Detailed Biochemical Modelling and Analysis
Methodologies for Industrial Biotechnology

A thesis submitted to The University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

2015
Liliana Angeles Martinez

School of Chemical Engineering and Analytical Science
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>4</td>
</tr>
<tr>
<td>List of Tables</td>
<td>5</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>6</td>
</tr>
<tr>
<td>Nomenclature</td>
<td>10</td>
</tr>
<tr>
<td>Abstract</td>
<td>15</td>
</tr>
<tr>
<td>Declaration</td>
<td>17</td>
</tr>
<tr>
<td>Copyright Statement</td>
<td>18</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>19</td>
</tr>
<tr>
<td><strong>Chapter 1. Introduction and Research Contribution</strong></td>
<td>20</td>
</tr>
<tr>
<td>1.1. Biotechnology in industrial processes</td>
<td>20</td>
</tr>
<tr>
<td>1.1.1. Bio-production of succinic acid from glycerol, a by-</td>
<td>20</td>
</tr>
<tr>
<td>product of the biodiesel production</td>
<td></td>
</tr>
<tr>
<td>1.2. Intracellular environment and its influence on the</td>
<td>25</td>
</tr>
<tr>
<td>metabolism</td>
<td></td>
</tr>
<tr>
<td>1.3. Mathematical representation of the metabolic systems</td>
<td>29</td>
</tr>
<tr>
<td>and the importance of modelling</td>
<td></td>
</tr>
<tr>
<td>1.4. Thesis objective and research contribution</td>
<td>31</td>
</tr>
</tbody>
</table>
## Chapter 2. Literature Review

### 2.1. Unstructured models

### 2.2. Structured models
- 2.2.1. Deterministic models
- 2.2.2. Stochastic models

### 2.3. Concluding remarks

## Chapter 3. Steady State Metabolic Network Analysis: a Thermodynamically Constrained Stoichiometric Model

### 3.1. Preamble

### 3.2. Publication 1
- 3.2.1. Supplementary Material for Publication 1

## Chapter 4. Crowding Conditions in the Thermodynamic Analysis of a Metabolic Network

### 4.1. Preamble

### 4.2. Publication 2
- 4.2.1. Supplementary Material for Publication 2

## Chapter 5. Crowding Conditions in the Intracellular Diffusion Process

### 5.1. Preamble

### 5.2. Publication 3

## Chapter 6. Conclusions and Future Work

### 6.1. Conclusions

### 6.2. Future work

## References

## Appendix. Scaled Particle Theory
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Transesterification reaction of a triglyceride producing biodiesel and glycerol</td>
<td>21</td>
</tr>
<tr>
<td>1.2</td>
<td>Schematic representation of the whole process for biodiesel production</td>
<td>22</td>
</tr>
<tr>
<td>1.3</td>
<td>Schematic representation of the spatio-temporal distribution of the cellular components</td>
<td>26</td>
</tr>
<tr>
<td>1.4</td>
<td>Available volume for a spherical test molecule of radius $r_T$ in a solution of background macromolecules of radii $r_B$</td>
<td>28</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2.1</td>
<td>Some simulator for the reaction dynamics</td>
<td>51</td>
</tr>
</tbody>
</table>
Abbreviations

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSA</td>
<td>Anomalous SSA</td>
</tr>
<tr>
<td>CA</td>
<td>Cellular Automata method</td>
</tr>
<tr>
<td>DCW</td>
<td>Dry Cell Weight</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EBA</td>
<td>Energy Balance Analysis</td>
</tr>
<tr>
<td>FBA</td>
<td>Flux Balance Analysis</td>
</tr>
<tr>
<td>FVA</td>
<td>Flux Variability Analysis</td>
</tr>
<tr>
<td>GCM</td>
<td>Group Contribution Method</td>
</tr>
<tr>
<td>kMC</td>
<td>kinetic Monte Carlo algorithm</td>
</tr>
<tr>
<td>LBM</td>
<td>Lattice Boltzmann Method</td>
</tr>
<tr>
<td>MC</td>
<td>Monte Carlo method</td>
</tr>
<tr>
<td>MILP</td>
<td>Mixed Integer Linear Problem</td>
</tr>
<tr>
<td>NLP</td>
<td>Non-Linear Problem</td>
</tr>
<tr>
<td>ODE</td>
<td>Ordinary Differential Equation</td>
</tr>
<tr>
<td>PDE</td>
<td>Partial Differential Equation</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>SPT</td>
<td>Scaled Particle Theory</td>
</tr>
<tr>
<td>SSA</td>
<td>Stochastic Simulation Algorithm</td>
</tr>
<tr>
<td>tFVA</td>
<td>Thermodynamic based Flux Variability Analysis</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricarboxilic acid cycle</td>
</tr>
<tr>
<td>TMFA</td>
<td>Thermodynamics-based Metabolic Flux Analysis</td>
</tr>
</tbody>
</table>

Metabolites

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>13dpg</td>
<td>3-Phospho-D-glyceroyl phosphate</td>
</tr>
<tr>
<td>2pg</td>
<td>D-Glycerate 2-phosphate</td>
</tr>
<tr>
<td>3pg</td>
<td>3-Phospho-D-glycerate</td>
</tr>
<tr>
<td>6pgc</td>
<td>6-Phospho-D-gluconate</td>
</tr>
<tr>
<td>6pgl</td>
<td>6-phospho-D-glucono-1,5-lactone</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>ac</td>
<td>Acetate</td>
</tr>
<tr>
<td>acald</td>
<td>Acetaldehyde</td>
</tr>
<tr>
<td>accoa</td>
<td>Acetyl-CoA</td>
</tr>
<tr>
<td>actp</td>
<td>Acetyl phosphate</td>
</tr>
<tr>
<td>adp</td>
<td>Adenosine diphosphate</td>
</tr>
<tr>
<td>atp</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>co2</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>coa</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>dhap</td>
<td>Dihydroxyacetone phosphate</td>
</tr>
<tr>
<td>e4p</td>
<td>Erythrose 4-phosphate</td>
</tr>
<tr>
<td>etoh</td>
<td>Ethanol</td>
</tr>
<tr>
<td>f6p</td>
<td>Fructose 6-phosphate</td>
</tr>
<tr>
<td>fdp</td>
<td>Fructose 1,6-bisphosphate</td>
</tr>
<tr>
<td>for</td>
<td>Formate</td>
</tr>
<tr>
<td>fum</td>
<td>Fumarate</td>
</tr>
<tr>
<td>g3p</td>
<td>Glyceraldehyde 3-phosphate</td>
</tr>
<tr>
<td>g6p</td>
<td>Glucose 6-phosphate</td>
</tr>
<tr>
<td>glyc</td>
<td>Glycerol</td>
</tr>
<tr>
<td>glyc3p</td>
<td>Glycerol 3-phosphate</td>
</tr>
<tr>
<td>h2o</td>
<td>Water</td>
</tr>
<tr>
<td>lac</td>
<td>Lactate</td>
</tr>
<tr>
<td>mal</td>
<td>Malate</td>
</tr>
<tr>
<td>nad</td>
<td>Nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>nadh</td>
<td>Nicotinamide adenine dinucleotide - reduced</td>
</tr>
<tr>
<td>nadp</td>
<td>Nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>nadph</td>
<td>Nicotinamide adenine dinucleotide phosphate - reduced</td>
</tr>
<tr>
<td>oaa</td>
<td>Oxaloacetate</td>
</tr>
<tr>
<td>pep</td>
<td>Phosphoenolpyruvate</td>
</tr>
<tr>
<td>pi</td>
<td>Phosphate</td>
</tr>
<tr>
<td>pyr</td>
<td>Pyruvate</td>
</tr>
<tr>
<td>r5p</td>
<td>Ribose 5-phosphate</td>
</tr>
<tr>
<td>ru5p</td>
<td>Ribulose 5-phosphate</td>
</tr>
<tr>
<td>s7p</td>
<td>Sedoheptulose 7-phosphate</td>
</tr>
</tbody>
</table>
succ  Succinate
xu5p  Xylulose 5-phosphate

