, Simões,
Aqueles que partilham a sublime profundeza.

Agradeço-Te o professor André que impulsionou este prezado projeto,
o santo orientador Francisco que o suportou durante 3 semestres,
e o lab esp(a)cial que o albergou com um amor desproporcional.

Agradeço-Te pelas leis incompreensíveis,
e por todos aqueles que rumam sentido à última certidão.
Jó perguntou porquê, e só Tu respondeste.

São estes que vivem no meu coração sem permissão.
Por isso Te peço que lhes agradeças por mim,
pois só Tu poderás conceder o amor que lhes desejo.

آمين

Abstract

With the growing interest in space exploration, ethical issues arise. The sudden human invasion of pristine extraterrestrial habitats could potentially destroy sensitive life forms, resulting in an ecological disaster. The Committee on Space Research issues goals for interplanetary missions with direct interest in chemical evolution such as biological load reduction, but these are costly and time-consuming. An accessible method to test the contamination potential involves genome-scale metabolic reconstructions. These have shown the ability to simulate and predict bacterial growth effectively and economically on different substrates. However, these models are based on optimal growth conditions, which are unlikely to be found in extraterrestrial environmental scenarios.

The project aims to adapt these models to a variety of non-optimal thermal conditions to analyze how the metabolism of potentially contaminating organisms react in extraterrestrial environments and predict contamination risk. The adapted model should be simple and quick to analyse with the provided tools. An *Escherichia coli* metabolic model is used as the project's backbone due to its simplicity, well-studied nature, and its role as a contaminant on Earth, easily utilizing humans as dissemination vectors.

Several methods using a mass action kinetics simplification were tested, involving thermodynamical data and kinetic parameters estimation to simulate metabolic network flows. However, data inconsistencies and unrealistic results highlighted the importance of a reliable database. An extended genetic algorithm was developed to address incoherence, estimating kinetic parameters that were used to simulate cellular response across temperature ranges from 10K to 800K. The analysis revealed that cellular osmolarity remains stable across temperatures, but kinetic constants vary significantly. A stress function was developed to quantify how much a cell needs to change to meet viability requirements, showing significant metabolic stress between 10K and 320K, and a sudden rise in stress levels above 620K. Due to inconsistency with observations, no final conclusions about potential extraplanetary survival were drawn.

Keywords: Interplanetary Contamination, Genome-scale Metabolic Models, *Escherichia coli*, Genetic Algorithm, Stress Function

Resumo

Com o recente crescimento no interesse da exploração e colonização espacial como a nova vaga de exploração lunar ou até a colonização de Marte por empreendimentos privados, levantam-se questões éticas quanto às suas implicações e responsabilidades. A repentina invasão humana sobre os imaculados habitats extraterrestres poderá provocar contaminação indesejada potencialmente destruindo as suas sensíveis formas de vida. A destruição deste património genético e funcional provar-se-ia um desastre ecológico e científico de escala planetária. Como podemos prevenir esta catástrofe? Que meios podemos oferecer? O Comité de Pesquisa Espacial requer que certas metas sejam cumpridas em missões interplanetárias de modo a evitar uma potencial contaminação direta. As suas recomendações dependem do tipo de missão e do corpo celestial explorado. Incluem montagem em sala limpa, redução da carga biológica, e métodos de esterilização. Estes protocolos provam-se bastante dispendiosos e demorados, podendo até comprometer a praticidade das missões. Podemos desenvolver um método acessível para testar o poder de contaminação de um potencial contaminador? Reconstruções metabólicas à escala genómica mostram-se atualmente capazes de simular e prever o desenvolvimento de bactérias de maneira eficaz, significante e módica. Experiências demonstram que o alastramento destes seres unicelulares é possível e preciso sobre diferentes substratos, conseguindo até prever como o ser se desenvolveria num novo substrato. Este tipo de modelo tem um valor crescente em engenharia genética industrial e na vertente da biologia sintética. No entanto, os seus resultados são relativamente ao crescimento sobre condições ótimas, as quais não serão encontradas noutros cenários ambientais, nomeadamente noutros planetas. Neste projeto pretende-se assim testar a adaptação deste tipo de modelos a grande variedade de condições térmicas não ótimas de modo a analisar como as reconstituições metabólicas de seres potencialmente contaminantes reagem nos ambientes extraterrestres e então prever o risco de contaminação. A adaptação do modelo por um terceiro deverá ser simples e rápida, dispondo de ferramentas de análise para auxiliar os seus próprios propósitos. No nosso caso, um modelo metabólico de escala de *Escherichia coli* será utilizado como *backbone* do projeto pelo fato de se tratar de um modelo simples, bem estudado, com os dados alvo melhor definidos e disponíveis, pois é um elemento contaminante na Terra, e utiliza o humano como vetor de transmissão.

Neste trabalho foram testados vários métodos que utilizaram uma aproximação da cinética enzimática a uma lei de ação de massas. Esta abordagem permite simplificar o modelo, resumindo os parâmetros enzimáticos em apenas um parâmetro cinético e a constante de equilíbrio da reação. O parâmetro cinético engloba características intrínsecas à reação como afinidade catalítica e velocidade enquanto que a constante

de equilíbrio considera a influência da temperatura na termodinâmica da reação diretamente e através da variação da energia livre de Gibbs. Esta última serve então como a ponte que necessitamos para adaptar o modelo a diferentes condições térmicas, atuando nas propriedades termoquímicas da reação, e pode ser estimada sabendo a variação de entropia e entalpia da reação e a temperatura a que ela ocorre. Uma vez que estes dados estão apenas parcialmente disponíveis, tivemos de utilizar aproximações que envolveram uma calculadora de termodinâmica bioquímica - eQuilibrator - e dados de uma investigação científica de modo a continuar o projeto. Por outro lado, os parâmetros enzimáticos não estão disponíveis, tampouco o parâmetro cinético que utilizamos na nossa aproximação. Os métodos desenvolvidos servem então para estimar os parâmetros cinéticos das reações e simular os respetivos fluxos da rede metabólica.

Primeiramente tentámos estimar este parâmetro algebricamente, revertendo a equação de lei de ação de massas. No entanto, este método provou-se inviável estimando valores negativos ou seja irrealistas. O problema provavelmente advém da incoerência dos nossos dados. Cada concentração é estimada através da análise de diferentes experiências com diferentes condições e protocolos laboratoriais, levando a uma incongruência experimental. O metabolismo dos organismos estudados é sensível a pequenas alterações nas condições daí não podermos esperar que os dados sejam correlacionados. Para resolver este problema implementamos um algoritmo genético que visa minimizar esta incoerência estimando um vetor de concentrações de metabolitos coesas na rede metabólica em estudo. Este algoritmo utiliza um parâmetro de *fitness* que aproxima as estimativas a uma amostra de fluxos matematicamente coerente obtido de uma análise direta do modelo metabólico. No entanto, os resultados mantiveram-se irrealistas mostrando a importância de uma boa base de dados para os estudos pretendidos.

Dada a ausência de dados, decidimos estender o algoritmo genético para os parâmetros cinéticos. Este mantém sempre uma relação com a constante de equilíbrio da reação. Por isso, para evitar que o algoritmo se enviesasse para constantes de equilíbrio absurdas, foram definidos valores limite a partir dos quais as reações passariam a ser irreversíveis. Os resultados para este método foram favoráveis e por isso este foi o método selecionado para analisar os dados de *Escherichia coli* numa gama diversa de temperaturas. Paralelamente, foi investigado ainda um terceiro método que tira proveito de uma construção matemática que emerge do nosso modelo se considerarmos que as concentrações de metabolito não são significativamente alteradas pela temperatura. No entanto, este método apenas funcionou para a temperatura ambiente, insinuando que a aproximação não é válida numa gama térmica maior.

Para analisar um modelo metabólico real *e_coli_core*, começámos por importar-lo e curá-lo para os nossos objetivos. Considerou-se apenas o metabolito central uma vez que apenas para estas reações temos acesso a dados termoquímicos e sabendo que a dinâmica enzimática é substancialmente diferente para outras reações como de transporte e intercâmbio com o ambiente. Posteriormente, simulou-se o desempenho do modelo para uma gama de temperaturas entre os 10K e os 800K, utilizando perto de 7000 amostras da análise de fluxo para o algoritmo genético. Destas, 1225 foram consideradas plausíveis e devidamente analisadas. Foi descoberto que a osmolaridade da célula mantêm-se relativamente estável para todas as temperaturas, descartando problemas de pressão osmótica. Por outro lado, as constantes

cinéticas variavam bastante com a temperatura. Foi então desenvolvida uma função de stresse que compara as constantes a uma dada temperatura com as constantes à temperatura ambiente, ao mesmo tempo que considera o peso da reação em causa na produção de biomassa. Assim, esta função serve como um quantificador de quanto a célula tem de se alterar de modo a corresponder aos requerimentos celulares, quer seja em recursos ou eficácia enzimática. Verificou-se um elevado stresse metabólico entre 10K e 320K, com uma descida abrupta para níveis basais até 620K, e uma subida repentina até 800K. Uma análise mais profunda mostra que estes incrementos em stresse devem-se à irreversibilidade da aconitase a baixas temperaturas e da desidrogenase do gliceraldeído 3-fosfato a altas temperaturas. De notar que ambas as enzimas catalisam reações chave nos ciclos de obtenção de energia celular.

Em princípio, esperar-se-ia que os valores da função de stresse à temperatura ambiente de 298K fossem basais. No entanto, não é isso que ocorre, realçando a natureza qualitativa da função em relação a um modo ótimo e a impossibilidade de fazer uma análise de sobrevivência noutros planetas. Contudo, é de notar que a temperatura ótima para a proliferação de *E. coli* é ligeiramente superior à ambiente, onde o nosso limite se estabelece. Poderá ser que estes microrganismos se desenvolvem a temperaturas menores com uma maior carga enzimática? O limite superior a 620K é bastante superior às mais altas temperaturas registadas suportadas por formas de vida extremas de cerca de 390K. No entanto, note-se que apenas parâmetros termodinâmicos estão a ser considerados e não mecanismos mais termo-sensíveis, como a estrutura de proteínas. Sabendo que os resultados são qualitativos e não permitem uma conclusão sobre a sobrevivência de organismos noutros planetas , apenas uma análise grosseira foi realizada sobre as temperaturas superficiais. Toda a gama de temperaturas marcianas indicam uma maior carga por parte do metabolismo, enquanto certas regiões da Lua apresentam temperaturas ótimas para o seu desenvolvimento, embora com alguma periodicidade.

Em resumo, foi desenvolvido um método que adapta modelos metabólicos de escala para qualquer temperatura utilizando um algoritmo genético e uma função de stresse que quantifica o nível de adaptação ao novo ambiente. Para isso, dados experimentais das concentrações de metabolito e dos parâmetros termoquímicos devem ser importados. Estes dados refletem a maior limitação do método pois a sua aquisição pode-se provar desafiante devido a incongruências e recursos. O aparente desempenho ótimo do organismo a altas temperaturas relembra-nos do fato que os resultados não são conclusivos. Uma célula é um mecanismo frágil, sensível, e altamente complexo. Logo, para uma previsão plausível da sua capacidade de adaptação a ambientes inóspitos, outros fatores para além da temperatura devem ser tomados em consideração. Para permitir essa exploração, deverão estar disponíveis dados e recursos adicionais. Não obstante, o presente trabalho permite abrir caminho para o desenvolvimento de métodos mais complexos e precisos, e desvendar o comportamento de organismos em ambientes extremos. Estes métodos podem ser cruciais para o avanço do conhecimento na vertente da engenharia genética e constituir uma ajuda fundamental e inovadora nos mais recentes empreendimentos espaciais.

Palavras Chave: Contaminação Interplanetária, Modelos de Escala Genómica, *Escherichia coli*, Algoritmo Genético, Função de Stresse

Contents

List of Figures

List of Tables

Acronyms

COSPAR	Committee on Space Research
GEM	Genome-scale Metabolic Model
FBA	Flux Balance Analysis
FVA	Flux Variability Analysis
ECMDB	E. Coli Metabolome Database
GAPD	Glyceraldehyde-3-phosphate Dehydrogenase
NAD+	Nicotinamide Adenine Dinucleotide
ATP	Adenosine Triphosphate

Symbols

f	Flux Rate
K_m	Michaelis Constant
V_{max}	Flux Maximum Velocity
$k^{\pm}$	(Forward/Reverse) Flux Rate
$K^{\ominus}$	Thermodynamic Equilibrium Coefficient
$\Delta G^{\ominus}$	Standard Gibbs Free Energy Variation
$\Delta H^{\ominus}$	Standard Enthalpy Variation
$\Delta S^{\ominus}$	Standard Entropy Variation
ϕ	Standard Gibbs Free Energy Variation and Standard Enthalpy Variation Ratio
R	Universal Gas Constant
T_0	Ambient Temperature
K	Degrees Kelvin
Osm	Osmole
M	Molar Concentration

Chapter 1

Introduction

1.1 Motivation

Biological contamination played a critical role in the evolution of mankind. It spurred technological advancement in the development of medicine and hygiene, design and planning of sanitation systems and even warfare. Life forms spread rapidly into any available habitat [Nisbet and Sleep, 2001], adapting to its particular requirements and continuously differentiating to create rich ecosystems. Somewhere in Earth's past, life originated, filling all niches from extreme conditions deep undergound [Heuer et al., 2020], to the highest mountaintops [Coleman, 2022]; from hot arid deserts to ice-cold dark cavernous systems [Orellana et al., 2018]. Estimates say that we breathe in between 1 and 10 fungal spores every time we inhale [Fröhlich-Nowoisky et al., 2009]. Would Earth's life characteristic expansibility apply to other worlds?

Interplanetary contamination is accounted for in the planning of space missions by the United Nations guidelines for planetary protection. Research lines on the ethical implications and responsibility of space exploration have been conducted [Rummel et al., 2012], some implying total protection of alien species, ecosystems, and even lifeless environments. Others, defend the value of extraterrestrial life as a need to ensure the scientific integrity of the discovery of life since they could represent totally new biology or similar biology that has developed in a totally different environment [Persson, 2012]. How can we avoid an ecological and scientific disaster? How will the upcoming human space exploration missions affect the native environments? Plans are well underway, with the recent success of Artemis I [Tribou, 2022], and there could be humans footprints on lunar and martian soil by the end of this decade [Harvey and Mann, 2022; Torchinsky, 2022].

Forward contamination is the transfer of life and other forms of contamination from Earth to another celestial body and current space missions are governed by the Outer Space Treaty and the Committee on Space Research (COSPAR) guidelines for planetary protection. They categorize the mission in five

categories depending on the risk of contamination. Direct interest locations for chemical evolution or the origin of life requirements include trajectory biasing, clean room assembly, bioburden reduction, spacecraft sterilization and if impact is a possibility, inventory of organics [OSMA, 2024]. All these measures require extensive resources compromising the mission objectives. However, little literature is found on bio-contamination experiments or modeling applied to other planets. Studies like this could help space agencies to more efficiently protect pristine alien environments while searching for faint signs of life.

1.2 Problem

How would biologically contaminant organisms behave in an alien environment? Would they be capable of survival or could even thrive in foreign habitats? How can this behaviour be predicted? *In-situ* experiments, in addition of being expensive and requiring high amount of preparation to accomplish such missions, could jeopardize the studied environment. *In-vitro* experiments are also expensive and extraterrestrial conditions can be challenging to recreate in the laboratory. That leaves the option of *in-silico* experiments. Inexpensive and practical, computational simulation may provide a good first approach to the interplanetary contamination problem.