**Reactions**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACALD</td>
<td>Acetaldehyde dehydrogenase</td>
</tr>
<tr>
<td>ACKr</td>
<td>Acetate kinase</td>
</tr>
<tr>
<td>ACYP</td>
<td>Acetyl phosphate phosphohydrolase</td>
</tr>
<tr>
<td>ALCD2x</td>
<td>Alcohol dehydrogenase</td>
</tr>
<tr>
<td>ATPm</td>
<td>ATP requirements for other reactions</td>
</tr>
<tr>
<td>ENO</td>
<td>Enolase</td>
</tr>
<tr>
<td>EX_ac(e)</td>
<td>Acetate exchange</td>
</tr>
<tr>
<td>EX_co2(e)</td>
<td>CO₂ exchange</td>
</tr>
<tr>
<td>EX_coa(e)</td>
<td>Coenzyme A exchange</td>
</tr>
<tr>
<td>EX_etoh(e)</td>
<td>Ethanol exchange</td>
</tr>
<tr>
<td>EX_for(e)</td>
<td>Formate exchange</td>
</tr>
<tr>
<td>EX_fum(e)</td>
<td>Fumarate exchange</td>
</tr>
<tr>
<td>EX_glyc(e)</td>
<td>Glycerol exchange</td>
</tr>
<tr>
<td>EX_h2o(e)</td>
<td>Water exchange</td>
</tr>
<tr>
<td>EX_lac(e)</td>
<td>Lactate exchange</td>
</tr>
<tr>
<td>EX_oaa(e)</td>
<td>Oxaloacetate exchange</td>
</tr>
<tr>
<td>EX_pi(e)</td>
<td>Phosphate exchange</td>
</tr>
<tr>
<td>EX_pyr(e)</td>
<td>Pyruvate exchange</td>
</tr>
<tr>
<td>EX_succ(e)</td>
<td>Succinate exchange</td>
</tr>
<tr>
<td>FBAr</td>
<td>Fructose-bisphosphate aldolase</td>
</tr>
<tr>
<td>FBP</td>
<td>Fructose-1,6-bisphosphatase</td>
</tr>
<tr>
<td>FDH3</td>
<td>Formate dehydrogenase</td>
</tr>
<tr>
<td>FRD</td>
<td>Fumarate reductase</td>
</tr>
<tr>
<td>FUM</td>
<td>Fumarase</td>
</tr>
<tr>
<td>G3PD2</td>
<td>Glycerol 3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>G6PDH2r</td>
<td>Glucose 6-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GAPD</td>
<td>3-Phosphate glyceraldehyde dehydrogenase</td>
</tr>
<tr>
<td>GLYK</td>
<td>Glycerol kinase</td>
</tr>
<tr>
<td>GND</td>
<td>Phosphogluconate dehydrogenase</td>
</tr>
<tr>
<td>Acronym</td>
<td>Name</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>LDH</td>
<td>Lactate dehydrogenase</td>
</tr>
<tr>
<td>MDH</td>
<td>Malate dehydrogenase</td>
</tr>
<tr>
<td>ME2</td>
<td>Malic enzyme</td>
</tr>
<tr>
<td>NADHm</td>
<td>NADH requirements for other reactions</td>
</tr>
<tr>
<td>NADPHm</td>
<td>NADPH requirements for other reactions</td>
</tr>
<tr>
<td>OAD</td>
<td>Oxaloacetate decarboxylase</td>
</tr>
<tr>
<td>PDH</td>
<td>Pyruvate dehydrogenase complex</td>
</tr>
<tr>
<td>PGI</td>
<td>Glucose-6-phosphate isomerase</td>
</tr>
<tr>
<td>PGK</td>
<td>3-Phosphoglycerate kinase</td>
</tr>
<tr>
<td>PGL</td>
<td>6-Phosphogluconolactonase</td>
</tr>
<tr>
<td>PGM</td>
<td>Phosphoglycerate mutase</td>
</tr>
<tr>
<td>PFK</td>
<td>6-Phosphofructokinase</td>
</tr>
<tr>
<td>PFL</td>
<td>Pyruvate formate lyase</td>
</tr>
<tr>
<td>PPCK</td>
<td>Phosphoenolpyruvate carboxykinase</td>
</tr>
<tr>
<td>PTAr</td>
<td>Acetate phosphotransferase</td>
</tr>
<tr>
<td>PYK</td>
<td>Pyruvate kinase</td>
</tr>
<tr>
<td>RPE</td>
<td>Ribulose 5-phosphate 3-epimerase</td>
</tr>
<tr>
<td>RPI</td>
<td>Ribose-5-phosphate isomerase</td>
</tr>
<tr>
<td>TALA</td>
<td>Transaldolase</td>
</tr>
<tr>
<td>TKT1</td>
<td>Transketolase</td>
</tr>
<tr>
<td>TKT2</td>
<td>Transketolase</td>
</tr>
<tr>
<td>THD2</td>
<td>NAD(P) transhydrogenase</td>
</tr>
<tr>
<td>TPI</td>
<td>Triose-phosphate isomerase</td>
</tr>
</tbody>
</table>
Nomenclature

Chapter 1 and 2

$a$ Activity of a metabolite
$C$ Concentration of an intracellular metabolite
$C$ Concentration vector of all intracellular metabolites
$C^s$ Standard concentration of an intracellular metabolite
$h$ Spectral dimension
$I$ Ionic strength
$k_i$ Rate constant of reaction $i$
$k_i^0$ Rate constant of reaction $i$ at initial time
$K_M$ Michaelis-Menten constant
$pH$ Potential of hydrogen
$r$ Radii of the molecules
$R$ Gas constant equivalent
$S$ Stoichiometric matrix of the metabolic network
$T$ Temperature
$v$ Rate or flux of reaction
$v$ Vector of the fluxes through the metabolic network
$V_{max}$ Maximum rate attainable
$\gamma$ Activity coefficient of a metabolite
$\Gamma_M$ The non-ideal factor
$\Delta_f G^0$ Standard Gibbs free energy of formation of a metabolite
$\Delta_r G$ Gibbs free energy of a reaction
$\phi$ Volume occupied by the background molecules

Chapter 3 and 4

$a$ Activity of a metabolite [dimensionless]
$A$ Constant of the extended Debye-Hückel equation equivalent to 0.510651 $[L^{-0.5} \text{mol}^{-0.5}]$
**Nomenclature**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B$</td>
<td>Constant of the extended Debye-Hückel equation equivalent to $1.6 , [\text{L}^{-0.5} , \text{mol}^{-0.5}]$</td>
</tr>
<tr>
<td>$c$</td>
<td>Subscript that indicates intracellular medium [dimensionless]</td>
</tr>
<tr>
<td>$C$</td>
<td>Concentration of an intracellular metabolite $[\text{mol} , \text{L}^{-1}]$</td>
</tr>
<tr>
<td>$C$</td>
<td>Concentration vector of all intracellular metabolites $[\text{mol} , \text{L}^{-1}]$</td>
</tr>
<tr>
<td>$C_{\text{max}}$</td>
<td>Maximum concentration of an intracellular metabolite $[\text{mol} , \text{L}^{-1}]$</td>
</tr>
<tr>
<td>$C_{\text{max}}$</td>
<td>Maximum concentration vector of all intracellular metabolites $[\text{mol} , \text{L}^{-1}]$</td>
</tr>
<tr>
<td>$C_{\text{min}}$</td>
<td>Minimum concentration of an intracellular metabolite $[\text{mol} , \text{L}^{-1}]$</td>
</tr>
<tr>
<td>$C_{\text{min}}$</td>
<td>Minimum concentration vector of all intracellular metabolites $[\text{mol} , \text{L}^{-1}]$</td>
</tr>
<tr>
<td>$C^{\text{st}}$</td>
<td>Standard concentration of an intracellular metabolite $[\text{mol} , \text{L}^{-1}]$</td>
</tr>
<tr>
<td>$d$</td>
<td>Parameter that indicates if a specific reaction is blocked (d=0) or at equilibrium (d=1) [dimensionless]</td>
</tr>
<tr>
<td>$d$</td>
<td>Parameter vector of all the reactions that indicates if a specific reaction is blocked (d=0) or at equilibrium (d=1) [dimensionless]</td>
</tr>
<tr>
<td>$e$</td>
<td>Subscript that indicates extracellular medium [dimensionless]</td>
</tr>
<tr>
<td>$E$</td>
<td>Error variable to consider the uncertainty in standard Gibbs free energy of the reaction estimated by the group contribution method $[\text{kcal} , \text{mol}^{-1}]$</td>
</tr>
<tr>
<td>$f$</td>
<td>Subscript that indicates the reaction used as objective function [dimensionless]</td>
</tr>
<tr>
<td>$\text{func}$</td>
<td>Objective function to optimise in the NLP formulation</td>
</tr>
<tr>
<td>$F$</td>
<td>Faraday constant equivalent to $2.306 \times 10^2 , [\text{kcal} , \text{mV}^{-1} , \text{mol}^{-1}]$</td>
</tr>
<tr>
<td>$i$</td>
<td>Subscript that indicates the pseudoisomer group of a metabolite $i$ [dimensionless]</td>
</tr>
<tr>
<td>$I$</td>
<td>Ionic strength $[\text{M}]$</td>
</tr>
<tr>
<td>$j$</td>
<td>Subscript that indicates the protonated species $j$ of a metabolite $i$ [dimensionless]</td>
</tr>
<tr>
<td>$k$</td>
<td>Subscript that indicates a specific reaction [dimensionless]</td>
</tr>
<tr>
<td>$m$</td>
<td>Number of intracellular metabolites in the metabolic network, the hydrogen ions $H^+$ is not included [dimensionless]</td>
</tr>
<tr>
<td>$M$</td>
<td>Molecular weight of metabolite $[\text{g} , \text{mol}^{-1}]$</td>
</tr>
<tr>
<td>$\text{M}$</td>
<td>Molecular weight vector of all metabolites $[\text{g} , \text{mol}^{-1}]$</td>
</tr>
<tr>
<td>$n$</td>
<td>Number of intracellular reactions in the metabolic network analysed [dimensionless]</td>
</tr>
</tbody>
</table>
Nomenclature

\(N\) Number of reactions for the transport of metabolites outside/inside the cell [dimensionless]

\(N_a\) Avogadro’s number equivalent to \(6.02214 \times 10^{23}\) [molecules \(\text{mol}^{-1}\)]

\(N_H\) Number of hydrogen atoms in a metabolite [dimensionless]

\(pH\) Potential of hydrogen [dimensionless]

\(pH_{ref}\) Reference \(pH = 7\) at standard biological conditions [dimensionless]

\(pK_a\) Acid dissociation constant [dimensionless]

\(r\) Radium of a molecule [cm]

\(r_{\text{random}}\) Random radii vector [cm]

\(R\) Gas constant equivalent to \(1.9858 \times 10^{-3}\) [kcal \(\text{mol}^{-1} \text{K}^{-1}\)]

\(S\) \(m \times (n+N)\) stoichiometric matrix of the metabolic network [dimensionless]

\(S_x\) Variable of the Scale Particle Theory equation, where \(0 \leq x \leq 3\)

\(SE\) Uncertainty associated to the standard Gibbs free energy of the reaction estimated by the group contribution method [kcal \(\text{mol}^{-1}\)]