Tools, like genome-scale metabolic models (GEMs), are useful to simulate how simple organisms produce biomass, necessary for their growth, reproduction, structure maintenance and adaptation to their environments. However, analysis of these models is often performed under *normal* conditions and does not consider the extreme environmental conditions that are expected on other celestial bodies.

1.3 Objectives

The main goal of this work is to create a method to constrain genome-scale metabolic models with non-standard temperature conditions of the environment. This method is then to be utilized to predict the survival of a microorganism on several realistic conditions found in other celestial bodies, namely the Moon and Mars, by estimating the reaction's kinetic parameters under said conditions and analyzing the stress that the organism must go through in order to adapt.

A secondary goal is to develop a pipeline that adapts already developed models for new temperature conditions in a user friendly way, by just inputting the necessary data and offer a range of analysing tools.

1.4 Significance

Studying cell proliferation in bodies of interest could help to assess (and ideally avoid) ethical and biological hazardous situations in the near future. However the adaptation of genome scale metabolic models into non-standard temperatures, and other environmental constraints, is also of great significance for in-world cases. GEMs in general are useful for practical understanding of the metabolism of microorganisms [Segrè et al., 2002]. Understating them in a wide range of conditions have implications for industry, especially in fermentation business, that is undergoing an increase in interest in matters such as engineered microorganisms. Adapted GEMs could discover unexpected conditions that yield better outputs. If a simple and effective pipeline for GEM thermodynamic adaptation is developed, i.e a pipeline that autonomously accepts a wide range of already developed models, the study of cellular metabolism in extreme conditions becomes possible as well as all systemic metabolic research benefits from the adapted models.

1.5 Thesis outline

To fulfill the proposed objectives, a case-study microorganism must be selected and its genome-scale metabolic model imported. Knowing the metabolic reactions, the respective kinetic parameters, that are dependant on the environmental conditions, are estimated and used to calculate the range of thermodynamically feasible reaction rates. These parameters are the connection between the environment and the microorganism, actuating directly on each metabolic reaction, dictating their direction and influence on the rewired metabolic network for the given abiotic factors. Therefore, it is pivotal to correctly estimate these parameters and understand how they depend on these factors.

For this the following thesis is organized in five chapters:

- Introduction introduces the research motivation, problem and objectives.
- Background clarifies previous development of the handled models and methods.
- Methods describes the developed methodology.
- Results and Discussion elucidates the obtained results and their analysis.
- Conclusion compiles the work and contributions, suggesting future work.

Chapter 2

Background

To simulate the behavior of a microorganism under a large spectrum of temperatures, a genome-scale metabolic model is a favorable choice since temperature influences directly the thermodynamics of biochemical reactions. A GEM is a network-based tool that collects all known metabolic information of a biological system, including the genes, enzymes, reactions, associated gene-protein-reaction rules, and metabolites [Passi et al., 2021]. So knowing the metabolism in detail, i.e the set of life-sustaining chemical reactions for a given organism, GEMs can be adapted to predict how these reactions and the system as a whole would behave under different temperatures, constrained by the underlying thermodynamic laws.

2.1 Flux Solution Space

A useful technique to analyze a GEM is to compute its flux solution space. GEMs formalize the metabolic network as a system of equations describing the concentration changes in time as the dot product of a matrix of the stoichiometric coefficients S and the vector of the unsolved fluxes $\vec{v}$. This matrix represents each compound in a row, while the columns correspond to the reactions. Each matrix element contains the stoichiometric coefficient of the row compound in the column reaction. The flux solution space calculation is obtained by solving the linear equation system $S \cdot \vec{v} = \vec{0}$, that equals the dot product of the stoichiometric matrix by the flux vector to a null vector, which is equivalent to consider that the metabolic model reaches a steady state. **The steady state is our condition for survival**, so there needs to exists at least one set of fluxes that achieve a steady state for our organism to survive. Empirically, organisms are shown to reach steady states when grown in stable environments [Segrè et al., 2002]. This is also mathematically proven since if a steady state isn't reached in a constant medium, there will be either metabolite depletion or accumulation, both harmful to the organism. The flux solution space calculations for an example microorganism are expatiated in Appendix A, as well as the underlying meanings of the model stoichiometric matrix and concentration vector and how they are reached upon. Note that this method is easily scalable so it can be used to analyse a wide range of metabolic network reconstructions.

2.1.1 Flux rate constraints

Usually, the flux solution space is multidimensional and containing infinite flux combinations, so that constraints are applied to rule out unrealistic solutions. It can be constrained by thermodynamics, experimental or growth media factors to better simulate real metabolism. Furthermore one or several objective functions can be defined to collapse the accurate but still infinite solution space into a single most probable solution of fluxes. The most common approach is Flux Balance Analysis (FBA), which uses a *biomass reaction* to find the steady state solutions that maximize biomass production. The biomass reaction, an artificial reaction that is added to most GEMs, simulates the reaction necessary to create a new copy of the organism. It consumes many compounds present in the metabolic network in adequate proportions to replicate the molecular composition of the organism's biomass. Therefore, a positive flux through this biomass reaction is our condition for organism proliferation. Experimental observations confirm that microorganisms, such as E. coli, maximize biomass production in their exponential growth phase under optimal conditions [Segrè et al., 2002]. Other functions can be considered, e.g energy production or metabolite production rate, depending on the case-study [Khannapho et al., 2008]. Simulations performed using FBA are computationally inexpensive and can calculate steady-state metabolic fluxes for large models (over 2000 reactions) in a few seconds on modern personal computers.

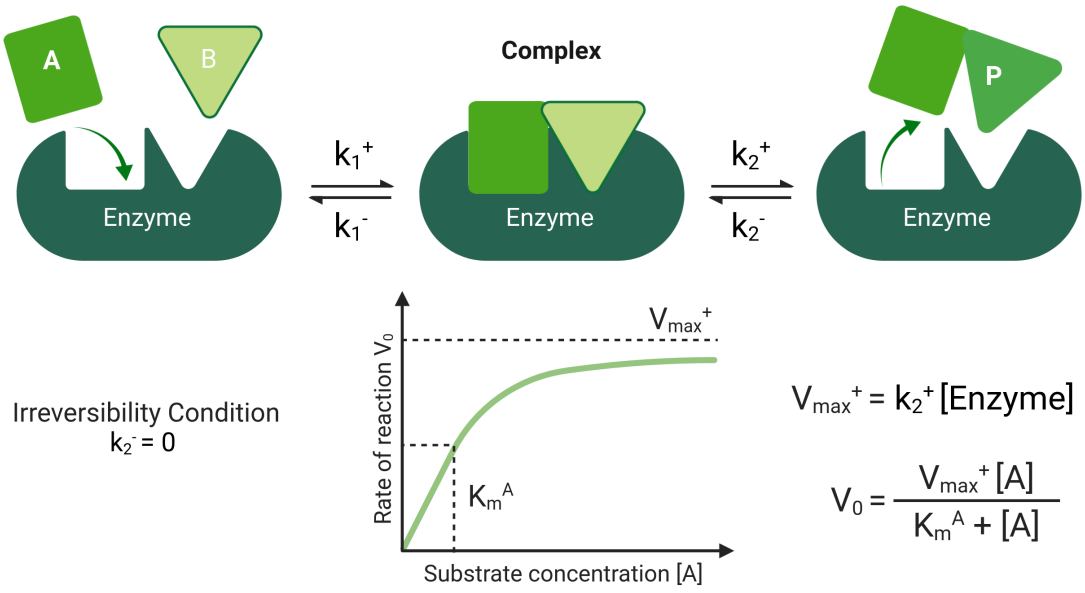


Figure 2.1: **Top:** The generalization of enzyme catalysis follows a two-step elementary reaction, described by laws of mass action. **Bottom:** If the reaction is irreversible in the forward direction, i.e the second step reverse rate constant is null, a Michaelis–Menten enzyme kinetics is met where reaction rate grows linearly at low substrate concentration and approaches a limiting rate (V^{+}_{max}) at saturation. The Michaelis constant (K_m) is defined as the substrate concentration at which reaction rate is half of V_{max}.

2.2 Enzyme kinetics

How the reaction flux rates depend on the kinetic parameters is dictated by enzyme kinetics. Most biological reactions are catalyzed by proteins, i.e enzymes. In the absence of enzymatic catalysis, most biochemical reactions are so slow that they would not occur under the mild conditions of temperature and pressure that are compatible with life. Enzymes accelerate the rates of such reactions by well over a million-fold, so reactions that would take years in the absence of catalysis can occur in fractions of seconds if catalyzed by the appropriate enzyme. Cells contain thousands of different enzymes, and their activities determine which of the many possible chemical reactions actually take place within the cell [Cooper, 2000].

Consider the following reaction catalyzed by enzyme E and kinetic parameters k_1 and k_2 for the forward reactions, and k_1^- and k_2^- for the reverse reactions:

$$\mathrm{E} + a\mathrm{A} + b\mathrm{B} \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} \mathrm{EA}_a\mathrm{B}_b \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{E} + p\mathrm{P} + q\mathrm{Q} \tag{R1}$$

In dilute solutions, elementary reactions (reactions with a single mechanistic step) empirically follow the law of mass action. This law states that reaction rates are proportional to the product of the concentrations of reactants, raised to the powers of their stoichiometric coefficients. Rate constants are the proportionality factors that intrinsically encode the energetic properties of the reaction. This principle is based on the assumptions of free (Fickian) diffusion and thermodynamically-driven random collision [Connors, 1990]. The rate of the reaction (f) in Reaction 1, assuming each step is elementary, can be described by:

$$f = \frac{\mathrm{d}\mathrm{P}}{\mathrm{d}t} = \frac{k_1^+ k_2^+ [\mathrm{A}]^a[\mathrm{B}]^b - k_1^- k_2^- [\mathrm{P}]^p[\mathrm{Q}]^q}{k_1^- + k_2^+ + k_1^+ [\mathrm{A}]^a[\mathrm{B}]^b + k_2^- [\mathrm{P}]^p[\mathrm{Q}]^q}[\mathrm{E}] \tag{2.1}$$

This reaction follows reversible Michaelis–Menten kinetics, represented in Figure 2.1. In this model, kinetic parameters are linked to the reaction's turnover number, indicating the maximum number of enzymatic reactions catalyzed per second, and Michaelis constants (K_m). The K_m value is experimentally determined as the concentration at which the enzyme reaction rate reaches half of its maximum velocity (V_{max}). Substituting with:

$$V_{max}^{\pm} = k_2^{\pm}[\mathrm{E}] \qquad K_m^+ = \frac{k_1^- - k_2^+}{k_1^+} \qquad K_m^- = \frac{k_1^- - k_2^+}{k_2^-}$$

an expression of the kinetic parameter dependent flux rate is achieved by Equation 2.2.

$$f = \left(\frac{V_{max}^+}{K_m^+}[\mathrm{A}]^a[\mathrm{B}]^b - \frac{V_{max}^-}{K_m^-}[\mathrm{P}]^p[\mathrm{Q}]^q\right) \cdot \frac{1}{1 + [\mathrm{A}]^a[\mathrm{B}]^b/K_m^+ + [\mathrm{P}]^p[\mathrm{Q}]^q/K_m^-} \tag{2.2}$$

Flux rates can be calculated by the reaction kinetic parameters, i.e maximum velocities, Michaelis constants, and concentration values. However due to research effort needed to estimate them experimentally, **kinetic parameters are not available** for many reactions let alone in different thermic conditions.

2.3 Reaction thermodynamics

At equilibrium, the rate of the reaction that consumes reagents into products is the same as that that consumes products into reagents. This occurs when:

$$v = 0 \quad \Leftrightarrow \quad \frac{V_{max}^{+} K_m^{-}}{V_{max}^{-} K_m^{+}} = \frac{[\mathrm{P}]_{eq}^{p}[\mathrm{Q}]_{eq}^{q}}{[\mathrm{A}]_{eq}^{a}[\mathrm{B}]_{eq}^{b}} = K^{\ominus} \tag{2.3}$$

showing that thermodynamics forces a relation between the values of the 4 rate constants by the thermodynamic equilibrium constant ($K^{\ominus}$) for the reaction. This constant is related to the standard Gibbs free energy change ($\Delta G^{\ominus}$) of reaction by Equation 2.4.

$$\Delta G^{\ominus} = -RT \ln K^{\ominus} \Rightarrow K^{\ominus} = \mathrm{e}^{-\frac{\Delta G^{\ominus}}{RT}} \tag{2.4}$$

where R is the universal gas constant and T the absolute temperature in Kelvin (K). External factors can therefore limit the ability of an enzyme to catalyse a reaction in both directions by altering the chemical reaction's kinetic parameters values, i.e f is positive if the reaction proceeds in the forward direction and negative otherwise. The standard Gibbs free energy variation also changes with temperature.

$$\Delta G^{\ominus}(T) = \Delta H^{\ominus}(T) - T\Delta S^{\ominus}(T) \tag{2.5}$$

$\Delta G^{\ominus}$ is dependant on the reaction's enthalpy variation ($\Delta H^{\ominus}$) and entropy variation ($\Delta S^{\ominus}$) [Smith et al., 2005]. The contribution of entropy variation is proportional to temperature. Note that enthalpy and entropy variations may also vary with temperature (Equation 2.5).

2.3.1 Group contribution method

The overall reaction energies can be calculated by the sum of the energies of the products minus the sum of the energies of the reactants. This is true for all Gibbs free, enthalpy and entropic energy changes. In the next equation there is the example for the Gibbs free energy for a reaction of type Reaction 2.

$$\Delta G_j^{\ominus} = \sum_{i=1}^{N} \Delta G^{\ominus}([\mathrm{X}_i]^{r_{ij}}) - \Delta G^{\ominus}([\mathrm{X}_i]^{p_{ij}}) \tag{2.6}$$

A similar analysis can be made to estimate enthalpic/entropic energies of the metabolites themselves. A group-contribution method, demonstrated in Figure 2.2, is used to predict metabolite energies thus reducing the number of needed data. Knowing the properties of the functional groups that compose the metabolite, a simple additive method can be used to estimate the overall metabolite properties. This simple form assumes that the property is strictly linearly dependent on the number of groups, and additionally no interaction between groups and molecules are assumed, which may lead to considerable error. For a rough estimation of thermodynamic properties in a research scenario where the number of metabolite species is high and available data is very limited and infeasible to obtain, this method is a valid approach.

However, the method is not the state of the art and others were considered for this research. One was the generation of the 3 dimensional conformation of the metabolites using the SMILES specification available in the ECMDB database with a quantum-mechanical calculation of minimal energy. This led to the problem that the conformation could not necessarily mean a global minima but a local low energy saddle point. To solve this, it's necessary to generate several configurations and evaluate the minimal. After the calculation the thermodynamical properties can be performed. However these estimations were deemed impractical due to long process time (days) for each one of the metabolites. Since the objective is to create a useful, fast pipeline for metabolism behaviour estimation under different temperatures, the quantum method was abandoned and a less accurate but faster, pragmatic group contribution preferred.

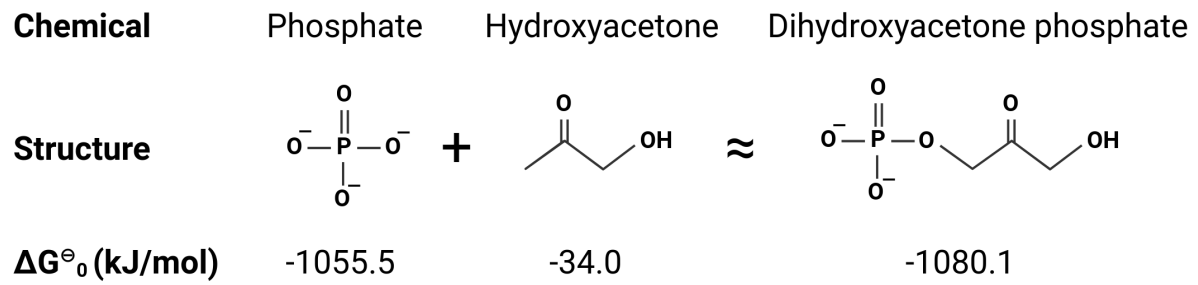


Figure 2.2: Group contribution method example for dihydroxyacetone phosphate (dhap_c in our model) composed of hydroxyacetone and phosphate functional chemical groups. A simple Gibbs energy variation estimation of the compound is accomplished with the sum of these chemical groups energies. The presented $\Delta G_0^{\ominus}$ were obtained by the eQuilibrator software.

2.4 Genetic algorithm

GEMs allow the estimation of steady state flux solutions, but rate constants and metabolite concentrations are not determined. Knowing the flux rate expressions and the relation between forward and reverse rate constants imposed by thermodynamic constants, one can estimate the concentrations and rate constants compatible with a given steady state flux solution. It is not possible to perform this estimation analytically, but numerical optimization algorithms can be employed to search for these values. Genetic algorithms commonly generate high-quality solutions to optimization and search problems by relying on biologically inspired operators such as, in our case, mutation, and selection [Mitchell, 1998]. It starts with a population of candidate solutions that iterate through the generations towards an objective function. For this, the fitness of each individual solution is determined, and the better fit ones are left to mutate and pass their genes (values) to the next generation. This generational process is repeated until a termination condition is reached. In our case, the termination condition means that the highest ranking solution's fitness reached a plateau such that successive iterations no longer produce better results.

Chapter 3

Methods

Our objective is to study the impact that temperature has on a microorganism's metabolism. First we need to select our model based on ambient temperature information, since its lacking for other temperatures. Then analyse how it behaves and adapts to new temperatures, namely its variables. Since we'll be studying the metabolism in the eye of the metabolic reaction's chemistry, the appropriate study variables are cellular metabolite concentrations and reaction's kinetic parameters. These are the variables that are ought to be estimated under normal temperatures in order to understand how they change under the new conditions, and if survival is possible.

3.1 The Model

What microorganism should be analysed? This is of importance since the choice must 1) have available metabolic data for analysis, 2) be a viable space contaminator. One would assume that extremophiles would be the most capable organisms at colonizing a rather extreme alien world. However, would they be the most capable at reaching such said world? The probability of them traveling to another planet would be far lesser than of an ordinary earth microorganism, simply by their relative scarcity. Also the conducted research on extremophiles lacks on metabolic data, mainly due to their living conditions, so the study could not take advantage of previous models, and would have to create a metabolic model from extremophile genetic sequences, something of interest and novelty but deviating from the original objective.

The study of the capability of an ordinary bacteria to colonize another astral body is of far more interest. By their abundance, these bacteria are much more likely to reach other worlds traveling in spacecraft or even in rocky fragments that eventually escape earth. Their models are already constructed, heavily studied and easily accessible. But are they viable contaminators? Could they compete with the much more versatile extremophiles in terms of alien survivability? Would they even survive such journeys to these worlds? Bacteria in space have been found to create aggregates to increase their survival rates. The outer cells, exposed to the hazardous space conditions, burn and die protecting the inner cells from

this conditions [Kawaguchi et al., 2020]. Besides that, the harsh conditions enhance the latent bacteria mutation rates meaning that a complex, potentially *extremophile*, society could form at another celestial body, even with *ordinary* bacterial ancestors. [Moeller et al., 2012]

So on what ordinary bacteria should we focus our efforts? The most common should be in our scope and also the ones most studied and with most reliable, sturdy models. Is it sufficient to be common on earth to easily reach other worlds? Surely they have a good chance of hiding in a spacecraft but reaching another planet they have one and one only chance of survival. What organisms will be constantly inputted into this alien environment? A lot of potentially contaminant organisms will hitchhike humans to other worlds. Microbes inhabit just about every part of the human body. These beings will be constantly released into the environment as long as humans are present and surviving. One type of microorganism that is specially of interest is gut microbiome bacteria. These are released everyday (hopefully) by humans in their waste and a lot of literature as been written on the subject due to their strong socioeconomic and general health value.

Extensive literature is written on the gram negative bacteria *Escherichia coli* and has available databases on genome-scale metabolic data, kinetic parameters and metabolite concentrations. Besides being an organism associated with contamination on earth, it is also common and naturally present in human microbiota.

3.1.1 Genome-scale Metabolic Model

The *e_coli_core* genome-scale metabolic model [Orth et al., 2010] from the BiGG Models database [King et al., 2015] was imported and handled using the `COBRApy` python package [Ebrahim et al., 2013]. It contains 72 metabolites and 95 reactions and was developed mainly for educational purposes, so it is a good model for the testing and development of our proposed approach. The model has enough reactions and pathways to enable interesting and insightful calculations, but it is also simple enough that the results of such calculations can be understood easily. [Orth et al., 2010].

In our study, only the cytoplasmic reactions are considered, leaving transport and exchange reactions unconstrained. This is because we are interested in how the central metabolism reacts to temperature change. Besides that, the thermodynamics and kinetics of transport reactions are wildly different and we don't possess the necessary means to estimate or simulate this. Also, the core metabolism isn't exclusive to *E. coli* and possibly shared with a wide range of microorganisms, making our research more inclusive and not necessarily focused on just one bacteria.

3.1.2 Reaction weight

Each reaction has a contribution *weight* to metabolism when it comes to survival. Some reactions play a crucial role in the central metabolism leading to a null maximal biomass reaction flux when shut down. This doesn't necessarily mean that the organism is dead but reproduction is impossible so it **cannot sustain contamination**. Others play secondary roles leading to less, or even equal but never null, maximal biomass reaction flux values when shutdown. Because of this, every reaction has a `weight` variable that we defined by the percentage of maximal flux of the biomass reaction that is lost due to the shutdown of the respective reaction. A weight of 1 implies an essential reaction to the biomass production of the organism and a weight of 0 implies that the reaction shut down doesn't influence cellular growth. To calculate the weight of each reaction, we use COBRApy's `single_reaction_deletion` function and analyse how much the flux declines.

3.1.3 Flux Equation

Temperature influences reaction equilibrium constants, which condition the ratio between kinetic rate constants. Therefore, to constrain reaction fluxes as a function of temperature, we need to express how fluxes depend on the rate constants. This is obtained through Equation 2.2. However data on reaction kinetic parameters and enzyme concentration lacks heavily. How can we address this fundamental issue in our research? By making the assumption that the system is always near a steady state, so that metabolite concentration variation is small, $\delta[\vec{\mathrm{X}}] \approx 0$, Reaction 1 behaves like an elementary reaction, and its reaction rate expression is well approximated by the law of mass action:

$$f = C\frac{V^+_{max}}{K^+_m}[\mathrm{A}]^a[\mathrm{B}]^b - C\frac{V^-_{max}}{K^-_m}[\mathrm{P}]^p[\mathrm{Q}]^q \ , \ C = \frac{1}{1 + [\mathrm{A}]^a[\mathrm{B}]^b/K^+_m + [\mathrm{P}]^p[\mathrm{Q}]^q/K^-_m} \tag{3.1}$$

Considering $C \approx$ constant, k^+_j and k^-_j constant are defined as an overall forward and reverse reaction rate constant, respectively, that considers enzymatic efficiency, abundance, and metabolite concentration: $k^+_j = CV^+_{max}/K^+_m$ and $k^-_j = CV^-_{max}/K^-_m$. These parameters provide the necessary bridge between the abiotic factors of the environment and the microorganism metabolic reactions.

The rate equation can then be generalized to apply it to every reaction in the model. First a generic form of the chemical equation of the j-th reaction is written.

$$r_{1j}\mathrm{X}_1 + r_{2j}\mathrm{X}_2 + ... + r_{Mj}\mathrm{X}_M \underset{k^-_j}{\overset{k^+_j}{\rightleftharpoons}} p_{1j}\mathrm{X}_1 + p_{2j}\mathrm{X}_2 + ... + p_{Mj}\mathrm{X}_M \tag{R2}$$

where j is the reaction index running from 1 to N, X_i the i-th chemical species, k^+_j and k^-_j the rate constants of the j-th reaction and r_{ij} and p_{ij} the stoichiometric coefficients of reactants and products, respectively, of a general description of a chemical reaction network that considers a number M of distinct

chemical species reacting via N reactions. The rate of a reaction can then be inferred by the law of mass action:

$$f_j = k_j^+ \prod_{i=1}^{M} [\mathrm{X}_i]^{r_{ij}} - k_j^- \prod_{i=1}^{M} [\mathrm{X}_i]^{p_{ij}} \tag{3.2}$$

Note that $k_j^+ = K_j^{\ominus} k_j^-$ by Equation 2.3. $K^{\ominus}$ is dependent on the reaction Gibbs free energy by Equation 2.4, an energy that can be estimated in order to proceed with the calculations. We therefore create Equation 3.3, **a simplified, temperature dependent flux rate equation that sums all the inaccessible kinetic parameters into a single kinetic constant and a calculable thermodynamic parameter**, the reaction equilibrium constant.

$$f_j(k_j^-, K_j^{\ominus}, [\vec{X}], T) = k_j^- K_j^{\ominus}(T) \prod_{i=1}^{M} [\mathrm{X}_i]^{r_{ij}} - k_j^- \prod_{i=1}^{M} [\mathrm{X}_i]^{p_{ij}} \tag{3.3}$$

Three parameters are required to estimate the organism's flux rate under the various temperatures: *rate constants*, *equilibrium constants*, and *metabolite concentrations*. However, there is no available database that provides all these values from one same condition. This is of rather importance since the metabolism is a dynamical therefore changing mechanism. If parts of data are collected in different conditions, they are by default independent and incoherent, contrary to what would be the case in a *real* system. To minimize error, **all data needs to be obtained at an ambient temperature of 298.15K and when the metabolic network is adapted for the maximal growth performance**, that is in the exponential growth phase [Segrè et al., 2002]. Methods to obtain the data for the three different parameters are necessary and the next three sections explain how the values are estimated.

3.2 Rate Constants

The estimation of reaction rate constants is key to understand how the organism reacts to non-standard temperatures, as previously stated, since it represents the enzymatic properties of the reaction and its values are influenced by temperature. Three methods were developed throughout this work due to the fact that, as work progressed, limitations were found and the way our problem was addressed had to change in order to meet the objectives. All approaches use Equation 3.3 as the underlying metabolic rate equation.

3.2.1 Algebraic Estimation

Flux distribution of the model under optimal conditions can be calculated using the same library which returns the optimal growth rate by setting the objective function as the model's `Biomass_Ecoli_core` maximum, as expatiated in subsection 2.1.1. Knowing this value as well as the equilibrium constant, metabolite concentration and stoichiometric coefficients, we can reverse the well established flux equation, in subsection 3.1.3, and algebraically calculate the rate constants. Fixing these constants for ambient

temperature, kinetic constants for new temperatures are calculated by multiplying by the respective reaction equilibrium constant at that temperature.

$$f_j = k_j^- K_j^{\ominus} \prod_{i=1}^{N} [\mathrm{X}_i]^{r_{ij}} - k_j^- \prod_{i=1}^{N} [\mathrm{X}_i]^{p_{ij}} \Rightarrow k_j^- = \frac{f_j}{K_j^{\ominus} \prod_i [\mathrm{X}_i]^{r_{ij}} - \prod_z [\mathrm{X}_i]^{p_{ij}}} \tag{3.4}$$

Equilibrium constants and metabolite concentrations are estimated by the methods explained at section 3.3 and section 3.4. Chemical species and reactions stoichiometric coefficients are directly imported from the model (refer to subsection 3.1.1). Fluxes calculated using FBA and biomass reaction maximization as the objective function. However, the flux distribution solution that maximizes biomass has null reaction rate values, yielding non-desired null rate constant values. Null values at ambient temperature will compromise the respective reaction in other temperatures since the calculation will always yield null values. This is a problem since reactions that aren't significant at ambient temperatures could eventually become crucial, while ambient significant ones shut down. To solve this problem, flux variability analysis (FVA) that returns the boundaries for the fluxes through each reaction that can, paired with the right combination of other fluxes, estimate solutions at a defined fraction of the optimal. Since biological systems are prone to variation, an 95% fraction of the optimum seams a valid approximation for the estimation of the reaction rate constants. The returned values define the interval of optimum flux for a given reaction. These can be automatically calculated using COBRApy `flux_variability_analysis` function.

After estimating the rate constants in a known temperature, their values are fixed and the fluxes depend on temperature by the equilibrium constant factor that translates reverse to forward ratio of rate constants. Several simulations are performed using different flux and concentration values from the estimated intervals. Even at ambient temperature a real metabolism can perform in different configurations, so the added noise in metabolite concentration and flux rates mimic these metabolic differences. In the case of metabolite concentrations the noise values must differ from a certain percentage of the original value, and the fluxes are randomly chosen from the calculated FVA flux range. This way, a viable flux range for a certain reaction at temperature T can be constructed and our flux solution space constrained. The solutions found within the constrained interval for every reaction are then viable steady state solutions for a metabolism at T.

3.2.2 Low Concentration Variation Approximation

The process described in the previous section can be condensed in a single mathematical representation. The "0" subscript is used to represent values at ambient temperatures and "T" subscripts values at a given T temperature.

$$f_{jT} = k_j^- \left[K_{jT}^{\ominus} \prod_{i=1}^{M} [\mathrm{X}_i]_T^{r_{ij}} - \prod_{i=1}^{M} [\mathrm{X}_i]_T^{p_{ij}} \right] = \frac{K_{jT}^{\ominus} \prod_i [\mathrm{X}_i]_T^{r_{ij}} - \prod_i [\mathrm{X}_i]_T^{p_{ij}}}{K_{j0}^{\ominus} \prod_i [\mathrm{X}_i]_0^{r_{ij}} - \prod_i [\mathrm{X}_i]_0^{p_{ij}}} f_{j0} \tag{3.5}$$

By our assumptions of how Gibbs free energy variation depends on temperature (refer to Appendix B), the reaction equilibrium constants acquire an interesting property where the constant at any temperature is equal to the constant at the standard temperature times the exponential of a function that depends on the given temperature, as in Equation B.7.

$$f_{jT} \sim e^{\phi_0 \left(\frac{1}{T_0} - \frac{1}{T}\right) \frac{\Delta G_e^{\ominus}}{R}} f_{j0} \tag{3.6}$$

If we assume that the concentrations at a given temperature are approximately equal to the concentrations at the standard temperature, i.e $[\mathrm{X}_i]_0 \sim [\mathrm{X}_i]_T$, we can simplify the equation in a way that the distribution flux stops depending on metabolite concentration values (Equation 3.6). This is of great use since the concentration values are easily corrupted (refer to Chapter 4) so simplifying our methods to not include this could mean coherent results and faster processing time. Like in the previous method, we algebraically calculate the flux range for the new temperature with simulations using the flux variability analysis range (note $f_{jT}(T_0) = f_{j0}$) and the flux solution space constrained based on the results.