\(tol\) Tolerance value of what is considered a non-zero Gibbs free energy [kcal \(\text{mol}^{-1}\)]

\(T\) Temperature of the medium [K]

\(v\) \(n \times 1\) vector of the fluxes through the metabolic network [mmol L\(^{-1}\) \(g_{\text{DW}}^{-1}\)]

\(v_{\text{max}}\) Maximum flux through a reaction [mmol L\(^{-1}\) \(g_{\text{DW}}^{-1}\)]

\(v_{\text{max}}\) Maximum flux vector of all reactions [mmol L\(^{-1}\) \(g_{\text{DW}}^{-1}\)]

\(v_{\text{min}}\) Minimum flux through a reaction [mmol L\(^{-1}\) \(g_{\text{DW}}^{-1}\)]

\(v_{\text{min}}\) Minimum flux vector of all reactions [mmol L\(^{-1}\) \(g_{\text{DW}}^{-1}\)]

\(\text{vol}_{av}\) Available volume in the system [L]

\(\text{vol}_{tot}\) Total volume of the system [L]

\(z\) Electrical charge of the metabolite species \(j\) [dimensionless]

\(\Delta_j G_H\) Gibbs free energy of formation of the hydrogen [kcal \(\text{mol}^{-1}\)]

\(\Delta_j G^*\) Standard Gibbs free energy of formation of a metabolite [kcal \(\text{mol}^{-1}\)]

\(\Delta_j G^{**}\) Transformed Standard Gibbs free energy of formation of a metabolite [kcal \(\text{mol}^{-1}\)]

\(\Delta_j G^{***}\) Transformed standard Gibbs free energy of a reaction formation of a
Nomenclature

metabolite [kcal mol$^{-1}$]

$\Delta G'$ Transformed Gibbs free energy of a reaction [kcal mol$^{-1}$]

$\Delta G'$ Transformed Gibbs free energy vector of all intracellular reactions [kcal mol$^{-1}$]

$\Delta G_{tr}'$ Total transformed Gibbs free energy of a transport reaction of a metabolite [kcal mol$^{-1}$]

$\Delta H^0$ Standard enthalpy of formation of a metabolite [kcal mol$^{-1}$]

$\Delta H^0$ Standard enthalpy of the reaction [kcal mol$^{-1}$]

$\Delta \varphi$ Difference between the internal membrane potential and the external one [mV]

$\gamma$ Activity coefficient of a metabolite [dimensionless]

$\eta$ Stoichiometric coefficient of a metabolite in a reaction [dimensionless]

$\rho$ Density number [molecules cm$^{-3}$]

$\nu$ Specific volume [cm$^3$ g$^{-1}$]

Chapter 5 and Appendix A

$\alpha_{sp}$ Activity of the species $sp$ [dimensionless]

$A_{sp}$ Area of a molecule of the species $sp$ [nm$^2$]

$C_{sp}$ Concentration of the species $sp$ [molecules nm$^{-2}$]

$C_{sp}^*$ Standard concentration of the species $sp$ [molecules nm$^{-2}$]

$d$ Direction chosen by the molecules to jump to neighboring voxel [dimensionless]

$D_{sp}$ Diffusion coefficient [nm$^2$ ms$^{-1}$]

$D_{sp}^*$ Diffusion coefficient in dilute solutions [nm$^2$ ms$^{-1}$]

$error_{sp}$ Relative error of molecules’ distribution predicted by kMC–cLBM or kMC–cLBM of the species $sp$ [%]

$F_{d,sp}$ Distribution function of the species $sp$ in the direction $d$ predicted by cLBM [molecules per voxel]

$F_{d,sp}^*$ Equilibrium distribution function of the species $sp$ in the direction $d$ [molecules per voxel]
Nomenclature

\( F_{d,sp}^{LB} \) Distribution function of the species \( sp \) in the direction \( d \) predicted by LBM [molecules per voxel]

\( h \) Planck’s constant [nm\(^2\) g s\(^{-1}\)]

\( i \) Index that identify the position of a voxel [dimensionless]

\( j \) Index that identify the position of a voxel [dimensionless]

\( J_{sp} \) Diffusive flux of molecules \( sp \) in 2D [molecules nm\(^{-1}\) ms\(^{-1}\)]

\( k_B \) Boltzmann constant equivalent to \( 1.3806 \times 10^{-23} \) [J K\(^{-1}\)]

\( m_{sp} \) Mass of a molecule \( sp \) [g molecule\(^{-1}\)]

\( next \) Subscript that indicates the target voxel where the molecules will move in the \( t+\Delta t \) [dimensionless]

\( P_{sp} \) Probability to find available space for species \( sp \) in the target voxel [dimensionless]

\( r_{sp} \) Radium of the species \( sp \) [nm]

\( sp \) Index that identify the molecule species \( sp \) [dimensionless]

\( t \) Time [s]

\( T \) Temperature of the medium [K]

\( w_d \) Weight factor for the calculation of the equilibrium function [dimensionless]

\( \Delta t \) Time increment [s]

\( \Delta W \) Work required to free the target space from background molecules [J molecule\(^{-1}\)]

\( \Delta x \) Size of the voxel in which the lattice is divided [nm]

\( \gamma_{sp} \) Activity coefficient of a molecule \( sp \) [dimensionless]

\( \mu \) Chemical potential [J molecule\(^{-1}\)]

\( \Omega_{d,sp}^{non} \) Non-reactive collision term [molecules per voxel]

\( \omega_{sp} \) Relaxation parameter [dimensionless]

\( \rho_{sp} \) Macroscopic density of species \( sp \) in a voxel [g per voxel]

\( \rho_{sp} \) Matrix with the macroscopic density of species \( sp \) in all voxels of the lattice [g per voxel]
Abstract

Many industrial processes use biological agents as catalysts. In this context, the study of the cellular metabolism becomes relevant for planning the best strategies (environmental and/or genetic modifications) to manipulate the cell in order to maximise the production of a metabolite of interest and minimise the by-products one. This increases the yield of the fermentation and reduces the cost of product recovery; thereby the profitability of the process is improved.

The intracellular reactions are carried out in a complex, crowded and heterogeneous medium composed by solid components (macromolecules, ions, enzymes, small solutes, etc.) in a fluid phase called cytoplasm, all of them enclosed within the cellular membrane. The interactions among the intracellular components (as well as with the extracellular environment) determine the behaviour of the organism. The modelling and simulations of these interactions help the understanding of the metabolism. The aim of this thesis is to provide generic tools for the analysis and simulation of metabolic systems under the intracellular environmental conditions. In particular, this research focuses on the estimation of metabolic fluxes and the simulation of the diffusion process.

The stoichiometric models have been widely used for the calculation of unmeasured fluxes in a metabolic network, assuming the system is at steady state. The addition of thermodynamic constraints allows only the prediction of fluxes that go in the direction of the Gibbs free energy drop. The Gibbs free energy change ($\Delta_G$) depends on the (intracellular) environmental conditions and determine the direction, feasibility and reversibility of the reactions involved in the pathways. The thermodynamically constrained stoichiometric model proposed here allows the estimation of the range of fluxes of a metabolic network, where the information about the presence of the enzymes that catalyse the reactions can be incorporated (if available). The effect of considering a zero flux reaction as blocked or at equilibrium on the flux predictions was
investigated, as well as the environmental conditions ionic strength, temperature and pH.

Additionally, since the solid components within the cell occupy about 40% of its total volume, these crowding conditions could alter the thermodynamic feasibility of the pathways. For this reason, the thermodynamically constrained stoichiometric model is extended to incorporate the crowding effect. The case study used in this work is the central carbon metabolic network of Actinobacillus succinogenes for the production of succinic acid from glycerol, a by-product in the biodiesel manufacture.

Moreover, the crowding conditions also affect the diffusion of the molecules. The prokaryotic cells have been widely used in fermentation processes for the production of metabolites of interest. In this type of cells the diffusion is the primary mean of the particles’ motion, so that the diffusion reduction due to the crowding conditions could affect the possibility of encounter among the reactants, decreasing the reactions’ rate and therefore the yield of the process. A methodology based on the Lattice Boltzmann Method (LBM) and the Scaled Particle Theory (SPT) is presented in this thesis for fast simulations of the diffusion of hard-disk molecules in 2D crowded systems, which also allows evaluating the effect of the molecules’ size on their diffusion.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Liliana Angeles Martinez
Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
First, I would like to thank to my family, especially to my mother Esperanza and my siblings Daniela and Daniel, for their unconditional support. Thank you for everything.

I would like to express my gratitude to my supervisor Dr. Constantinos Theodoropoulos for his guidance, help and support through this research.

In addition, I want to thank to Dr. Edgar Salgado Manjarrez and Professor Dr. Alexander Gorban for their valuable comments and suggestions related to this work. Thanks also to Dr. Michael Binns for his computational advices, and to Dr. Anestis Vlysidis for providing his experimental results of the fermentation of Actinobacillus succinogenes.

This journey would not have been possible without the financial support of the National Council of Science and Technology of Mexico (CONACyT, No. 213602).

Finally, I want to thank to my colleagues of the Centre for Process Integration, especially to Lluvia for her comments, and to Ioannis for his help with Fortran codes. Thanks also to my friends Francisco, Gerardo, Estefania, Mary, and Miguel for their friendship during my stay in Manchester.
1.1. Biotechnology in industrial processes

Biotechnology can be defined as the use of living organisms, including microorganisms, fungi, plants, algae, or their cellular components, for the conversion of raw materials into beneficial products (materials, chemicals, and/or energy) for mankind (Nair, 2008).

Even though the biotechnology has been practiced since ancient times for the preparation of beverages, bread, yogurt and other fermented products, the new tools and technologies currently available allow monitoring, control and improvement of processes, e.g. genetic manipulation to change or enhance the capabilities of the organisms involved in such processes.