3.2.3 Genetic Algorithm

We ultimately employed a genetic algorithm based method. The algorithm starts by selecting a sample from the solution space and setting a temperature. The metabolite concentrations are set for the case study and noise is added to it. After this the metabolite concentration productories for each reaction, i.e $\prod_i [\mathrm{X}_i]_T^{r_{ij}}$ and $\prod_i [\mathrm{X}_i]_T^{p_{ij}}$ are computed.

It continues by estimating the kinetic parameters. These are initially set to 1 and iterated to approximate to the value that minimizes the weighted euclidean distance to the sample from the flux sample space by Equation 3.2. These samples are just punctual solutions of the flux solution space calculated by the method explained at section 2.1 and are easily attainable since COBRApy has a integrated function `sample` that returns n random samples of the solution space. We start by generating all necessary samples *a priori* since the calculation requires some computational power. This way all samples are ready for the genetic algorithm and repeated samples are avoided.

The parameters are always related by the reaction's equilibrium constant. However, high $K^{\ominus}$ values skew the estimations. To solve this, if equilibrium constants surpass a threshold above 10^3 or below 10^{-3} the reaction is deemed as irreversible in the forward or reverse direction, respectively. In other words, when the rate constant in one direction is one thousand times greater than in the other. This way, the favorable direction rate constant is let to iterate free from its thermodynamic dependence by the opposite side, while the other is set to 0 meaning that no metabolites are exhausted in that direction. This threshold is arbitrary and just signifies that one of the constants is much larger that the other, i.e $k_j^{\pm} >> k_j^{\mp}$, so that the transformation of products to reagents is insignificant to the transformation of reagents to products or

vice-versa. The reversibility is defined by a instance variable `flxSgn` that is `none` for reversible reactions and takes the values of 1 or -1 if the reaction is considered irreversible in the forward or reverse direction. The equilibrium constants of the reactions are calculated at the given temperature by Equation B.6.

The genetic algorithm initiates with a initial metabolite concentration vector and initial kinetic constants vector, where one is related by the other by the reactions equilibrium constant at the temperature that the algorithm is being run. Each value in the vector represents an individual that differs by the default noise that actuates on 50% of the metabolites deviating them by a factor of 10% for both metabolite concentrations and kinetic parameters. These values were reached by trial and error and were aimed to create a substantially different combination of initial concentrations while not deviating much from the ambient temperature vector. It then iterates through the generations (refer to subsection 3.2.3) where in each individual it calculates its fitness. The fitness is dependent to the weighted euclidean distance between the flux distribution vector computed with the individual parameters and the sample flux distribution. The distance is weighted to represent a percentage change for the respective reaction. This way higher flux reactions don't bias the results. The individual with the lowest weighted distance is considered the best fit and it becomes the parent of the next generation. Its children are created by adding noise to the concentration and kinetic parameters vectors and the algorithm should be run for several generations in order to achieve better results. Each generation the noise has a maximum of 20% deviation from the base value in 10% of the metabolite concentrations and a maximum of 20% deviation in 40% of the reactions kinetic constants. Once again these values were obtained in a qualitative manner for this specific model in a trial and error fashion. The chosen stop parameter is when the simulation fitness stabilizes. Here, the best fit metabolite concentration, forward and reverse kinetic parameters vectors for the given temperature are returned.

3.3 Thermodynamic Equilibrium Constants

The estimation of the thermodynamic equilibrium constants is of pivotal importance to our methodology since it defines the relation between forward and reverse rate constants, and reaction irreversibility. The `eQuilibrator` library is used for the estimation of Gibbs energies and posterior calculation of the reaction's thermodynamic equilibrium constants. However, not all COBRApy `Reaction` objects are compatible with the library's functions. So these are translated into equilibrator_api `PhasedReaction` objects, that are read by the `physiological_dg_prime` function. This function calculates the $\Delta G'$ of a single reaction, assuming that all aqueous reactants are at 1 mM, gas reactants at 1 mbar and the rest at their standard concentration. This is important for posterior dimensional coherence, since concentrations are in M.

This yields the Gibbs free energy at 298.15K, however this energies depend on temperature by Equation 2.5. Entropy and enthalpy are known to depend themselves on temperature, but at this stage they are considered to remain constant under different temperatures. Yet, having only access to the Gibbs free energy and temperature values, Equation 2.5 remains unsolvable. If we assume that the enthaply/entropy ratio also remains constant we can solve the equation if enthalpy values are known. By the calculations instantiated at Appendix B, the reaction's Gibbs free energy at a specified temperature is given by Equation 3.7 where the ratio between the variation of Gibbs free energy and variation of enthalpy at ambient temperature (ϕ_0) is accounted for. A Standard Gibbs Energies and Enthalpies of Formation at 298 K, 1 atm, pH 7, and 0.25 M Ionic Strength table [Hammes and Hammes-Schiffer, 2015] provides access to Gibbs free and enthalpy energies of the several metabolites, thus allowing for the calculation of ϕ_0. This way the ratio of enthaply/entropy for each reaction can be calculated, by summing the energies of each compound metabolite of the reaction considering their stoichiometric coefficients by Equation 2.6. Some metabolites however were not present in the paper, so further curation was needed. For the metabolites that were the product of known energy metabolites reactions, its energies were estimated using group contribution method (refer to subsection 2.3.1). For the remaining metabolites a manual research of the energies was conducted or set to a default value of -300 kJ/mol. After estimating $\Delta G^{\ominus}$ for different temperatures, the respective $K^{\ominus}$ can be calculated by Equation 2.4.

$$\Delta G^{\ominus}_{T} = \Delta G^{\ominus}_{0eQuilibrator}\left(\phi_0 - \frac{T}{T_0}\phi_0 + \frac{T}{T_0}\right) \tag{3.7}$$

3.4 Metabolite Concentrations

Metabolite concentration data is essential for estimating rate constants in our methodology. Since our study organism is *Escherichia coli*, the *E. Coli* Metabolome Database (ECMDB) was used. It contains a compilation of metabolomic data and metabolic pathway diagrams about *Escherichia coli* (strain K12, MG1655) as well as additional material derived from hundreds of textbooks, scientific journals, metabolic reconstructions, and other electronic databases [Sajed et al., 2016].

All the obtained metabolite molar concentrations were measured on the Mid-Log growth Phase, i.e when the bacteria were growing at maximal rate, and at 37ºC or 310.15K. However, other experimental conditions might differ. Even though some of the abiotic parameters are registered, e.g temperature and pH, microbiological experiments are very sensitive to the unmeasured factors. This creates a major discrepancy in the concentration vectors since the values in the database come from different experimental environments and, most certainly, different experimental conditions. One could even argue that different *E. coli* strains, using different metabolic pathways, could be used in each measurement. To address this issue, the genetic algorithm instantiated at subsection 3.2.3 is used to mitigate the discrepancies in the values and, most importantly, create a viable, coherent concentrations vector for our model.

Metabolite concentration information was imported via a HTTP GET request of the https://ecmdb.ca/concentrations url that leads to the ECMDB concentrations browse. This way, the concentration values at exponential growth phase can be imported. For metabolites with several values, they are averaged out. Metabolites with no concentration information are set to a default concentration of 0.2 mM. One metabolite that requires special attention is water. Cytoplasmic water concentration is calculated at 41 M [Cayley et al., 1991] and remains fixed in the algorithmic run.

3.5 Survival Simulation

Having gathered all the essential components for the developed methods, a survival analysis proceeds. The simulated minimum and maximum rates for the reactions defined at subsection 3.2.1 and subsection 3.2.2, construct a possible flux rate interval at a given temperature, for our first two constant estimation methods. These intervals are applied for the reactions in a random flux solution space sample and the biomass flux for each sample is penalized by the weight of each reaction outside the interval. This way, crucial thermodynamically unfeasible reaction fluxes yield null biomass rates, while less important ones present a proportional penalty. Note that the flux solution space is a mathematical construction that emerges from the network, therefore blind to abiotic factors and suitable for an unbiased thermic analysis. Repeating the process for different samples of the solution space simulates how different metabolic configurations would behave at the given temperature, and how much biomass could it produce.

The third method instantiated at subsection 3.2.3, which simultaneously estimates rate constants and metabolite concentrations, yields three parameters after the optimization stop that can be analysed for their survival impact: simulation optimal fitness, and respective metabolite concentration, forward and reverse rate constants. If fitness decreases significantly for some temperatures, i.e the minimal weighted distance between the simulation and sample flux distributions increases, it could mean that the algorithm is not able to find the necessary parameters to approximate to the sample. This indicates a mathematical incapacity for the model to find rate constant values that simultaneously can generate the steady state reaction fluxes of that sample and are in agreement with the thermodynamic constraints at that temperature. But the optimization can yield a solution that is very close to the sample fluxes while being biochemically unfeasible. For this the calculated *optimal* parameters are of use. In the case of metabolite concentration, osmolarity can be analysed. Osmolarity, or osmotic concentration, is the measure of the accumulated concentration of all solutes, defined as the number of osmoles (Osm) of solute per litre (L). We calculate the simulation osmolarity by simply summing all the metabolite concentrations except water and a threshold can be defined from *E. coli* osmolarity measurements. Above 300 mOsm, the observed cytoplasmic osmolarity [Stock et al., 1977], growth rate starts decreasing linearly with increasing osmolarity, reaching zero at an osmolarity of approximately 1.8 Osm [Cayley et al., 1991]. Variation in osmolarity affects osmotic pressure that could potentially damages the cellular walls. While no quantitative calculations on osmotic pressure were performed, a reference osmolarity at T_0 can be determined to account

for the variation from that referenced state. A T_0 reference metabolite concentration vector is initially calculated to determine the respective osmolarity.

The variation of the rate constants can also be analysed to infer cellular performance. Our kinetic parameters depend on enzyme concentration and enzyme efficiency (refer to section 2.2), therefore demonstrating the enzimatic load on the organism. All reactions are enzimatically mediated and the synthesis of these enzymes is energy and resource intensive. A rise in rate constant values under non-standard temperature means either an increase in enzyme concentrations, a resource intense process, or enzyme efficiency, a rare, evolutionary trial and error process. A rate constant decrease does not pose these problems since the halt of enzyme production or enzyme inhibition is easily achieved by the cell [Newton et al., 2018]. In a nutshell, enzyme activity is limited by the spatial and temporal constraints of the cell, by its kinetics, and by the energy and resource availability. Considering this, the sum of the overall forward and reverse kinetic constants can qualitatively estimate resource usage, or overall *stress*, that the organism undergoes to reach a certain state. By computing the ratio between a given temperature stress and an ambient temperature reference stress, we can infer the enzyme load variation for that temperature. Another consideration is the weight of the reaction at play. Reactions that do not affect the biomass function and pose a stressful cellular environment in non-standard temperatures, can be rendered down by the cell's regulatory systems. So we designed Equation 3.8 as a **stress function that analyses the overall enzimatic load that the organism must undergo in order to reach the desired state at a given temperature**, considering this effect for each and every one of the targeted reactions.

$$\text{stress}_T = \frac{1}{N} \sum_{j}^{N} \frac{k_{jT}^{+} + k_{jT}^{-}}{k_{j0}^{+} + k_{j0}^{-}} \text{ weight}_j \tag{3.8}$$

3.6 Software

This section instantiates how the pipeline code is constructed. Python software was the preferred language since it has a wide library portfolio for data handling, e.g pandas for analysis [McKinney, 2010] and JSON for storage, and the existence of Python ready models and libraries, e.g COBRApy for metabolic constraint-based reconstruction and flux analysis [Ebrahim et al., 2013] and eQuilibrator for calculation of the reaction's $\Delta G^{\ominus}$ [Noor et al., 2013], for our specific case study. Other reasons include the simplicity, user friendly nature and universality of this programming language. Motifs that do not directly influence this decision regarding the primary objective, but go in hand with the construction of an easily usable software. The project's code, functions, classes, methods, and files are available as a GitHub repository[1].

[1] https://github.com/lucasmonteiro859/ThermoGEM.git

3.6.1 Pipeline

Firstly a BiGG model `.xml` file is imported using the COBRApy `read_sblm_model()` function. The model contains reactions of interest and reactions that are not of interest for our analysis. Due to this reason the target reactions must be selected. This requires that the user defines by a dictionary where the keys are the reactions id's and the values are the respective boolean of is target. There is the exception of the biomass reaction. This reaction must have the value of `"bioMRcn"` string. These are the parameters necessary to create an object of class `Metabolism`. This object represents the metabolism of the modeled organism and its parameters must be set in order to run its methods. The reaction's weights are set by `Metabolism` method's `setWghs`, metabolite concentrations by `setCncs`, ambient temperature Gibbs free energies by `setTrgRcnGec0s` and ϕ_0 values by `setTrgRcnHgr0s`. All methods require a dictionary where the keys are the reactions id's and the values the respective values. Flux samples are required for the different methods so a `getFlxSmp(n)` is available, returning n flux solution space samples. The `genAlg(smp, tmp)` runs the subsection 3.2.3 method for a *smp* sample in a *tmp* temperature, returning optimal solution parameters - minimal distance, temperature, metabolite concentration and rate constant vectors.

While the supercited functions allow for the running of the genetic algorithm, a set of analysing and visualization tools are also available:

- `avgDevAvgDst` - for sample distance variance over temperature analysis
- `calRefs0` - for calculating reference vectors of a set of samples
- `cncDevT` - for sample metabolite concentration variation over temperature
- `plotGecs` - for $\Delta G^{\ominus}$ and $K^{\ominus}$ temperature variation visualization
- `plotDstsHist` - for ambient temperature distance visualization
- `plotAvgDevAvgDstsHist` - for distance temperature variance visualization
- `plotCncDevs` - for metabolite concentration temperature variation visualization
- `plotKinStr` - for kinetic parameter stress visualization

Overall, the repository should provide the tools that are required to replicate this thesis results and to apply the same approach to other organisms genome scale metabolic reconstructions.

Chapter 4

Results and Discussion

This section explains how the results were obtained, the results themselves, their interpretation, and subsequent discussion. Since we followed three different methods, that evolved sequentially in an attempt to solve the previous method's problems, the process is developed one method at a time following this project's setbacks and overall reasoning in chronological order. The first logical step is getting the model computationally ready for use. The pipeline at subsection 3.6.1 is followed to import the reaction weights, metabolite concentrations, reaction free Gibbs energy variation and ϕ_0 values obtained by subsection 3.1.2, section 3.4, and section 3.3 respectively. All these are essential to parameterize our model and implement the *E. coli* metabolism performance temperature dependent simulations. The model's reactions and metabolite characteristics are tabulated in Appendix C.

4.1 Equilibrium Constants Dependence on Temperature

Temperature can affect physical and chemical processes occuring in living organisms through multiple mechanisms. In this work we opted to study just one of these mechanisms: how temperature influences the metabolic network through its impact in the thermodynamics of biochemical reactions. This thermic effect emerges from the fundamental thermochemistry that depends on the energy differences between products and reactants. These temperature dependent reaction energies influence the value of the equilibrium constant, which in turn conditions the reaction kinetic parameters. The forward and reverse rate constants are related to each other by the reaction equilibrium constant (Equation 2.3). So to include the behaviour of reaction's thermochemistry, we calculate $K^{\ominus}$ values by Equation 2.4 using the $\Delta G^{\ominus}$ estimation of Equation 3.7 for the desired temperature. Depicted in Figure 4.1, three types of metabolic reactions can be distinguished based on the variation of equilibrium constants with temperature. Those who quickly exchange from irreversible in one direction to the other, those who remain reversible over the range of temperatures and those in between who remain reversible for some range of temperatures, slowly favoring one reaction direction over the other, eventually leading to irreversibility. The last category can

be divided into the reactions where the equilibrium constants increase or decrease with temperature, respectively. Note that in the first type, the small range of temperatures in which the reactions remain reversible, i.e roughly from 280K to 320K or 7ºC to 47ºC, is the standard temperature values found on earth where we know that metabolic systems can reach a steady state.

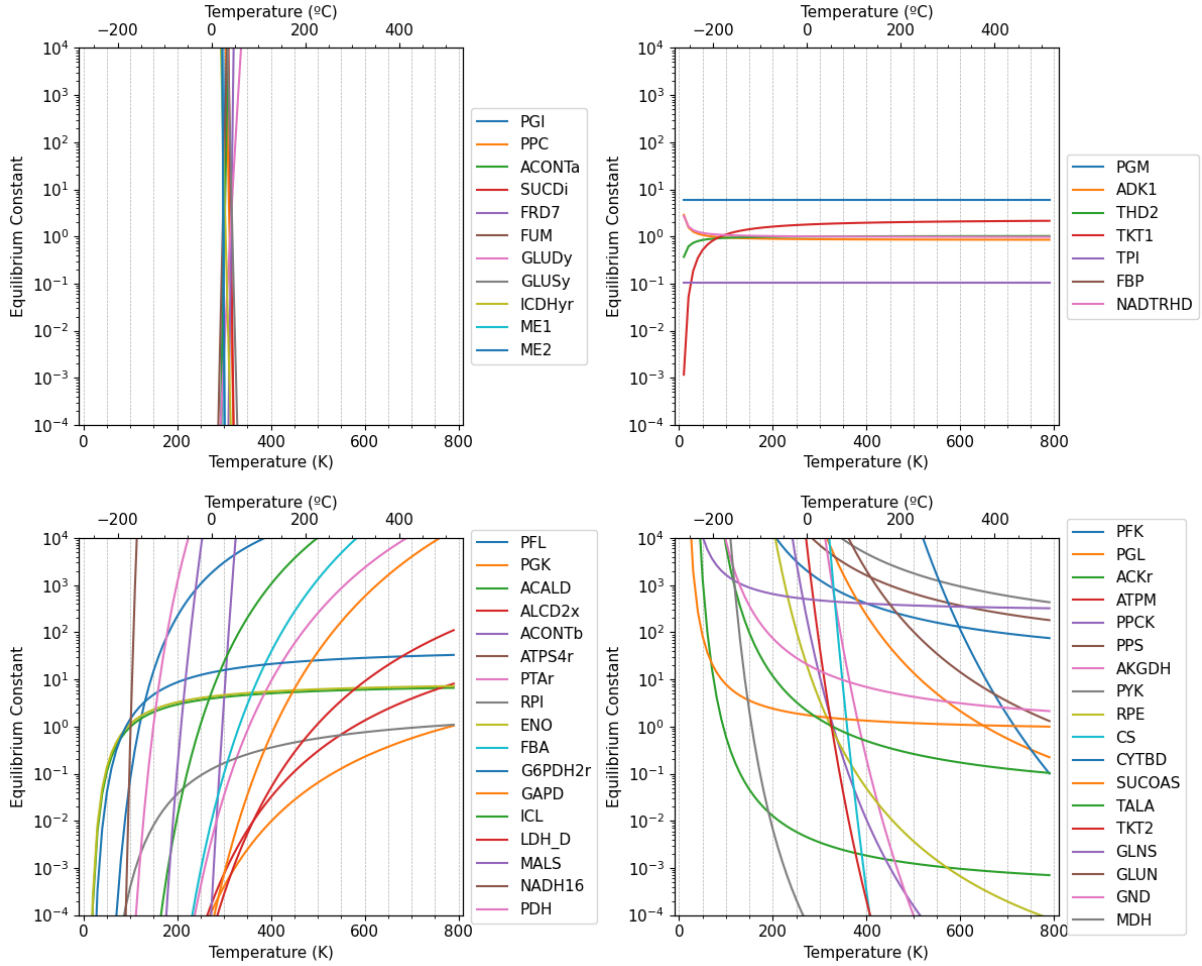


Figure 4.1: Equilibrium constant estimations for our model's reactions from 10K to 800K. According to our assumptions, $\Delta G^{\ominus}$ grows linearly with temperature affecting $K^{\ominus}$ behaviour. Plots show how the $K^{\ominus}$ values change with temperature. Each subplot represents a different subgroup of the model's reactions that show a similar behavior: $K^{\ominus}$ quickly changes irreversibility from one direction to the other over a range of $\sim 40K$, between 280K and 320K, by surpassing the $10^{\pm 3}$ threshold (**Top Left**), remains reversible for a wide range of temperatures (**Top Right**), slowly changes reversibility from reverse to forward (**Bottom Left**) or vice-versa (**Bottom Right**).

4.2 Algebraic and Low Variation Methods

Having our estimation for the reactions' equilibrium constants at different temperatures, values for metabolite concentrations and rate constants were still lacking. Therefore the methods in section 3.2 were employed. A direct algebraic estimation of the rate constants using Equation 3.4 and the imported metabolite concentration values yielded absurd results at ambient temperature, with extremely high or low values, emphasizing the incoherence in our data. We hypothesized that we could correct the incoherences in the metabolite concentration vector through an optimization step using a genetic algorithm.

We implemented a genetic algorithm to optimize metabolite concentrations and calculate the rate constants with Equation 3.4. During the optimization, fitness of metabolite concentration values is defined by the distance between the reaction flux vector computed with these concentrations and the model *optimal* flux randomly generated by FVA. 100 optimizations were run for 400 generations, a number of generations qualitatively determined to reach a fitness plateau. The best fit metabolite concentration, forward and reverse kinetic parameters vectors are returned at algorithm termination for the ambient temperature and fixed for the temperature range simulations where we introduce the equilibrium constant and a up to 50% perturbation to 10% of the metabolite concentration vector.

The method laid out in subsection 3.2.1 was run from 10K to 800K. The results show an irregular pattern where at around 100 degrees kelvin the biomass function increases from 0 to its maximum, excluding some intervals. One of this intervals is at the standard temperature, temperature that, in the way the model is constructed, should yield a positive flux in the biomass function. A further analysis revealed **negative rate constants values deeming this method unsuitable**. Rate constants represent unidirectional reaction rates when reactants have unit concentrations, so negative values pose invalid solutions. These results show again the incoherence in our data and the need for other methodology.

We tried to perform a low concentration variation approximation as a solution to this problem. As demonstrated in subsection 3.2.2, we can take advantage of the mathematical expression for reaction fluxes to erase the dependence on metabolite concentrations. We ran this method in the same thermic interval as before. The simulations matched flux distributions for temperatures near the standard temperature but yielded null stationary states for all temperatures that differed 2 degrees kelvin from this. *E. coli* is a thermic stress tolerant bacteria and the literature provides insight on how its flux distribution varies with temperature [Strocchi et al., 2006], so this method was also deemed unsuitable since the flux distribution should match or at least allow for biomass production at higher, or lower, temperatures. One could argue that **the metabolite concentration low variation approximation is false** and that concentrations are indeed changed by the thermic stress.

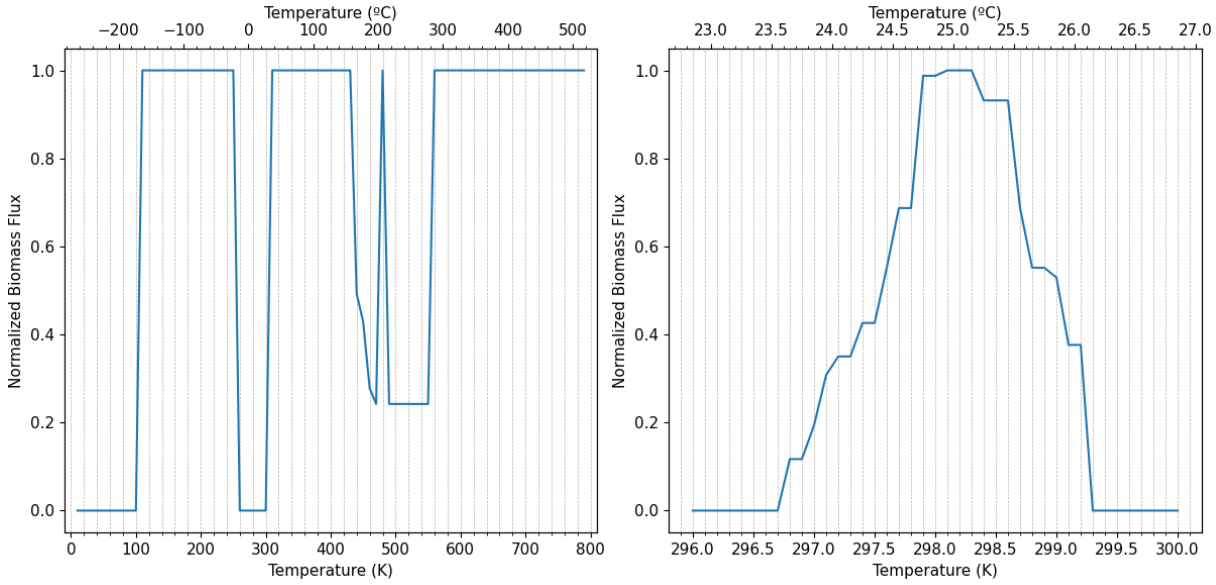


Figure 4.2: Normalized biomass flux results from 10K to 800K obtained with algebraic estimation (**Right**) or low concentration variation approximation (**Left**). Algebraic estimation revealed negative kinetic constants, and null biomass flux at ambient temperature, rendering it unfeasible. Low concentration variation approximation only produces biomass near standard temperature, contrary to observations, so this approximation is not valid.

4.3 Rate Constants and Metabolite Concentrations Joint Optimization

In an effort to better estimate the rate constant values, we continued by extending the already implemented genetic algorithm, to simultaneously estimate the rate constants. This differs from the first method because the **rate constants are free to adjust, although always maintaining their equilibrium constant relation** at the given temperature. Before, the optimization fitness was calculated by the distance to a optimum flux at standard temperature, fixing the parameters for other temperature simulations. Now the simulation is run at a given temperature and the fitness calculated by the distance to a flux solution space random sample, estimating the parameters at *that* temperature.

As explained in section 3.5, three variables can be analysed to infer the survival of the organism. At standard ambient temperature, solutions should meet our law of mass action approximation and a maximum fitness solution, i.e null distance in the genetic algorithm, should be found. Besides distance, osmolarity of less than 300 mOsm at standard temperature should be observed. Also, high deviation from the standard temperature metabolite concentration or rate constant values could mean a stressful situation for the metabolism to adapt.

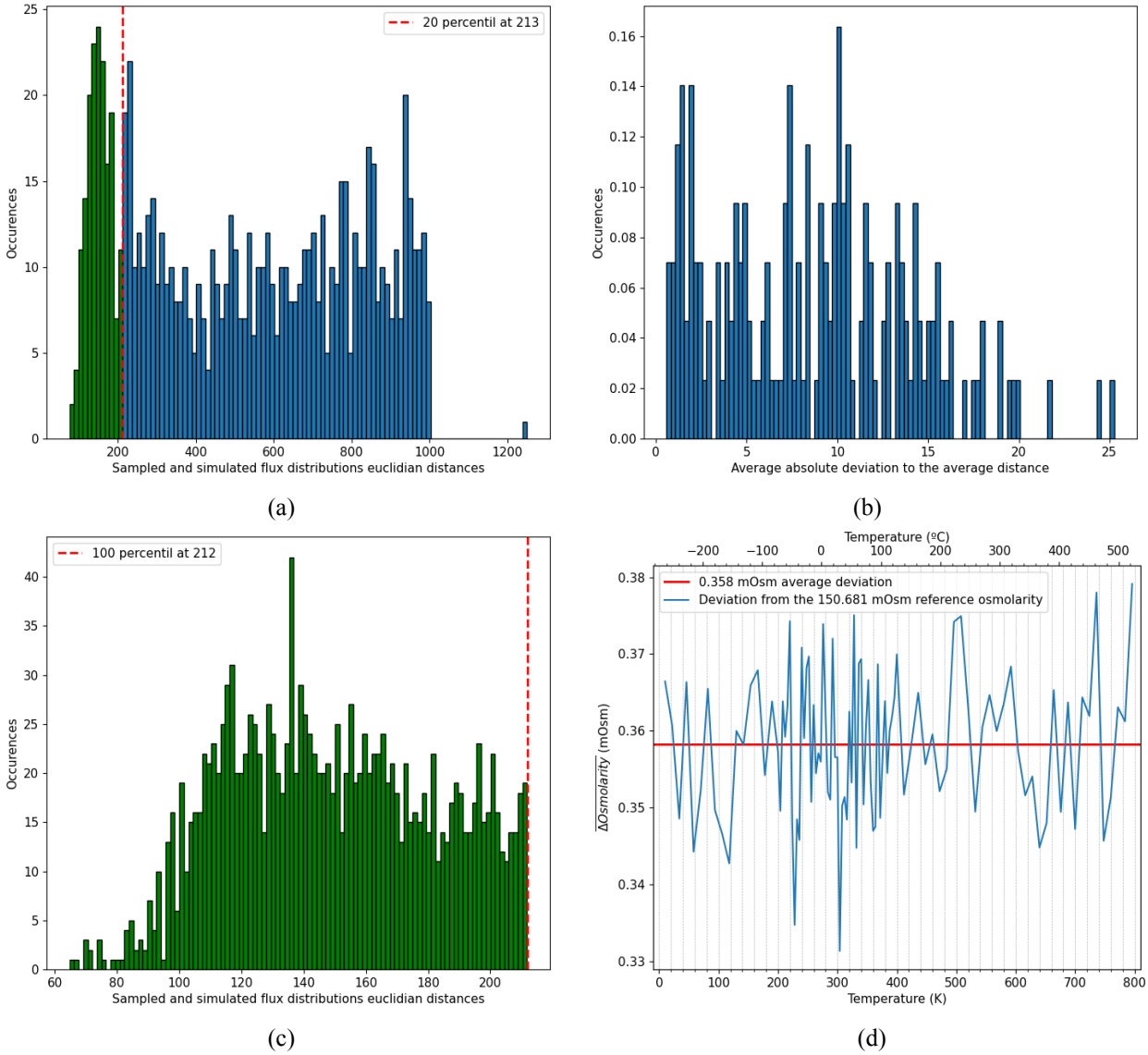


Figure 4.3: Distance and osmolarity analysis for two solution space samplings. **Top Left (a)**: The histogram of the euclidean distances between 863 sampled and optimized flux distributions allowed the definition of a viable distance threshold. **Top Right (b)**: The low average absolute deviation from the average minimal distance for the 20% fittest samples shows that distance doesn't vary with temperature in this model, thus making it an inadequate parameter for survival analysis. **Bottom Left (c)**: With threshold defined at 213, a 1584 sample optimizations were accomplished. **Bottom Right (d)**: Their reference osmolarity was calculated at 150 mOsm, and all the absolute variation from this value were between 0.33 and 0.38 mOsm in a temperature range between 10K and 800K. Since the variation values are low, osmolarity was also deemed an inadequate parameter for survival analysis.