The applications of biotechnology in the industry are wide, ranging from the production and development of drugs, vaccines, cosmetics, detergents, chemicals, food products, gas and energy, the reduction of pollutants and waste generation, to the revalorisation of by-products given in other part of the process, etc. A clear example of the latter is the conversion of crude glycerol (a by-product of the biodiesel industry) into added-value chemicals, e.g. succinic acid.

1.1.1. Bio-production of succinic acid from crude glycerol, a by-product of the biodiesel production

With the increasing energy demand due to a growing world population, and the rising prices of the crude oil (projected as $150 per barrel in 2035, EIA, 2014), the
governments strive to exploit and support the use of alternative energy sources with the aim to reduce the dependence on the fossil fuels (petroleum, coal, and natural gas). It is expected that the fossil fuels cover the energy demand for at least the next 50 years, but eventually they will be depleted (Smil, 2005).

There is an increasing interest in the biomass-based production of alternative energy sources, e.g. bioethanol, biobutanol and biodiesel (Demirbas, 2009). In particular, the European Union has mainly supported the biodiesel production, generating in 2013 9.57 Mt (EU Biofuels Annual 2014).

Biodiesel is a compound made from renewable raw materials like vegetable oils and animal fats. It is typically synthesised through the transesterification of the oil or fat with methanol, forming also the by-product glycerol (Figure 1.1), which represent 10% (w/w) of the production of biodiesel (Da Silva et al., 2009; Yang et al., 2012).

Biodiesel combustion emissions contain smaller amount of carbon monoxide, CO₂ and unburnt hydrocarbons compared to those from fossil fuels. Besides, the CO₂ produced is balanced with the CO₂ consumed by the plant crops used as raw material for the biodiesel production (Vlysidis, 2011a; European Biodiesel Board, 2011). These advantageous features, besides the tax incentives, have contributed to increase the interest in the use and production of biodiesel worldwide.

One of the strategies proposed to improve the profitability of biodiesel production (Figure 1.2) is the purification and sale of the by-product glycerol, or its use as cheap
raw material for the synthesis of value added products such as 1,3-propanediol, citric acid, ethanol, propionic acid, hydrogen, lipids, lactic acid, succinic acid, among others, which also solves the problem of disposal of excess glycerol associated with the biodiesel production (Da Silva et al., 2009; Vlysidis et al., 2011a, b, c, d; Yang et al., 2012).

Succinic acid is a dicarboxylic acid used to produce additives for the food (e.g. flavours and antimicrobials) and pharmaceutical industries, detergents, surfactants, plant growth stimulants, biodegradable polymers, among others (Song and Lee, 2006; Da Silva et al., 2009; Beauprez et al., 2010).

Nowadays, succinate is mainly synthesised chemically from liquid petroleum gas (LPG) or petroleum oil, except for that used in the food industry which is produced by fermentation of sugars (Song and Lee, 2006; McKinlay and Vieille, 2008; Da Silva et al., 2009). However, the utilization of a biotechnology process rather than a chemical one has the advantages of reducing the energy cost related to the use of high pressures and temperatures, avoiding the multistep process, using renewable raw material and carbon dioxide (a greenhouse gas).

Microorganisms like Actinobacillus succinogenes, Mannheimia succiniciproducens, Anaerobiospirillum succiniciproducens, Corynebacterium glutamicum, Aspergillus spp., Fusarium spp., Penicillium simplicissimum and Saccharomyces cerevisae, and recombinant Escherichia coli, among others (Song and Lee, 2006; Beauprez et al., 2010; Blankschien et al., 2010), either wild strain or genetically modified, are capable
to synthesise succinate, an intermediate of the tricarboxilic acid cycle (TCA cycle), from a variety of substrates, most of them under anaerobic conditions.

The biochemical reactions that are active in a metabolic pathway, and used for synthesis of any metabolite, e.g. succinate, lactate, fumarate, etc., depend on the kind of microorganism and the environmental conditions.

In the past few years, our research group has been studying the bioconversion of the glycerol into succinic acid (Vlysidis et al., 2011a,b,c,d) by *Actinobacillus succinogenes*. With the aim that in the future the computational tools presented in this thesis contribute to improve the yields of this process, the *A. succinogenes*’ metabolic system is used as case study in this work.

*Actinobacillus succinogenes*

*A. succinogenes* is a capnophilic, mesophilic, Gram negative, facultative anaerobic bacterium isolated from bovine rumen, and member of the *Pasteurellaceae* family (Corona-Gonzalez et al., 2008; Beauprez et al., 2010).

Its ability to grow on a broad range of substrates such as glucose (McKinlay et al., 2007; McKinlay and Vieille, 2008; Li et al., 2010), lactose (Guettler et al., 1999), fructose (Van der Werf et al., 1997), sorbitol (Van der Werf et al., 1997; Li et al., 2010), mannitol (Van der Werf et al., 1997), corn fibre (Chen et al., 2010), glycerol (Vlysidis et al., 2011a; Carvalho et al., 2014), among others, synthesising high concentrations of succinate as the main product make this microorganism an interesting biocatalyst for the industrial production of succinic acid.

Moreover, kinetic studies show its tolerance to high concentrations of acid products (succinate, formate, acetate) and glucose, having found that the critical concentrations at which the growth was completely inhibited were 45 g L\(^{-1}\) mixture of acids and 158 g L\(^{-1}\) glucose (Lin et al., 2008; Corona-Gonzalez et al., 2008). This is beneficial for batch fermentations inasmuch as it is possible the use of high initial substrate concentration,
allowing also extend the fermentation time (where the microorganism is still growing despite the fact of the high concentration of acids produced).

\(^{13}\)C-labeling studies on the metabolic network of *A. succinogenes* suggest that two enzymes of the TCA cycle (isocitrate dehydrogenase and alfa-ketoglutarate dehydrogenase) are missing or are inactive (McKinlay et al., 2007; Beauprez et al., 2010), while the flux through the Entner-Doudoroff pathway is negligible and the glyoxylate cycle is not active during growth on glucose, in contrast to the Embden-Meyerhof-Parnas and Pentose Phosphate pathways whose activity was found (Van der Werf et al., 1997; McKinlay et al., 2007; McKinlay and Vieille, 2008).

Several strategies have been proposed to try to stimulate the synthesis of succinate (not all them with the expected results), for example:

- The use of more reduced substrates than glucose, like glycerol or sorbitol (Li et al., 2010, Vlysidis et al., 2011b), resulting in an increment of reduced products like succinate and ethanol.
- The addition of high concentrations NaHCO\(_3\) to the medium causing the fluxes through malic enzyme and OAA decarboxylase to decrease. This is because the synthesis of succinate needs reductants (NAD(P)H) which are produced during the formation of the by-products formate, acetate and ethanol (McKinlay and Vieille, 2008). Nevertheless, under limiting CO\(_2\) environment the production of ethanol is favoured rather than the succinate one, because there is no CO\(_2\) fixation during the formation of ethanol (Van der Werf et al., 1997).
- The addition of H\(_2\) to the medium, as the use of a reduced substrate, increases the flux to succinate due to the availability of the reducing power. However, the combination of high concentration of NaHCO\(_3\) and the addition of H\(_2\) to the medium stimulates more the succinate production (Van der Werf et al., 1997; Song and Lee, 2006; McKinlay and Vieille, 2008).

The experimental results obtained from strategies such as those listed above, together with the simulation of the metabolism provide a deeper understanding of the intracellular processes.
In general, the study and simulation of the metabolism of a microorganism, and the
factors that affect it such as the intracellular conditions, can help to determine the best
strategy (genetic modification and/or changes in the experimental conditions) to
maximise the production of the metabolite of interest and minimise the by-products
ones.

1.2. Intracellular environment and its influence on the metabolism

The cytoplasm is a heterogeneous fluid medium containing macromolecules, solutes,
metabolic ions, enzymes, skeletal proteins and membranes, whose concentrations are
not high (e.g. the concentration of protein, RNA, and DNA in *Escherichia coli* are 200
to 320 mg mL\(^{-1}\), 75 to 120 mg mL\(^{-1}\), and 11 to 18 mg mL\(^{-1}\), respectively (Elowitz et al.,
1998)) but all together occupy about 40% of the total volume of the cell. For this reason
the medium is referred as “crowded” rather than “concentrated” (Minton, 2001;
Chebotareva et al., 2004).

In eukaryotes different processes and reactions take place in cellular compartments
called organelles. Cytoskeletal networks and motor proteins are responsible for the
active transport of molecules and the cytoplasm mixing. In prokaryotes, however, there
are no such structures so the biochemical processes are carried out in the cytoplasm,
and the diffusion is the primary means of intracellular motion (Elowitz et al., 1998).

Nevertheless, a prokaryotic cell is not just a “bag” with enzymes and substrates
swimming freely. Due to the electrostatic interactions among ions and charged moieties
of molecules, the formation of macromolecular agglomerates and multienzymatic
complexes (Huthmacher et al., 2008; Zhang, 2011), e.g. metabolons, that may affect the
spatial distribution of the cell is possible.

The metabolons are non-covalent complexes of sequential enzymes of a metabolic
pathway, which protect unstable metabolites, facilitate the channelling of a metabolite
from the active site of one enzyme to another, and avoid the consumption of a
metabolite by competing pathways, among others (Huthmacher et al., 2008; Zhang,
Examples of metabolons are found in glycolysis (Kurganov et al., 1985), Krebbs cycle (Lyubarev and Kurganov, 1989) and oxidative pentose phosphate pathway (Debnam et al., 1997).

It has been suggested that these macromolecular agglomerates bring some order to the cytoplasm, i.e. help the formation of transient hyperstructures, super-crowded zones or subdomains where particular metabolic processes are carried out, followed by dilute zones or pools of metabolites and other low molecular weight species where they can easily diffuse and move from one region to another (Figure 1.3) (Amar et al., 2008; Spitzer and Poolman, 2009; Zhang, 2011).

These macromolecular association phenomena are related to the properties of the intracellular medium such as ionic strength, pH, crowding conditions, temperature.

The ionic strength ($I$) is a measure of the concentration of ions present in the medium, which also affects the conformation of proteins, the formation of micelles, the solubility of ions, the thermodynamic activity and diffusion of the charged species, among others.
(Spitzer and Poolman, 2009).

\( \text{pH} \) may modify the electric charge of ions and macromolecules by the protonation of a group or residues, affecting in this way the associations of DNA and proteins (Spitzer and Poolman, 2009), and also modifying the equilibrium constant of a reaction (Alberty, 2003), or even the transportation of a metabolite across a cellular membrane permeable to only one type of protonated species (Jol et al., 2010).

Furthermore, changes in temperature can modify the reaction rate and the equilibrium constant (Alberty, 2003; Leskovac, 2003).

The effects of macromolecular crowding are related to the steric repulsion of the molecules, which avoids two molecules occupying the same space at the same time. The excluded volume is the fraction of the total volume that the centre of mass of a molecule cannot reach at a time \( t \), while the volume that may be occupied by it is called available volume (Minton, 2001; Minton, 2006).

The available volume depends on the size and shape of the test and background molecules as well as the number of molecules present in the total volume. For example, when a spherical test molecule of radius \( r_T \) is added to a solution of “background” particles (i.e. those that do not participate in a particular reaction) of radii \( r_B \), if \( r_T \ll r_B \) the available volume (white regions in Figure 1.4A) will be greater than the case where \( r_T \sim r_B \) (white regions in Figure 1.4B).

The excluded volume (grey regions in Figure 1.4) can be easily identified by drawing a spherical shell with radius equal to the sum of \( r_T \) and \( r_B \) on each background particle, i.e. the minimum distance between the centres of the two molecules, then the available volume for the test species is that volume not occupied by these shells (Minton, 2001).
The molecular crowding may affect the rate and equilibrium of the reactions which involve changes in the available volume, i.e. the products are more compact than the reactants, for example reactions of self- or heteroassociation, adsorption, folding/unfolding protein, etc., including those reactions where the enzyme has a significant conformational change (Minton, 2001).

Crowding has complex and opposite effects on the rate of biochemical reactions:

- The thermodynamic activities of the reactants increase, enhancing in this way the rate of the reaction (which is proportional to the activity) and modifying the equilibrium of the reaction (Chebotareva et al., 2004; Minton, 2006).

- The diffusional mobility of the molecules decreases, so that the possibility of the meeting of the reactants and therefore the reaction rate is also reduced (Chebotareva et al., 2004). The molecular diffusion is an anomalous process in a heterogeneous medium such as the cellular environment (Verkman, 2002).

The predominant effect on the reaction rate depends on the volume occupied by the background molecules ($\phi$): at small values of $\phi$ the reaction rate is enhanced with the increase of $\phi$, and at high $\phi$ values the reaction rate decreases due to the increase in the viscosity of the medium and the decrease of the diffusion (Minton, 2001; Chebotareva et al., 2004).
The simulation of the intracellular environment can be complicated due to complex interactions among the cellular components, and also because the total number of individual molecules (enzymes, metabolites, molecules of water, etc.) can become quite large. Therefore, choosing the appropriated model (with the corresponding assumptions, simplifications and experimental validation), is essential for obtaining useful information that help to understand and manipulate the metabolism.

1.3. Mathematical representation of the metabolic systems and the importance of modelling

The cell is a complex structure delimited by a membrane, that contains a fluid phase (cytoplasm), macromolecules, enzymes, DNA, RNA, solutes, etc. The way those elements interact with each other (regulation, signalling, metabolic reactions, etc.) determines the behaviour of the cell.

The cellular metabolism consists of several, coupled enzymatic reactions (grouped in metabolic pathways) catalysing the conversion of a substrate $S_i$ into a product $P_n$, where the products of an enzyme are the substrates of the next ones (Mendes, 2004).

Mathematical models have been proposed to represent, in a simplified way, the interactions among the cellular components. This in order to understand and predict the behaviour of the cell under certain conditions, identifying the experimental strategies to manipulate the microorganism for the benefit of industrial process or for a new product development, thereby avoiding run experiments with undesirable results.

These models require experimental information as inputs, e.g. reaction constants, and have to be validated with experimental observations. However, the collection of information is not an easy task as some experiments are often done under unnatural conditions, so that they do not recreate the complex cellular environment. For example, enzymatic assays where either the presence of other enzymes are avoided to not affect the activity of the enzyme of interest, or other enzymes are added in quantities that may not exist in the cell but that optimise the assay (Mendes, 2004).
The modelling and simulation of metabolic pathways have many applications in the field of System Biology, Bioprocesses Engineering, and Metabolic Engineering, for example (Mendes, 2004; Saavedra et al., 2007; Llaneras, 2010):

- To study the effect of a perturbation in the enzymatic activity on the metabolic fluxes of a pathway.
- To study the regulation of metabolic pathways by comparing the simulation results with experimental data.
- To calculate the variation in the enzymatic concentrations required (and made by genetic engineering) to produce changes in a particular metabolite concentration.
- To identify steps or reactions with stronger negative effect in the metabolic network of a parasite, this information may be used for drug design targets.
- To identify steps or reactions with more influence (positive or negative) on the formation of a metabolite of interest and the by-products.
- To identify by simulation the environmental conditions that maximise the formation of a metabolite of interest without spending the material resources of the laboratory.

According to the phenomena we want to describe and the information available, one can choose the type of model and the assumptions required to better describe/solve the problem and compute useful results in a short computational time.

A model can be classified as (Blanch and Clark, 1996; Llaneras, 2010):

- **Unstructured** model, where the cell is considered a black-box catalyst that converts the substrates into products.
- **Structured** model, when the internal processes or reactions within the cell are taken into account.
- **Non-segregated** model, when the population is considered as homogeneous.
- **Segregated** model, distinguishes the characteristics among individuals of a population, e.g. mutations or aging of the cell.
- **Kinetic** model follows the dynamic or time evolution of intracellular processes; these models require reaction rates and other kinetic parameters (or at least their order of magnitude) obtained from experiments.
- **Stoichiometric** model assumes that the intracellular processes are at steady state, so there is no accumulation of metabolites with the cell.
\begin{itemize}
\item **Deterministic** model, when there is a relationship between the cause and effect, and the randomness is not involved in the prediction of the results or states, e.g. the models formulated as ordinary differential equations.
\item **Stochastic** model, when the result of an event depends not only on the initial condition but also on the random variables, e.g. Monte Carlo method.
\end{itemize}

Different methodologies are found in the literature to simulate the metabolic fluxes and the diffusion process within the cell. This thesis focuses on the improvement and implementation of two thermodynamically constrained stoichiometric models: One of them incorporates the crowding effect on the thermodynamic feasibility of the intracellular reactions. Furthermore, a kinetic model is developed based on the Lattice Boltzmann Method, which incorporates the effect of macromolecular crowding.

### 1.4. Thesis objective and research contribution

The aim of this thesis is to create computational tools for the simulation of the biochemical processes carried out within the cell. Even though the complete representation of the intracellular environment is not possible at the moment due to the complex interactions among the cellular components, this research focus on the estimation of thermodynamically feasible metabolic fluxes and the diffusion of molecules in a crowded media such as that present in the cell.

The previous works related to the modelling and simulation of intracellular processes are reviewed in Chapter 2.

The thesis is presented in “alternative format” as a series of papers submitted for publication in scientific journals. The research contributions can be summarised as follows:

1. The thermodynamically constrained stoichiometric models are useful tools for the determination of flux distributions in metabolic networks. These models are based on mass conservation and thermodynamic restrictions, assuming a well-mixed
system at steady state, and volumeless metabolites. The presence or absence of an enzyme determines if the catalysed reaction is blocked (where $\Delta G'$ can take any value) or at equilibrium ($\Delta G'=0$) if a zero flux is measured or estimated. Since there is not a direct relationship between flux and the $\Delta G'$ value of a reaction, and due to the uncertainties associated with thermodynamic information and the metabolites’ concentration, these models can estimate a zero flux through the mass balance only to “relax” the thermodynamic constraints and achieve a better maximum/minimum value, despite the presence of the enzyme. In this context, a methodology is proposed for the incorporation of the information related with the experimental detection of the enzymes catalysing the metabolic reactions. The central carbon metabolism of Actinobacillus succinogenes is used as case study to show the effect of considering a zero flux reaction as blocked or at equilibrium. Also the influence of the environmental conditions on the estimation of the range of fluxes is investigated. For this purpose, a methodology is presented to clarify and synthesise the steps for the estimation of the standard Gibbs free energy of formation of metabolite $i$, at different values of $I$, $T$, and $pH$. This research work is presented in Chapter 3.

2. The assumption that the cellular components are volumeless molecules is a simplification that could hide the effects of the crowding conditions prevailing in the cell. The reduction in the available space due to the presence of macromolecules and other solutes causes an increase in the activity of the metabolites. Strictly speaking the Gibbs free energy of a reaction is a function of the activity of the reactants, so that changes in the crowding conditions could alter the thermodynamic feasibility of a reaction. However this effect is not explicitly taken into account by the thermodynamically constrained stoichiometric models available in the literature. Based on the Scaled Particle Theory (SPT) (Reiss et al., 1959; Lebowitz et al., 1965), a methodology is proposed to incorporate the crowding conditions on a thermodynamically constrained stoichiometric model. The central carbon network of A. succinogenes for the production of succinic acid from glycerol and the lactic acid pathway are used as case studies. The research shows the importance of the crowded regions in the cell to convert a thermodynamically infeasible pathway into a feasible one, and also the crowding effect on the estimation of the flux distribution.
compared with the other environmental conditions. This study is presented in Chapter 4.