To understand the overall performance of the model, we did a preliminary analysis of the genetic algorithm in a temperature range of 10K to 800K and a spacing of 40K for 863 flux solution space samples. No null distances were found so only those with the smaller distances are deemed viable metabolic flux solutions. Figure 4.3a histogram shows all minimum distances at standard temperature and we determined the threshold distance at the 20% highest fitness samples. This corresponded to the samples that showed distances below 212 at T_0. The variation of distance with temperature for each solutions was calculated as the average absolute deviation to the global distance average across all temperatures. All solutions presented a low variation with temperature (Figure 4.3b). This shows that the **viable samples, by optimization fitness remain viable for all studied temperatures**. This happens because the genetic algorithm parameters are free to adjust themselves to unrealistic values in order to meet the required fluxes. 95% of the samples met the 300 mOsm osmolarity requirement. Therefore, to discriminate the survivability at different temperatures, we needed to analyse how temperature influences metabolite concentration and the rate constants.

For this reason, we reran the genetic algorithm for 6574 samples of the solution space for a range of temperatures between 10K and 800K with a spacing of 12K, and a spacing of 4K between 200K and 400K. Only the results for the samples that showed minimal distances below 212 at T_0 were analysed. That meant a reduction of 81% leaving us with a sample set size of 1225 samples. The same analyses were performed to the new set of samples, and similar results were obtained (Figure 4.3c). The distance values for the same sample do not vary significantly with temperature, with an average absolute deviation of around 8 for averages in the order of 100. As such, this parameter is not the determining factor for the survival of our model across the temperature range studied.

We continued to analyse the model's parameters by the methodology at section 3.5. Reference osmolarity was determined to be ~ 150 mOsm by the average of the osmolarities of all samples. Reported osmolarities of cells are at 300 mOsm. Since our model is a core model containing an incomplete set of metabolic reactions and metabolites it is natural that the model osmolarities are lower than the experimentally determined values. The average variation for every temperature was measured and osmolarity values did not vary significantly, with a variation of 0.36 ± 0.02 mOsm as represented in Figure 4.3d. Since **our model's osmolarity isn't dependent on temperature**, it cannot be regarded as significant when studying how temperature influences the metabolism. The remaining factor to analyse is the magnitude of rate constants, analyzed through a stress function. All samples presented high stress at low temperature and a decline around the standard temperature, where a low stress *plateau* starts. Here, at near optimal conditions, the stress isn't null since reactions always deviate from the reference state due to the algorithm's noise (the stochastic nature of the simulations is observed at Figure 4.4). The low stress plateau continues to around 620K where a steep increment is observed and high stress regime is achieved. To smooth out the results and understand the impact of a wide range of temperatures in metabolism, we averaged the stress functions over all samples creating an overall stress quantifier, represented in Figure 4.5.

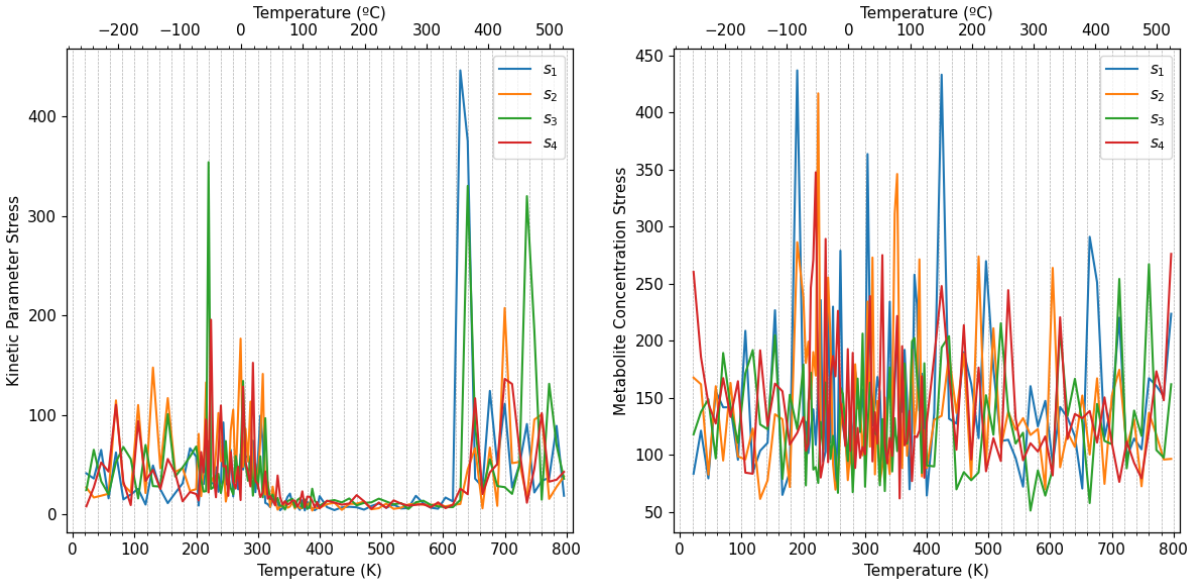


Figure 4.4: Absolute kinetic parameter and metabolite concentration stress for samples s_1, s_2, s_3, and s_4. **Left**: Kinetic parameter stress is lower in the 320K to 620K temperature range. **Right**: Metabolite concentration stress remains stable around 150 mOsm. Individual sample stress is irregular due to the model's stochacity, sometimes surpassing critical values e.g the 300 mOsm observed cytoplasmic osmolarity. However this irregularities are not significant when a large sample set is analysed and an overall behaviour of the model becomes apparent.

4.4 Considerations

The two-stepped characteristic of the stress function becomes readily apparent. What is happening around 320K and 620K to create such a drastic change in stress? Decomposing the stress function into the reaction's respective average stress it was found that two metabolic reactions constituted the majority of the stress levels at the low and high temperature ranges *respectively*. At low temperatures Aconitase, the tricarboxylic acid cycle reaction that catalyses the conversion of citrate to isocitrate [Flint and Allen, 1996], and at high temperatures glyceraldehyde-3-phosphate dehydrogenase (GAPD), the sixth step of glycolysis coverting glyceraldehyde-3-phosphate to D-glycerate 1,3-bisphospate in the presence of nicotinamide adenine dinucleotide (NAD+) and inorganic phosphate, mediating the formation of NADH and adenosine triphosphate (ATP) [Tristan et al., 2011], the energy currency molecules of the cell.

A closer analysis reveals that both reactions become irreversible at the temperatures where the metabolism abruptly adapts. Due to the threshold values of the equilibrium constant, i.e at 10^{-3} and 10^{3}, the reactions pass from reversible to irreversible, leading to changes in the metabolism even thought the ratio between the rate constants in the forward and reverse directions is higher than 10^{3}. GAPD becomes irreversible at 620K. It is expected that when such an essential reaction for the cellular metabolism is deemed thermodynamically infeasible, the cell becomes not viable. Aconitase abruptly changes from irreversible in the forward to reverse direction between 280K and 320K. The stress is minimal when the reaction is irreversible in the citrate to isocitrate direction, i.e the convectional triboxylic acid direction.

However, we should not neglect that stress levels are not optimal under the range of temperatures where life on earth usually occurs. The growth of *Escherichia coli* cells is only impaired at temperatures below 294K and stops at 280.5K [Strocchi et al., 2006] and our results clearly go astray from these observations. This deviation could be due to the fact that the model used represents just the central metabolism of the bacteria. More complex metabolic networks could adapt new pathways to address the abrupt change that could be caused in the metabolism if certain reactions are deemed unfeasible. Moreover the highest recorded temperatures for cellular survival is near 390K [Beulig et al., 2022] while our results remain thermodynamical optimal from these temperature until 620K. The emphasis is on the *thermodynamical* term since only thermodynamical parameters are being considered while other thermo-sensitive mechanisms like protein folding or substrate affinity for transport systems [Nedwell, 1999] are not. The error from our energy estimations could also deviate the values at which these key reactions are reversible. eQuilibrator provides the errors of their calculations and we consider a ϕ_0 error of 10% to plot Figure 4.5 that illustrates how the statistically possible deviations could influence the stress experienced in these limit zones. In the particular case of Aconitase, it shows that the temperature lower limit at which it becomes irreversible is still above 300K where the cell should not have a strongly impaired performance. It is very important to note that the computed stress function has no direct physical meaning. Its values are a **relative estimation of the metabolic change needed to meet the metabolic network steady state requirements** and how this change implies an increase in enzymatic activity. There is not a direct quantitative translation into survival likelihood or growth impairment. When high stress function values are obtained we can only conclude that the metabolic network is compromised and requires an outstanding adaptation. We do not know the stress value above which this adaptation is not possible.

The results show three well-defined stress regimens in the large spectrum of temperatures, which fail to represent experimental observations accurately and suggest that the methods used require further development. Hence, the analysis of survivability at other celestial bodies will be only qualitative and speculative instead of a well grounded prediction of microorganism survival in these habitats. The moon is the closest and most requested target for interplanetary missions, being the only celestial body that humans have stepped on besides Earth. Its surface presents abrupt temperature variation. During nighttime the temperature is relatively constant below 100K throughout the globe. At sunrise the temperature

sharply increases and stabilizes above 300K for regions below 60º latitude until sunset, due to the lack of atmosphere [Williams et al., 2017]. According to our results, thermodynamically feasible metabolisms could develop during lunar daytime. Further hypothesising, lifeforms could enter latent phases during the approximately 15 days of nighttime and develop during the, same length, daytime. It is known that *E. coli* can survive such periods of latency [Takano et al., 2023]. Other interesting case-study is Mars. Proofs of an ancient global ocean, atmosphere, and magnetic shield, put Mars as the planet with the potential of having Earth-like lifeforms, as of now or in the past. Due to the farther distance from the sun and a denser atmosphere, its temperatures are lower and more evenly distributed throughout the martian day. The highest temperature record was 308 K in the martian equatorial summer by the rover Spirit [NASA, 2007]. So never surpassing the temperatures that we find thermodynamically optimal for an organism.

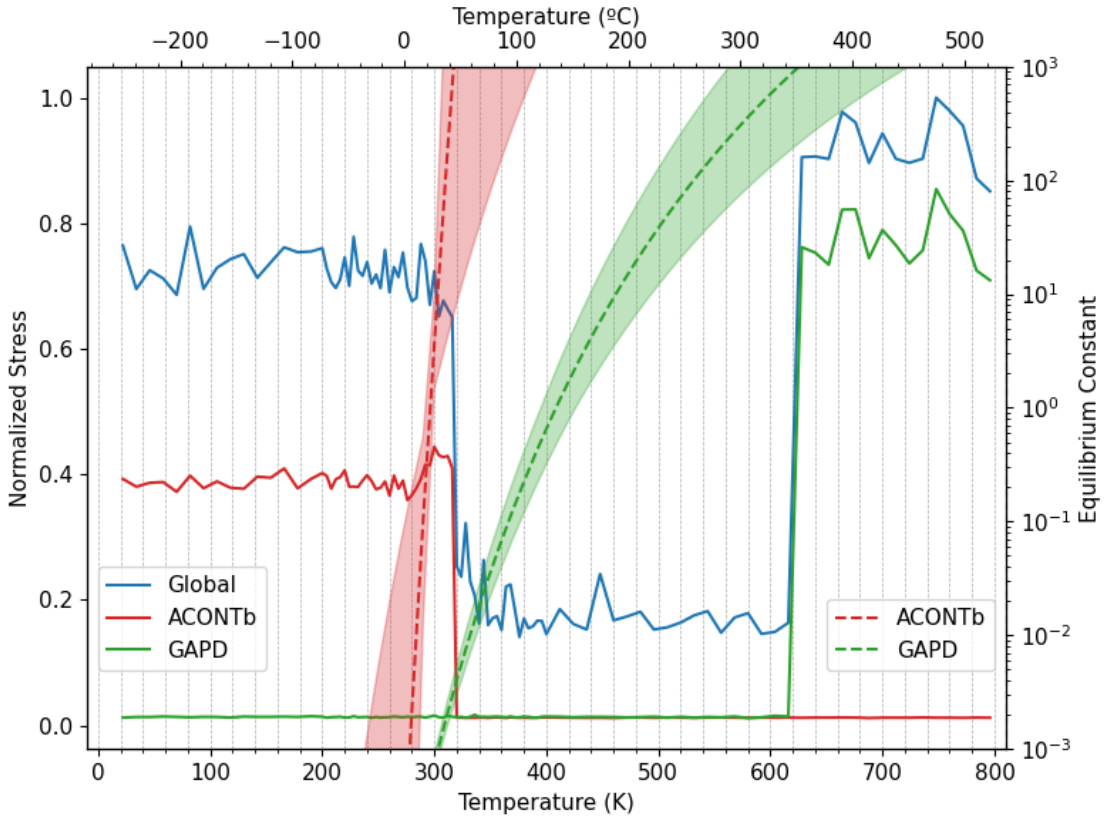


Figure 4.5: Normalized average kinetic parameter stress of aconitase (ACONTb, red continuous line), glyceraldehyde-3-phosphate dehydrogenase (GAPD, green continuous line), and the sum for all model's reactions (blue continuous line), for 1584 simulations over a range of temperatures between 10K and 800K. The rise in stress is due to metabolic changes when reactions change their reversibility, i.e their equilibrium constant is higher than 10^3. This is evidenced by equilibrium constant evolution over temperature (dashed lines). Shaded areas represent estimation error.

Chapter 5

Conclusion

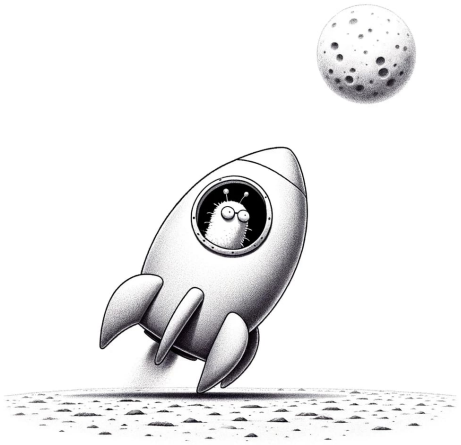
We have developed a methodology to evaluate the likelihood of survival of an organism at different temperatures due its impact in the thermodynamics of biochemical reactions. This methodology evaluates random samples of steady state metabolic flux solutions obtained from a genome-scale metabolic model of the organism. First, our method estimates the equilibrium constants of the biochemical reactions in the model at temperatures of interest. Then, for each sampled flux solution, it estimates the optimal metabolite concentrations and rate constants that better approximate the sampled flux solution using a genetic algorithm. This estimation guarantees that the ratios of forward and reverse rate constants imposed by equilibrium constants at that temperature are obeyed. The likelihood of survival for that sample is then evaluated according to 1) how well the sampled flux solution was approximated, 2) the osmolarity induced by the optimized metabolite concentrations, and 3) the presence of very high rate constant values (in comparison to the values at standard temperature) in the optimized solution, which suggests the need to a large increase in enzymatic activity to enable the metabolic network to reach a steady state. In our work, the first two factors did not allow to discriminate the different sampled flux solutions at different temperatures, and only the third factor (rate constant magnitude) allowed an evaluation of survival at different temperatures. Several hurdles were found along the way and solutions were developed throughout this thesis. The main issue was that the initial metabolite concentration values were found to be inconsistent, probably due to different experimental conditions used in their measurements. Our initial proposal to overcome this issue was to perform an optimization of metabolite concentrations with a genetic algorithm. The values continued to be inconsistent, so the genetic algorithm was extended to estimate also the model's rate constants. Such parameters are thermodynamically dependent on each other so they require the knowledge of the reaction's Gibbs free energy variation. Unfortunately, Gibbs energies databases are scarce, so we used the eQuilibrator API and thermodynamical values of several reactions found in the literature to estimate Gibbs energies at standard temperature, using the group contribution method when needed. The energies for other temperatures were linearly derived from the Gibbs energy at ambient temperature, depending on the enthalpy/entropy ratio of the respective reaction. Finally, a stress function was developed that compared the optimized rate constants at a given temperature with reference standard

temperature values, giving a higher weight to reactions that had a higher impact in the biomass reaction flux. This stress function originated a square-like step temperature profile with a thermodynamical stress minima between 320K and 620K. This thermodynamical baseline temperature range represents the range in which the metabolic system is able to reach the sampled steady state solutions without the need for costly enzymatic activity adaptations. Outside this stress minima range, reactions might become irreversible and deem biomass syntheses impossible, or at least require more resources. This happened in our *E. coli* model where GAPD reaction, a driver of glycosis, became irreversible at 620K, and cellular energy production could no longer be adequately regulated.

Several questions were evoked throughout the thesis. What should be an optimal threshold value for reaction reversibility, and, should it be dependent on the reaction and/or temperature? The genetic algorithm's parameters deserve a more in depth research and development. What rates of mutation are optimal for our model? Or number of simulations? Should we consider solution breeding? Can we elevate the efficiency of the code be it by optimization or automatization, e.g automatic algorithm stopping mechanism? Other implementations could be suggested, such as to include k_m and k_{cat} reaction parameters and compute reaction fluxes with enzymatic rate laws instead of the used mass action law approximation. Additionally, we could consider samples of steady state flux solutions from mutant metabolic models. In this way we could study new flux distributions that could prove to be more adapted to different temperatures. The impact of each reaction in the biomass reaction flux could also be different in these mutant models. It seams reasonable to remove singular or multiple reactions since the shutdown of metabolic pathways may occur as an adaptation strategy that is more easily implemented than the creation of new chemical reactions, as discussed before.

However, the main difficulty for advancing with this project remains the data. For the model to yield practical predictions we need to gather precise, coherent data for both metabolite concentration and reaction thermochemical data. A secondary step would be to better understand how these variables depend on temperature, in order to estimate model performance in a wide thermodynamical range, be it empirically or theoretically by implementing methods that currently are not feasible, e.g quantum-mechanical computations. If these different suggestions are successful and a good method is achieved, the analysis of the model with planetary data becomes of importance. We previously discussed how the Moon surface temperature varies wildly and Mars subsurface temperatures may present closer to Earth conditions. Global temperature maps are available and dynamical analysis could be performed to address this issues. Also, the physical conditions don't depend only on temperature but on other factors like pressure and ion concentration. Crucial points for life e.g water freezing point that impacts reaction speed and the stiffness of the cell, depend on all of these factors. Although an important first step, the focused analysis on temperature is incomplete. Ion strength plays a role in Gibbs energy calculation as well as in transport mechanisms so its consideration is needed for a more precise method. How could we implement these variables in our model and interpret the more complex stress profile. Could we link it with cell survival?

Nonetheless, with the brewing of new space colonization plans by several space agencies and private companies, the need to study biological contamination potential arises, be it for health [Schiwon et al., 2013; Gilbert et al., 2020], ethical or purely scientific reasons. This work is an initial step towards understanding how organisms behave under non-standard temperatures, and if they have the potential to contaminate such environments. We obtained rather uncertain results but it allowed to grasp the coming hurdles and possible solutions for a project of this kind. It also gives us a first glimpse of how simple models behave at other temperatures. Such capability is important to understand extraplanetary biological contamination scenarios but it also has the potential to optimize industrial processes, create bio-remediation techniques, development of biosensors, etc. Synthetic biology seeks to create new biological parts, devices, and systems, or to redesign systems that are already found in nature. A mastery of metabolic thermodynamics could increase these processes efficiency or even design new metabolic pathways and areas not yet explored in the field. A strong body of investment is currently focused on this field [Bryan, 2021] and new methods and models like the one presented here could play a role in the advancements. With that said, the project opens the way for the development of abiotic factor dependent genome-scale metabolic models in the quest of understanding how organisms react to change, be it in the genetic engineering vanguard or humanity's near future space saga.



The E. colonizer

References

Beulig, F., Schubert, F., Adhikari, R. R., Glombitza, C., Heuer, V. B., Hinrichs, K.-U., Homola, K. L., Inagaki, F., Jørgensen, B. B., Kallmeyer, J., and et al. (2022). Rapid metabolism fosters microbial survival in the deep, hot subseafloor biosphere. *Nature Communications*, 13(1). 30

Bryan, E. (2021). The synthetic biology revolution - investing in the science of sustainability. 35

Cayley, S., Lewis, B. A., Guttman, H. J., and Record, M. (1991). Characterization of the cytoplasm of escherichia coli k-12 as a function of external osmolarity. *Journal of Molecular Biology*, 222(2):281–300. 19

Coleman, J. (2022). Mount everest is teeming with life, from fungi to butterflies. *National Geographic*. 1

Connors, K. A. (1990). *Chemical Kinetics: The study of reaction rates in solution*. Wiley-VCH. 7

Cooper, G. M. (2000). *The Central Role of Enzymes as Biological Catalysts*. Sinauer Associates, Inc, 2nd edition. 7

Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). COBRApy: COnstraints-based reconstruction and analysis for python. *BMC Systems Biology*, 7(1):74+. 12, 20

Flint, D. H. and Allen, R. M. (1996). Iron−sulfur proteins with nonredox functions. *Chemical Reviews*, 96(7):2315–2334. 29

Fröhlich-Nowoisky, J., Pickersgill, D., Despres, V., and Pöschl, U. (2009). High diversity of fungi in air particulate matter. *Proceedings of the National Academy of Sciences of the United States of America*, 106:12814–9. 1

Gilbert, R., Torres, M., Clemens, R., Hateley, S., Hosamani, R., Wade, W., and Bhattacharya, S. (2020). Spaceflight and simulated microgravity conditions increase virulence of serratia marcescens in the drosophila melanogaster infection model. 6(1):4–4. 35

Hammes, G. G. and Hammes-Schiffer, S. (2015). *Appendix 4 Standard Gibbs Energies and Enthalpies of Formation at 298 K, 1 atm, pH 7, and 0.25 M Ionic Strength*, pages 465–466. John Wiley Sons, Ltd. 18, 43

Harvey, A. and Mann, A. (2022). Nasa's artemis program: Everything you need to know. *Space.com*. 1

Heuer, V., Inagaki, F., Morono, Y., Kubo, Y., Spivack, A., Viehweger, B., Treude, T., Beulig, F., Schubotz, F., Tonai, S., Bowden, S., Cramm, M., Henkel, S., Hirose, T., Homola, K., Hoshino, T., Ijiri, A., Imachi, H., Kamiya, N., and Hinrichs, K.-U. (2020). Temperature limits to deep subseafloor life in the nankai trough subduction zone. *Science*, 370:1230–1234. 1

Kawaguchi, Y., Kawaguchi, Y., Shibuya, M., Kinoshita, I., Yatabe, J., Narumi, I., Shibata, H., Hayashi, R., Fujiwara, D., Murano, Y., Hashimoto, H., Imai, E., Kodaira, S., Uchihori, Y., Nakagawa, K., Mita, H., Yokobori, S.-i., and Yamagishi, A. (2020). Dna damage and survival time course of deinococcal cell pellets during 3 years of exposure to outer space. 12

Khannapho, C., Zhao, H., Bonde, B. K., Kierzek, A. M., Avignone-Rossa, C. A., and Bushell, M. E. (2008). Selection of objective function in genome scale flux balance analysis for process feed development in antibiotic production. *Metabolic Engineering*, 10(5):227–233. 6

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., and Lewis, N. E. (2015). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44(D1):D515–D522. 12

McKinney, W. (2010). Data structures for statistical computing in python. In van der Walt, S. and Millman, J., editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56. 20

Mitchell, M. (1998). *An introduction to genetic algorithms*. MIT. 9

Moeller, R., Reitz, G., Nicholson, the PROTECT Team, W. L., and Horneck, G. (2012). Mutagenesis in bacterial spores exposed to space and simulated martian conditions: Data from the expose-e spaceflight experiment protect. *Astrobiology*, 12(5):457–468. PMID: 22680692. 12

NASA (2007). Extreme planet takes its toll. 31

Nedwell, D. (1999). Effect of low temperature on microbial growth: Lowered affinity for substrates limits growth at low temperature. *FEMS Microbiology Ecology*, 30(2):101–111. 30

Newton, M. S., Arcus, V. L., Gerth, M. L., and Patrick, W. M. (2018). Enzyme evolution: Innovation is easy, optimization is complicated. *Current Opinion in Structural Biology*, 48:110–116. 20

Nisbet, E. and Sleep, N. (2001). The habitat and nature of early life. *Nature*, 409:1083–91. 1

Noor, E., Haraldsdóttir, H. S., Milo, R., and Fleming, R. M. T. (2013). Consistent estimation of gibbs energy using component contributions. *PLOS Computational Biology*, 9(7):1–11. 20

Orellana, R., Macaya Ramos, C., Bravo Cortes, G., Dorochesi, F., Cumsille, A., Valencia, R., Rojas, C., and Seeger, M. (2018). Living at the frontiers of life: Extremophiles in chile and their potential for bioremediation. *Frontiers in Microbiology*. 1

Orth, J. D., Fleming, R. M. T., and Palsson, B. (2010). Reconstruction and use of microbial metabolic networks: the core escherichia coli metabolic model as an educational guide. *EcoSal Plus*, 4(1). 12

OSMA (2024). Planetary protection. 2

Passi, A., Tibocha-Bonilla, J., Kumar, M., Tec Campos, D., Zengler, K., and Zúñiga, C. (2021). Genome-scale metabolic modeling enables in-depth understanding of big data. *Metabolites*, 12:14. 5

Persson, E. (2012). The moral status of extraterrestrial life. *Astrobiology*, 12:976–984. 1

Rummel, J., Race, M., and Participants, G. (2012). Ethical considerations for planetary protection in space exploration: A workshop. *Astrobiology*, 12. 1

Sajed, T., Marcu, A., Ramirez, M., Pon, A., Guo, A. C., Knox, C., Wilson, M., Grant, J. R., Djoumbou, Y., and Wishart, D. S. (2016). ECMDB 2.0: A richer resource for understanding the biochemistry of E. coli. *Nucleic Acids Res*, 44(D1):495–501. 18

Schiwon, K., Arends, K., Rogowski, K., Fürch, S., Prescha, K., Sakinc, T., Van Houdt, R., Werner, G., and Grohmann, E. (2013). Comparison of antibiotic resistance, biofilm formation and conjugative transfer of staphylococcus and enterococcus isolates from international space station and antarctic research station concordia. *Microbial ecology*, 65. 35

Segrè, D., Vitkup, D., and Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23):15112–15117. 3, 5, 6, 14

Smith, J. M., Van Ness, H. C., and Abbott, M. M. (2005). *Introduction to Chemical Engineering Thermodynamics*. McGraw-Hill Education, 7th edition edition. 8

Stock, J., Rauch, B., and Roseman, S. (1977). Periplasmic space in salmonella typhimurium and escherichia coli. *Journal of Biological Chemistry*, 252(21):7850–7861. 19

Strocchi, M., Ferrer, M., Timmis, K., and Golyshin, P. (2006). Low temperature-induced systems failure in escherichia coli: Insights from rescue by cold-adapted chaperones. *PROTEOMICS*, 6(1):193–206. 25, 30

Takano, S., Takahashi, H., Yama, Y., Miyazaki, R., Furusawa, C., and Tsuru, S. (2023). Inference of transcriptome signatures of escherichia coli in long-term stationary phase. *Scientific Reports*, 13(1). 31

Torchinsky, R. (2022). Elon musk hints at a crewed mission to mars in 2029. *npr*. 1

Tribou, R. (2022). With success of artemis i, when will nasa fly artemis ii? *Phys.org*. 1

Tristan, C., Shahani, N., Sedlak, T. W., and Sawa, A. (2011). The diverse functions of gapdh: Views from different subcellular compartments. *Cellular Signalling*, 23(2):317–323. 29

Williams, J.-P., Paige, D., Greenhagen, B., and Sefton-Nash, E. (2017). The global surface temperatures of the moon as measured by the diviner lunar radiometer experiment. *Icarus*, 283:300–325. 31

Appendix A

Flux Solution Space Calculation

Consider a simple metabolic model containing five metabolites $\mathrm{X}_i,\ i = 1, 2, 3, 4, 5$ and four reaction processes $r_i,\ i = 1, 2, 3, 4$ and three transport processes between two compartments, *exterior* and *cytoplasm*, represented in Figure A.1. These processes are defined with v_i as the flux rate of the i-th reaction process and b_i of the i-th transport process. The following process and mass-balance expressions can be derived.

$$
\begin{aligned}
r_1 &: \ \mathrm{X}_1 \overset{v_1}{\rightleftharpoons} \mathrm{X}_2 \\
r_2 &: \ \mathrm{X}_1 \overset{v_2}{\rightarrow} \mathrm{X}_3 \\
r_3 &: \ \mathrm{X}_2 + \mathrm{X}_3 \overset{v_3}{\rightleftharpoons} \mathrm{X}_4 \\
r_4 &: \ \mathrm{X}_3 \overset{v_4}{\rightarrow} 2\mathrm{X}_5 \\
t_1 &: \ \overset{b_1}{\rightarrow} \mathrm{X}_1 \\
t_2 &: \ \mathrm{X}_4 \overset{b_2}{\rightleftharpoons} \\
t_3 &: \ \mathrm{X}_5 \overset{b_3}{\rightarrow}
\end{aligned}
\qquad
\begin{aligned}
\frac{\mathrm{d}\mathrm{X}_1}{\mathrm{d}t} &= -v_1 - v_2 + b_1 \\
\frac{\mathrm{d}\mathrm{X}_2}{\mathrm{d}t} &= v_1 - v_3 \\
\frac{\mathrm{d}\mathrm{X}_3}{\mathrm{d}t} &= v_2 - v_3 - v_4 \\
\frac{\mathrm{d}\mathrm{X}_4}{\mathrm{d}t} &= v_3 - b_2 \\
\frac{\mathrm{d}\mathrm{X}_5}{\mathrm{d}t} &= 2v_4 - b_3
\end{aligned}
$$

And the system is written in matrix notation where $\mathbf{M}$ is the concentration vector, $\mathbf{S}$ the stoichiometric matrix and $\vec{\mathbf{v}}$ the vector of unknown fluxes.

$$
\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathrm{X}_1 \\ \mathrm{X}_2 \\ \mathrm{X}_3 \\ \mathrm{X}_4 \\ \mathrm{X}_5 \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \ \Leftrightarrow \ \frac{\mathrm{d}M}{\mathrm{d}t} = S \cdot \vec{v}
$$

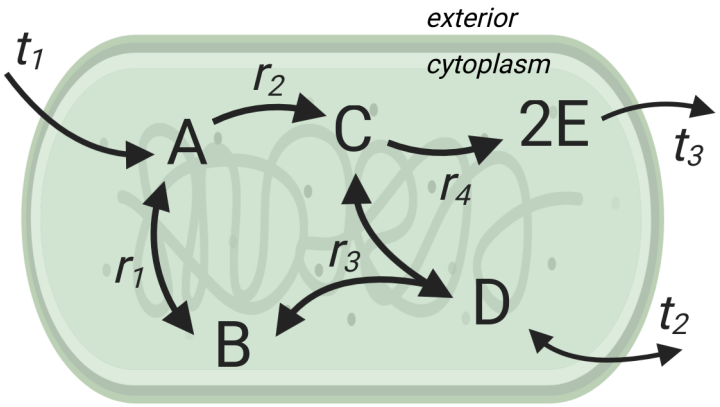


Figure A.1: Example of a simple metabolic model of a microorganism with five metabolites, four reaction and three transport processes, between two compartments.