3. The crowding conditions not only affect the thermodynamic feasibility of the reactions, but also the reaction rate and the diffusion process. A reduction in the molecules’ diffusion has been experimentally observed in crowded media, this can be explained by the presence of background molecules that hinder the motion of a tracer. Monte Carlo (MC) algorithm has been widely used (Reese et al., 2001; Hari et al., 2009; Pastor et al., 2011; Vilaseca et al., 2011a,b; Fragkopoulus et al., 2012) for the simulation of the molecules’ diffusion under crowding conditions. MC is easy to implement but could be computational expensive for long time simulations and/or a large number of molecules. On the other hand the deterministic Lattice Boltzmann Method (LBM) tracks the motion of collections of point-like molecules allowing faster simulations, but neglecting the crowding effects. In this work, a methodology that couples the SPT and LBM is proposed for diffusion-simulations of different size molecules in 2D crowded media. Selected cases studies show the effectiveness of this crowding-adaptation of LBM to predict the same behaviour than kinetic Monte Carlo simulations (which is used for validation purposes) but with a significant reduction in the computation time. This work is presented in Chapter 5.

Chapters 3–5 contain a preamble and the corresponding paper submitted to scientific journals with the discussion of the corresponding work. Finally, the conclusions of these studies are summarised in Chapter 6.
Chapter 2

Literature Review

The modelling and simulation of metabolic systems help understanding the interactions among the cellular components, the processing of experimental information, and also the prediction of the cellular behaviour under certain environmental conditions.

However, due to the complexity of the cellular interactions (regulation, signalling, reactions, etc.), the knowledge gaps in the metabolism, and the lack of experimental data, the mathematical representation of the complete life processes is not possible at the moment. Therefore, the appropriate model has to be chosen according to the available information, the purpose of the study and the level of detail required in the results in order to accurately represent the intracellular process or phenomenon to be studied.

As mentioned in Chapter 1, the models can be classified as structured, unstructured, segregated, non-segregated, kinetics, stoichiometric, deterministic, or stochastic models depending on the information required and the assumptions made by the model. A non-exhaustive but illustrative review of the models and methodologies used to simulate metabolic systems is presented below.

2.1. Unstructured models

One of the most popular models applied in the industry to simulate the microbial metabolism is the unstructured non-segregate model. This type of model is the simplest one, it considers the cell as an entity (or black box) that converts the substrates into final products. In this model, the intermediary intracellular reactions (or metabolic reactions) are not taken into account, thus their reaction rates are not required. Besides,
it is assumed that there are no differences among the individual cells, e.g. those given for mutations.

A clear example of this type of models is the one given by the well-known Monod equation that describes the microbial growth, or the Luedeking-Piret model (Luedeking and Piret, 1959) for the product formation rate.

The kinetic simulation using this type of models often requires the solution of simultaneous Ordinary Differential Equations (ODEs). In these equations, some kinetic parameters (estimated from experimental data) are essential, e.g. the maximum specific growth rate and the substrate saturation constant (for the Monod expression), or the growth and non-growth associated coefficients (for the Luedeking-Piret model). Inhibition terms can also be considered in these equations.

Examples of unstructured non-segregated models for the batch fermentation of Actinobacillus succinogenes can be found in (Corona-Gonzalez et al., 2008; Lin et al., 2008; Vlysidis et al., 2011b).

The distinction among individual cells, e.g. the age and/or mass of the cells, can be also incorporated resulting in unstructured segregated models. The solution of these type of models require the knowledge of the distribution function of the mass (or age) cell population, for example, and the probability that once the cell division takes place, the cell’s daughter will have a mass between $m$ and $m+\Delta m$. However, these distributions functions are not always available and/or are difficult to determine experimentally (Blanch, 1981).

While unstructured models are useful for the design and optimisation of fermentation processes, only the structured models (that takes into account the intracellular reactions) could help to gain a deep understanding of the metabolism and the further identification of key reactions for the overproduction of a metabolite of interest.
2.2. Structured models

The cellular metabolism consists of a series of enzymatic reactions connected with one another by intracellular metabolites. The prediction of how the fluxes are distributed in the metabolic network is the first step for understanding the metabolism and planning the best strategy for its manipulation (e.g. environmental changes, genetic modification), which aims to improve the yield of the fermentation process or the development of new products.

The simulation of coupled enzymatic reactions, such as those present in the metabolic pathways, can be done using either stochastic models or deterministic models. Stochastic models are those where a previous event has no correlation with the subsequent one, so the next state of the system has to be determined probabilistically. On the other hand, deterministic models are those with a direct relationship between the cause and effect, therefore, in the phenomena simulation a particular set of initial conditions produce always the same results.

2.2.1. Deterministic models

Deterministic approaches are based on the law of mass action, where the reaction rate is proportional to the concentration of the reactants and products and the rate constant. In these cases, a system of differential equations is built to represent the dynamics of the metabolites involved in the metabolic pathways analysed (Berry, 2002).

This is the case of the Michaelis-Menten model (Eq. 2.1) that describes the rate of formation of the product $P$ of the enzymatic reaction $E + S \leftrightarrow ES \rightarrow E + P$. This equation is derived from a set of ODEs assuming the steady state of the enzyme-substrate complex $ES$ (i.e. it remains constant over time) (Leskovak, 2003).

$$
\nu = \frac{d[P]}{dt} = k_{cat} [E]_o \frac{[S]}{K_M + [S]} = V_{max} \frac{[S]}{K_M + [S]}
$$

Eq. 2.1
where $V_{max}$ is the maximum rate attainable, the brackets indicate the concentration of the species, and $K_M$ is the Michaelis-Menten constant (equivalent to the substrate concentration at which the reaction rate $\nu$ is $V_{max}/2$).

However, these deterministic approaches consider that the system is continuous, the reactants have small size, and that these reactants are diluted and well mixed. Thus, the considered medium is homogeneous (Aranda et al., 2006) with conditions that do not correspond to the intracellular environment (see section 1.2).

For the reasons mentioned above, some authors have modified the mass action law to incorporate the macromolecular crowding effects observed experimentally in environments such as the cytoplasm. For example, Laurent (1971) proposes to replace the Michaelis-Menten constant ($K_M$) by the apparent Michaelis-Menten constant ($K_M' = \Gamma_M K_M$) where the non-ideal factor ($\Gamma_M$) is the ratio between the available volume of the reactants and the products. Instead, Minton (1981) simulates the reaction kinetics correcting the rate constants ($k_i$) of the ODEs with non-ideal factors ($k_i' = \Gamma k_i$) (Schnell and Turner, 2004).

Despite these attempts to integrate the macromolecular crowding effects on the classical kinetic approach (mass action law), these models still assume a homogeneous medium. Under this scenario, the non-classical kinetic approaches (the fractal-like kinetics and the power-law approximation) have emerged to consider the heterogeneity of the medium.

From his experimental observations of exciton-fusion reaction in heterogeneous environments, Kopelman (1986, 1988) proposed fractal-like kinetics based on the mass action law, where the rate constants ($k_i$) change over time following the expression $k_i(t) = k_i^{0} t^{h}$. Here, $k_i^{0}$ is the rate constant at $t = 1$ and $h$ is the spectral dimension.

In the power-law approximation proposed by Savageau (1969), the reactant concentrations are raised to no-integer powers (i.e. kinetic order) in the ODEs. This model has only been validated using computer simulations for homodimeric reactions.
The disadvantage of the non-classical kinetic approaches is that they need either experimental data or results from computer simulations (using the Monte Carlo algorithm) to find the values of $k_i^0$, spectral dimension and kinetic order (Schnell and Maini, 2003; Schnell and Turner, 2004; Aranda et al., 2006).

Although it is possible to obtain experimental data of a single reaction to calculate the above parameters, the problem is more complicated for a multi-enzymatic system under conditions similar to the cell, leaving the simulation of the reaction dynamics with the Monte Carlo algorithm as the most viable alternative.

On the other hand, even though the values of the parameters were available, due to the deterministic approaches are based on the mass action law, it would be necessary to solve simultaneously a large number of ODEs, which represents a large computational cost.

There are softwares available to simulate metabolic pathways, for example GEPASI (http://www.gepasi.org/) that predicts the flux and metabolite concentrations with high degree of accuracy. For these predictions, the rate constants are required. However, these kinetic parameters are obtained from experiments with purified enzymes sometimes in unnatural media, so that inaccuracies in the results are expected. The reason for this inaccuracy is that the environment and the interactions with other intracellular components can modify the behaviour of the enzyme.

Another way to consider the heterogeneous nature of the intracellular medium is to formulate the model as a system of Partial Differential Equations (PDEs) (e.g. Slepchenko and Loew, 2010) where the diffusion term allows the simulation of the spatial variations of the species’ concentration and the effect of the reduced species’ diffusion due to the intracellular crowding conditions (that could imply the use of a spatially variable diffusion coefficient).

The methodologies based on the solution of ODEs or PDEs (if spatial variations are
considered) are considered to be at macroscopic level since they assume that the system is continuous and follows the dynamics of the metabolite’s concentration. Unlike macroscopic scale, the microscopic approaches follow the movement and interaction of each molecule or individual entity in the system, e.g. the Monte Carlo algorithm or the Cellular automata (CA). Instead, at the intermediate mesoscopic scale, a collection of individual entities are replaced by a distribution function expressing the average properties of the system, e.g. density, temperature, etc.