Note that **S** represents the totality of the network reactions. The rows represent the compounds of the reactions, while the columns correspond to the reactions themselves. The stoichiometry itself is the quantitative relationship between substrates of a chemical reaction.

Upon reaching a steady state, the change in compound concentration over time is null, such that $S \cdot \vec{v} = \vec{0}$. Solving this equation, the solution space is defined.

Appendix B

Gibbs Free Energy Variation Temperature Dependence

The question is what is the change of Gibbs energy of a reaction at a given temperature T. If the reactants and products are all in their thermodynamic standard states then:

$$\Delta G_T^{\ominus} = \Delta H_T^{\ominus} - T\Delta S_T^{\ominus} \tag{B.1}$$

However if the enthalpic and entropic changes at a given temperature are unknown, the calculation is impossible. In this work, it was possible the access to two different ways to retrieve information that could help solve this problem.

The first of them was the `eQuilibrator` library that estimates reaction Gibbs free energy change at room temperature, i.e 298K. The other is a table with metabolite Gibbs free energy and enthalpic change [Hammes and Hammes-Schiffer, 2015]. Both have their pros and cons. `eQuilibrator` is a method that estimates for a wide range of reactions in an easy and rapid manner using numerical and statistical algorithms, providing coherent and *smooth* values. However it is prone to estimation errors and yields just one of the parameters. The paper table provides actual measurements of Gibbs free energy *and* enthalpy changes, so that the calculation of entropic change is possible. On the other hand, it is very limited and the values are for metabolites, i.e not reactions, so further calculations and estimations for metabolites not included in the dataset are needed and, once again, these are prone to errors. It is worthy of note that, upon calculating the reaction energies, the values of the dataset present a great variation. This variation can be due to the posterior estimations needed to calculate the values.

Considering the last paragraph, two considerations were made. The first is that the changes in enthalpy and entropy in different temperatures remain approximately equal. Note that this assumption is sufficient to calculate Equation B.1 using the paper's data. However, as supercited, the values showed inconsistency. Besides that, ignoring a developed method for estimation as `eQuilibrator` would under-

mine this research. So a second assumption is made. This is that the ratio between enthalpic and entropic changes remains constant and equal to the values in the paper. This way, the measurements of the paper's table, ally with the estimation power of the library and produce smooth values for the next calculations.

First the assumptions are defined mathematically.

$$\text{Assumption 1:} \quad \Delta H_T^{\ominus} \approx \Delta H_0^{\ominus} \text{ and } \Delta S_T^{\ominus} \approx \Delta S_0^{\ominus} \tag{B.2}$$

$$\text{Assumption 2:} \quad \frac{\Delta H_{paper}^{\ominus}}{\Delta S_{paper}^{\ominus}} = \frac{\Delta H_0^{\ominus}}{\Delta S_0^{\ominus}} \tag{B.3}$$

Then the calculations can depart from what is known of the provided data. The objective of these are to calculate $\Delta H_T^{\ominus}$ and $\Delta S_T^{\ominus}$ to substitute in Equation B.1.

$$\begin{cases} \Delta G_{eQuilibrator}^{\ominus} = \Delta H_0^{\ominus} - T_0 \Delta S_0^{\ominus} \\ \Delta G_{paper}^{\ominus} = \Delta H_{paper}^{\ominus} - T_0 \Delta S_{paper}^{\ominus} \end{cases} \overset{B.3}{\Leftrightarrow} \begin{cases} \Delta S_0^{\ominus} = \frac{1}{T_0}(\Delta H_0^{\ominus} - \Delta G_{eQuilibrator}^{\ominus}) \\ \Delta G_{paper}^{\ominus} = \Delta H_{paper}^{\ominus}\left(1 - T_0 \frac{\Delta S_0^{\ominus}}{\Delta H_0^{\ominus}}\right) \end{cases} \Leftrightarrow$$

$$\begin{cases} \Delta S_0^{\ominus} = \frac{1}{T_0}(\Delta H_0^{\ominus} - \Delta G_{eQuilibrator}^{\ominus}) \\ \Delta H_0^{\ominus} = \frac{\Delta H_{paper}^{\ominus}}{\Delta G_{paper}^{\ominus}} \Delta G_{eQuilibrator}^{\ominus} \end{cases} \Leftrightarrow \begin{cases} \Delta S_0^{\ominus} = \frac{1}{T_0}(\frac{\Delta H_{paper}^{\ominus}}{\Delta G_{paper}^{\ominus}} - 1)\Delta G_{eQuilibrator}^{\ominus} \\ \Delta H_0^{\ominus} = \frac{\Delta H_{paper}^{\ominus}}{\Delta G_{paper}^{\ominus}} \Delta G_{eQuilibrator}^{\ominus} \end{cases} \tag{B.4}$$

The ratio between enthalpic and Gibbs free energy in this work is calculated by the values in the paper and further estimations. However, future researches can possess more precise values and this calculations recycled so the ratio is defined as an adimensional $\phi_0 \equiv \frac{\Delta H_{paper}^{\ominus}}{\Delta G_{paper}^{\ominus}}$. By the assumption B.2, the system of equations B.4 is accounted for in Equation B.1, and the final Gibbs free energy change dependence on temperature equation is defined as follows:

$$\Delta G_T^{\ominus}(T) = \left(\phi_0 - \frac{T}{T_0}\phi_0 + \frac{T}{T_0}\right)\Delta G_{eQuilibrator}^{\ominus} \tag{B.5}$$

Using equation Equation 2.4, the dependence of equilibrium constants on temperature is calculated by Equation B.6. Defining $\Delta G_{eQuilibrator}^{\ominus} \equiv \Delta G_e^{\ominus}$:

$$K_T^{\ominus} = exp\left(-\frac{\Delta G_T^{\ominus}}{RT}\right) = exp\left(\phi_0 \frac{\Delta G_e^{\ominus}}{RT_0} - \phi_0 \frac{\Delta G_e^{\ominus}}{RT} - \frac{\Delta G_e^{\ominus}}{RT_0}\right) \tag{B.6}$$

Knowing that the $K_0^{\ominus} = exp\left(-\frac{\Delta G_0^{\ominus}}{RT_0}\right)$ and the estimation for $\Delta G_0^{\ominus} = \Delta G_e^{\ominus}$, then it is seen that the equilibrium constant at a given temperature depends on the equilibrium constant at the known temperature T_0. This observation is important for the low concentration variation approximation at subsection 3.2.2.

$$K_T^{\ominus}(T) = exp\left[\phi_0\left(\frac{1}{T_0} - \frac{1}{T}\right)\frac{\Delta G_e^{\ominus}}{R}\right]K_0^{\ominus} \tag{B.7}$$

Appendix C

Model Reactions and Metabolites

Table C.1: Our core model reaction's characteristics

Id	Name	Weight	$\Delta G_0^{\ominus}$ (kJ/mol)	ϕ_0
ACALD	Acetaldehyde dehydrogenase (acetylating)	0.000	-3.56	-0.51
ACKr	Acetate kinase	0.000	13.99	-0.46
ACONTa	Aconitase (half-reaction A, Citrate hydro-lyase)	1.000	8.47	109.71
ACONTb	Aconitase (half-reaction B, Isocitrate hydro-lyase)	1.000	-1.62	-159.25
ADK1	Adenylate kinase	0.000	0.33	-0.31
AKGDH	2-Oxogluterate dehydrogenase	0.018	-28.21	4.48
ALCD2x	Alcohol dehydrogenase (ethanol)	0.000	18.06	2.07
ATPM	ATP maintenance requirement	0.000	-46.76	0.64
ATPS4r	ATP synthase (four protons for one ATP)	0.572	46.76	0.64
CS	Citrate synthase	1.000	-38.76	5.96
CYTBD	Cytochrome oxidase bd (ubiquinol-8: 2 protons)	0.758	-85.01	1.71
ENO	Enolase	1.000	-3.81	-0.46
FBA	Fructose-bisphosphate aldolase	0.194	6.05	9.77
FBP	Fructose-bisphosphatase	0.000	-28.95	0.00
FRD7	Fumarate reductase	0.000	73.61	18.77
FUM	Fumarase	0.068	-3.42	-212.60
G6PDH2r	Glucose 6-phosphate dehydrogenase	0.012	-6.82	-0.44
GAPD	Glyceraldehyde-3-phosphate dehydrogenase	1.000	18.28	3.70
GLNS	Glutamine synthetase	1.000	-15.30	0.10
GLUDy	Glutamate dehydrogenase (NADP)	0.026	16.30	21.14
GLUN	Glutaminase	0.000	-31.46	1.57
GLUSy	Glutamate synthase (NADPH)	0.000	-47.76	16.12

Id	Name	Weight	$\Delta G_0^{\ominus}$ (kJ/mol)	ϕ_0
GND	Phosphogluconate dehydrogenase	0.012	-6.80	1.16
ICDHyr	Isocitrate dehydrogenase (NADP)	1.000	-11.71	57.72
ICL	Isocitrate lyase	0.000	-7.58	-5.00
LDH_D	D-lactate dehydrogenase	0.000	20.63	2.52
MALS	Malate synthase	0.000	-38.65	-2.32
MDH	Malate dehydrogenase	0.055	26.50	-1.08
ME1	Malic enzyme (NAD)	0.000	-5.01	542.61
ME2	Malic enzyme (NADP)	0.000	-5.01	587.92
NADH16	NADH dehydrogenase (ubiquinone-8 & 3 protons)	0.758	-131.76	-0.51
NADTRHD	NAD transhydrogenase	0.000	-0.00	40.09
PDH	Pyruvate dehydrogenase	0.088	-34.23	-1.00
PFK	Phosphofructokinase	0.194	-17.81	0.64
PFL	Pyruvate formate lyase	0.000	-19.81	-0.67
PGI	Glucose-6-phosphate isomerase	0.012	2.64	534.59
PGK	Phosphoglycerate kinase	1.000	19.55	1.62
PGL	6-phosphogluconolactonase	0.012	-25.38	1.84
PGM	Phosphoglycerate mutase	1.000	-4.47	0.00
PPC	Phosphoenolpyruvate carboxylase	0.004	-40.27	22.90
PPCK	Phosphoenolpyruvate carboxykinase	0.000	-6.49	10.78
PPS	Phosphoenolpyruvate synthase	0.000	-22.06	0.67
PTAr	Phosphotransacetylase	0.000	8.67	6.40
PYK	Pyruvate kinase	0.010	-25.03	0.64
RPE	Ribulose 5-phosphate 3-epimerase	0.009	-3.39	12.63
RPI	Ribose-5-phosphate isomerase	1.000	4.47	1.69
SUCDi	Succinate dehydrogenase (irreversible)	0.068	-73.61	18.77
SUCOAS	Succinyl-CoA synthetase (ADP-forming)	0.018	-1.21	1.62
TALA	Transaldolase	0.010	-0.94	11.23
THD2	NAD(P) transhydrogenase	0.000	0.00	40.09
TKT1	Transketolase	0.010	-1.50	-0.42
TKT2	Transketolase	0.008	-10.30	11.99
TPI	Triose-phosphate isomerase	0.194	5.62	0.00

Table C.2: Our core model metabolite's characteristics

Id	Name	Concentration (mM)
13dpg_c	3-Phospho-D-glyceroyl phosphate	0.200
2pg_c	D-Glycerate 2-phosphate	0.224
3pg_c	3-Phospho-D-glycerate	0.200
6pgc_c	6-Phospho-D-gluconate	1.021
6pgl_c	6-phospho-D-glucono-1,5-lactone	0.200
ac_c	Acetate	0.770
acald_c	Acetaldehyde	0.195
accoa_c	Acetyl-CoA	0.421
acon_C_c	Cis-Aconitate	0.042
actp_c	Acetyl phosphate	1.226
adp_c	ADP C10H12N5O10P2	0.312
akg_c	2-Oxoglutarate	0.200
amp_c	AMP C10H12N5O7P	0.167
atp_c	ATP C10H12N5O13P3	7.586
cit_c	Citrate	7.399
co2_c	CO2 CO2	1.200
coa_c	Coenzyme A	2.746
dhap_c	Dihydroxyacetone phosphate	0.355
e4p_c	D-Erythrose 4-phosphate	0.001
etoh_c	Ethanol	0.200
f6p_c	D-Fructose 6-phosphate	0.055
fdp_c	D-Fructose 1,6-bisphosphate	0.200
for_c	Formate	0.198
fum_c	Fumarate	0.212
g3p_c	Glyceraldehyde 3-phosphate	0.234
g6p_c	D-Glucose 6-phosphate	0.680
gln__L_c	L-Glutamine	3.940
glu__L_c	L-Glutamate	96.600
glx_c	Glyoxylate	0.200
h2o_c	H2O H2O	41 000
h_c	H+	0.010
h_e	H+	0.010
icit_c	Isocitrate	0.200
lac__D_c	D-Lactate	0.202

Id	Name	Concentration (mM)
mal__L_c	L-Malate	2.820
nad_c	Nicotinamide adenine dinucleotide	3.605
nadh_c	Nicotinamide adenine dinucleotide - reduced	0.115
nadp_c	Nicotinamide adenine dinucleotide phosphate	0.002
nadph_c	Nicotinamide adenine dinucleotide phosphate - reduced	0.235
nh4_c	Ammonium	1.000
o2_c	O2 O2	0.047
oaa_c	Oxaloacetate	0.032
pep_c	Phosphoenolpyruvate	0.660
pi_c	Phosphate	1.000
pyr_c	Pyruvate	8.050
q8_c	Ubiquinone-8	0.202
q8h2_c	Ubiquinol-8	0.200
r5p_c	Alpha-D-Ribose 5-phosphate	1.008
ru5p__D_c	D-Ribulose 5-phosphate	1.092
s7p_c	Sedoheptulose 7-phosphate	0.200
succ_c	Succinate	0.877
succoa_c	Succinyl-CoA	0.687
xu5p__D_c	D-Xylulose 5-phosphate	1.008

Nature is ever without action; yet nothing is left undone

Tao Te Ching 37