CA consists of a regular lattice where each site/cell has one of a finite number of states (e.g. on/off, bound/unbound). The system evolves according to some rules or mathematical functions that change the state of the each site for time \( t+\Delta t \) using the state’s information of the current sites and its neighbours at time \( t \). CA has been used in the simulation of diffusion and reaction processes at microscopic and mesoscopic scale, and it can be also formulated as a stochastic method (Weimar, 2002; Kier et al., 2005). However, this approach requires several simulation runs to estimate average values of the properties of the system (therefore the total computational time is increased), and only at the microscopic scale the excluded volume restrictions are explicitly considered since at most one molecule can occupy a site.

An alternative for the simulation of the diffusion and reaction processes at mesoscopic level is with Lattice Boltzmann Method (LBM), which is described below.

**Lattice Boltzmann Method**

LBM (McNamara and Zanetti, 1988) allows simulating the motion of collection of molecules on a regular lattice at every time step in a fast way. In this method, simple mathematical expressions that conserve mass and momentum are employed, which can reproduce the results of the macroscopic PDE for fluid flows or reaction-diffusion systems (Alemani, 2007). Due to its simplicity, this method has become popular for the dynamic simulation of fluids (Chen and Doolen, 1998; 44), but also has been applied to simulate reaction-diffusion systems (Dawson et al., 1993; Chen et al., 1995; Li et al., 2001), traffic flow models (Wagner et al., 1996), suspension of colloids (Ladd and
Verberg, 2001), flows in porous media (Dardis and McCloskey, 1998), multiphase and multicomponent systems (Shan and Chen, 1993).

The common ways to simulate the solute transport are 1) active solute component, where the solute is considered as another fluid so that a multicomponent system has to be solved (Shan and Chen, 1993), or 2) passive solute component, where the molecules (with no velocity) are carried by the fluid (Dawson et al. 1993; Alemani et al., 2012). However, both cases assume point-like molecules so there are no restrictions on the maximum number of solute molecules that can be present in the same fraction of space, and on the position exchange among two particles, neglecting in this way the effect of crowding conditions on diffusion-reaction process.

LBM has also been used to model the fluid motion in particle suspensions simulations where each molecule is represented by a hard sphere model, which could become computational expensive for the simulation of many molecules (Wang et al., 2010; Aidun and Clausen, 2010).

Stoichiometric models are used to estimate the fluxes/velocities of the reactions when the reaction rate constants are not available and the metabolic system can be considered at steady state and well-mixed. These models are described next.

**Stoichiometric models**

As mentioned before, the dynamics of the intracellular reactions can be represented by a system of ODEs

\[
\frac{dC}{dt} = S \cdot v
\]

Eq. 2.2

where the stoichiometric matrix \( S \) contains the information of the intracellular metabolites and reactions involved in the metabolic network, \( C \) is the vector of the intracellular metabolites concentration, and \( v \) is the flux vector.
It is assumed a balanced growth condition (Stephanopoulos et al., 1998; Provost and Bastin, 2004), which states that all cellular components grow at a constant rate. Thus, the intracellular metabolites concentration does not change in time (i.e. the system is at steady state) during the exponential growth or in a continuous culture. Eq. 2.2 reduces to the well-known mass balance constraint:

$$\mathbf{S} \cdot \mathbf{v} = 0$$  \hspace{1cm} \text{Eq. 2.3}

Eq. 2.3 is the base of the stoichiometric models that allow the estimation of unmeasured metabolic fluxes from the knowledge of stoichiometry of the system and some experimentally measured fluxes, such as the substrate uptake rate. Some examples of these type of models are Metabolic Flux Analysis, Flux Spectrum Approach, Flux Balance Analysis, Extreme Pathways, Elementary Flux Modes (Klamt and Stelling, 2003; Llaneras and Pico, 2008; Lee et al., 2011).

The pathways’ information, such as the metabolic reactions and the stoichiometric coefficients of the metabolites (or reactants), can be found in databases such as KEGG (http://www.genome.jp/kegg/) and BioCyc (http://www.biocyc.org). Other databases provide the information about the enzymes, e.g. BRENDA (http://www.brenda-enzymes.info), or genome annotation, e.g. PEDANT (http://pedant.gsf.de/) and CMR (http://cmr.jcvi.org).

The predictive stoichiometric models use optimisation techniques to analyse the flux distribution under certain conditions. For this, an objective function and the flux constraints such as the mass balance (Eq. 2.3) are required.

The objective function could be the maximisation of the biomass (Edwards et al., 2001; Deutscher et al., 2006), the minimisation of the ATP production for optimal energy efficiency (Ramakrishna et al., 2001; Vo et al., 2004), or the minimisation of the flux distribution changes given by a gene knockout, compared with that estimated for the wild type (Segre et al., 2002).

For example, the maximisation of the biomass (or the growth rate) allows determining
the robustness of the system, the essential reactions and their corresponding genes. This has a relevant application in the design of drugs for the identification of the target reactions that, once deleted or blocked, cause the prediction of zero biomass (Hu et al., 2007).

The mass balance constraint cannot capture all the complex interactions among the cellular components; therefore the prediction of a unique flux distribution, i.e. one single flux value per reaction, at certain environmental conditions is not possible, obtaining instead a range of fluxes.

In order to narrow the range of fluxes predicted, and therefore the space solution, other restrictions have been incorporated to the mass balance. For example, transcriptional regulation constraints (Covert et al., 2001; Covert and Palsson, 2002), crowding conditions (Beg et al., 2007; Vazquez et al., 2008), thermodynamics, gene expression, physiological flux limits for a particular microorganism or experimentally measured fluxes, gene knockout assumptions (Burgard et al., 2003).

In particular, the mass conservation and thermodynamic constraints are used in the proposed models presented in this thesis.

Thermodynamic constraints

The second law of thermodynamics states that a spontaneous process proceeds downhill, i.e. from a high energy state to a lower energy one. This loss of the available energy indicates the tendency of the system to equilibrium.

A system at equilibrium cannot perform any work, which in the case of a biological system represents the death of the organism that is unable to synthesise the essential compounds for cell maintenance. The energy required by the cell to maintain its nonequilibrium state (and therefore the driving forces of the process, e.g. the difference in the concentration) is taken from the surroundings in the form of substrates and/or sunlight.
In this respect, the thermodynamic constraints determine the feasibility and
directionality of the (spontaneous-) intracellular reactions according to their
 corresponding Gibbs free energy change ($\Delta_r G$). So that the direction of a reaction with
negative $\Delta_r G$ is in the formation of products (forward direction), a positive $\Delta_r G$
indicates backward direction, and a $\Delta_r G = 0$ means that the reaction is in equilibrium
therefore the net flux ($\nu$) through it is zero too.

The Gibbs free energy change of a reaction $\Delta_r G$ is given by differences between the
Gibbs free energy of formation of the products and reactants, and depends on the
intracellular conditions like pH, ionic strength ($I$), temperature, crowding conditions,
and the concentrations of the metabolites.

$$
\Delta_r G = \sum_{i=1}^{\text{Products}} \eta_i \left[ \Delta_f G_i^0 + RT \ln(a_i) \right] - \sum_{i=1}^{\text{Reactants}} \eta_i \left[ \Delta_f G_i^0 + RT \ln(a_i) \right]
$$

Eq. 2.4

where $\eta_i$ is the stoichiometric coefficient, and $\Delta_f G_i^0$ is the standard Gibbs free energy
of formation of a metabolite $i$. The activity of a metabolite is defined as $a_i = \gamma_i C_i / C_i^\alpha$,
where $C_i^\alpha$ is the standard concentration of the metabolite $i$. The activity coefficient $\gamma_i$
takes into account the interactions between the molecules, e.g. electrostatic and/or
steric, that causes deviations from the ideal behaviour of solutions. If the $\gamma_i$ value is
constant under the environmental conditions, then this term can be absorbed by the
$\Delta_f G_i^0$ value, so that the Eq. 2.4 become a function of the metabolites concentration $C_i$
instead of $a_i$.

The thermodynamic analysis of metabolic system has been used to determine the
feasibility, directionality and reversibility of the reactions (Yang et al., 2005; Henry et
al., 2006; Henry et al., 2007; Hoppe et al., 2007), eliminating infeasible loops in the
network (Beard et al., 2002; Beard et al., 2005), which reduces the solution space, and
also detecting putative regulatory sites (Kummel et al., 2006; Henry et al., 2007).

The regulatory sites in a metabolic pathway are identified as being those reactions with
a large $\Delta_r G$ (positive or negative), so they are relatively insensitive to changes in the metabolites concentrations since they cannot reach the equilibrium under the intracellular environmental conditions, therefore they could be regulated at transcriptional level.

Attempts to integrate the thermodynamics into stoichiometric models range from (1) the allocation of the direction and reversibility/irreversibility of a reaction based on in vitro experimental results (Fleming et al., 2009), (2) assuming irreversibility when molecules of ATP are involved in the reaction (Fleming et al., 2009), (3) first estimate the range of fluxes using only the mass balance constraint, then with these flux ranges identify the direction of essential reactions for the biomass growth, and finally determine through the independent calculation of $\Delta_r G$ whether the directions proposed were thermodynamically feasible (Henry et al., 2006), (4) assuming a reaction as irreversible when its standard Gibbs free energy of a reaction, i.e. $\Delta_r G^0 = \sum_{j=1}^{\text{Products}} \eta_i \Delta_r G_t^0 - \sum_{i=1}^{\text{Reactants}} \eta_i \Delta_r G_t^0$, is higher than a particular value (Binns et al., 2011), (5) calculating the maximum and minimum $\Delta_r G$ based on the range of intracellular concentration of the reactants and products to determine the reversibility of a reaction (Mavrovouniotis, 1993).

The drawback of these approaches is that they do not solve the mass balance and thermodynamic constraints simultaneously, and even if the direction of some reactions could be set based on experimental observations or the $\Delta_r G$ value, there are uncertainties in the feasible direction of many other reactions.

For example, if reaction 1 requires a specific concentration ratio between reactants and products in order to be feasible in forward direction, may be these metabolites concentration values do not favour the feasibility of reaction 2 in forward direction. Therefore, the flux range of reaction 2 must only include the values on the backward direction. This could change the flux distribution of other reactions in a non-obvious way, especially in large-scale metabolic networks.

Among the methodologies that solve simultaneously both mass and thermodynamic
constraints are: Energy Balance Analysis (EBA) (Beard et al., 2002), and Thermodynamics-based metabolic flux analysis (TMFA) (Henry et al., 2007). The latter is very similar to the model proposed by Hoppe et al. (2007).

EBA couples the mass and thermodynamic constraints using a nonlinear optimisation problem that, based on the null space of the stoichiometric matrix, allows the prediction of thermodynamically feasible flux distributions without previous knowledge of Gibbs free energy of formation \( \Delta f G_i \) of the metabolites involved in the biochemical reactions.

This feature becomes very important when no reliable thermodynamic data are available. However, since the environmental conditions directly affect the \( \Delta f G_i \) value, this information must be provided if the impact of ionic strength, \( pH \), temperature, and/or crowding conditions on the range of fluxes predicted needs to be also evaluated.

On the other hand, TMFA makes possible to estimate the thermodynamically feasible flux distribution and the intracellular metabolite concentrations by solving a Mixed Integer Linear Problem (MILP). In this methodology, the \( \Delta f G_i^0 \) values are essential for the thermodynamic analysis, so that the effect of environmental conditions, ionic strength and \( pH \) is easily taken into account.

The TMFA’s formulation of the optimisation problem allows the prediction of global optima. In this methodology, every flux reaction is split into its forward and backward components, i.e. two variables represent the net flux through that reaction, where at most one of the two components can have a nonzero value at the same time. This makes the stoichiometric matrix larger, and since there are more variables, more optimisations have to be done. In order to obtain the flux range (the minimum and maximum value) of a reaction, a total of 4 optimisations (2 for each variable or flux component) have to be carried out.

In these type of methodologies, if during the optimisation the \( \Delta f G_i \) of a reaction \( i \) is predicted zero, then the flux \( v_i \) is constrained to zero. However, it is assumed that a
zero flux indicates that the reaction is blocked, i.e. the enzyme that catalyses it is absent or inactive, therefore $\Delta_i G_i$ can take any value.

The estimation of an equilibrium reaction ($\Delta_i G_i = 0$) requires a specific ratio between the reactants and products concentration. However, due to the uncertainties associated with the measure or estimation of metabolites concentrations and $\Delta_i G_i^0$, it is possible that a $v_i = 0$ can be reached first during optimisation by the mass balance than a $\Delta_i G_i = 0$ for the thermodynamic constraint, even if the presence of an enzyme $i$ has been experimentally detected.

On the other hand, the crowding conditions on the stoichiometric models has been incorporated as a volume restriction of the maximum number of enzymes that can fit inside the cell (Beg et al., 2007; Vazquez et al., 2008), which in turn constrain the maximum metabolite formation rate (i.e. flux) attainable for reactions in the network. However, the crowding effect on the thermodynamic analysis of the metabolic pathways has not been reported.

Regarding the collection of the thermodynamic data of the metabolites, there are several difficulties, for example:

- The thermodynamic information can be found in databases such as Alberty (2003), or Li et al. (2011). However, the experimental standard Gibbs free energy of formation ($\Delta_i G_i^0$) at physiological conditions ($pH$ 7, 298 K, $I = 0$) is not always available for all the metabolites. In these cases, and if the molecular structure is known, the Group Contribution Methods (GCM) can be used as an alternative, e.g. those proposed by Mavrovouniotis (1990), Jankowski et al. (2008), or Noor et al. (2013). In GCM, the molecule is decomposed into smaller groups or substructures, where each group is associated with a Gibbs free energy, and then the sum of the energy of all the groups gives the standard Gibbs free energy of formation of the molecule. Due to the uncertainty associated with the estimation of $\Delta_i G_i^0$, it is recommended to use only one GCM (or database) for all the molecules involved in a reaction. In this way, the inaccuracies in the formation/destruction of a group can
be cancel out.

- Some metabolites dissociate in water, forming one or more protonated species whose proportion in the solution depends on the pH, where each species has its own $\Delta_f G^0$. The set of protonated species of a metabolite is called pseudoisomer group. In order to simplify the calculation, Henry et al. (2007) proposed the use of the $\Delta_f G^0$ of predominant protonated (ionic) species. However, this approximation could be questionable when the logarithm of the acid dissociation constant $pK_a$ of the metabolite is close to the medium pH, so that the concentration of two protonated species is similar, e.g. $HA \xrightarrow{pK_a} A^-$. In this context, Alberty (2003) proposed the estimation and use of the $\Delta_f G^0$ of the pseudoisomer group for those reactants. In this thesis (Chapter 3 and 4) each metabolite is represented by its pseudoisomer group.

- The $pK_a$'s for all the metabolites are not available, but they can be estimated from their chemical structure with the software MarvinSketch ver. 5.5.1.0, 2011, ChemAxon (http://www.chemaxon.com).

- The effect of the temperature in the metabolites’ $\Delta_f G$ may be taken into account using the standard enthalpy of formation of the species. Unfortunately, the enthalpy value is not available for all the metabolites, and in some cases they cannot be estimated by the GCM (Domalski and Hearing, 1993; Cohen and Benson, 1993; Constantinou and Gani, 1994; Domalski, 1996; Politzer et al., 2005) because they do not consider groups with phosphorous, which is a common chemical element in biological compounds.

**Approximations for the estimation of the activity coefficient under crowding conditions**

As mentioned in previous section, the activity coefficient $\gamma$ determines the non-ideality of the solutions including those given by the crowding conditions on the activity of the molecules (which affects the thermodynamic equilibrium and rate of the reactions, as well as the diffusion process).
Ogston (1958) proposed a way to approximate $\gamma$ of hard-sphere particles in a medium containing straight fibres of negligible thickness randomly distributed, where the number of fibres in a given space is determined by a Poisson distribution. In this case, the natural logarithm of $\gamma$ is a linear function of the concentration of all solute species present in the analysed system (the solvent is considered as a continuum).

An alternative for the estimation of $\gamma$ is the use of the Scaled Particle Theory (SPT).

**Scaled Particle Theory**

SPT (Reiss et al., 1959; Lebowitz et al., 1965) describes the interactions among hard-spheres molecules that avoids the overlapping of two spheres in the same space, neglecting the interactions given by dipole, electrostatic, hydrogen-bonding, hydrodynamic, etc. However, it is assumed that dipole, electrostatic, hydrogen-bonding interactions determine the volume of the system or fluid, i.e. how close the molecules are in the space due to the repulsion/attraction forces.

The mathematical formulation of SPT is based on the estimation of the work required to insert a test sphere-molecule $i$ in a fluid (or mixture) of hard-spheres molecules, allowing the calculation of the activity coefficient of the test molecule $\gamma_i$ (which is equivalent to the inverse of the probability to find enough available space in the fluid to fit molecule $i$).

The logarithm of $\gamma_i$ is expressed as a series of powers of the radius of molecule $i$, and is a function of the specific volume, number and radii of the background molecules of mixture. In fact the SPT predictions are very sensitive to these parameters (Zimmerman and Trach, 1991; Tang and Boomfield, 2000). In this regard Berg (1990) suggests the explicit incorporation of the fluid (or solvent) molecules in the analysis, e.g. water molecules for aqueous solutions, instead of considering as a continuum.

SPT has been applied for the estimation of thermodynamic activity of globular proteins.
(Minton, 1981), as well as the crowding effect on their stability (Minton, 2000), studies of solvation (Tang and Bloomfield, 2000), to simulate the kinetics of adsorption (Minton, 2001b) and enzymatic reactions (Smolen and Keizer, 1990), diffusion of proteins in membranes (Minton, 1989), the estimation of the osmotic pressure (Minton, 1995).

2.2.2. Stochastic models

Since the reaction rate depends on the probability that an enzyme and its substrates collide, it seems that a stochastic approach is an accurate alternative to model the dynamics of intracellular reactions (Aranda et al., 2006). In these reactions, the diffusion of the molecules, which determines the collision frequency, is an anomalous process in heterogeneous media such as the cytoplasm (Verkman, 2002).

Although the master equation allows determining the stochastic time evolution of a chemical reaction system, its analytical solution is limited to a few problems, and the numerical solution is not easy to compute (Gillespie, 1977).

Under this scenario, Gillespie (1977) proposes the Stochastic Simulation Algorithm (SSA) to simulate the dynamics of coupled reactions in a homogeneous medium using two random numbers to choose which reaction occurs and the waiting time between reactions. Although some modifications have been suggested to improve its efficiency (Gillespie, 2001; Tian and Burrage, 2004), it still remains the assumption of perfect mixing and homogeneity.

Latter, Nicolau and Burrage (2008) set the called Anomalous SSA (ASSA) to integrate the effects of heterogeneous and crowded media. Here, the Monte Carlo algorithm is used to calculate the parameter $h$ of the time dependent rate constant $k_i(t) = k_i^0 t^{-a}$, which replaces the rate constant in the SSA methodology.

The Monte Carlo (MC) method is a powerful technique widely used to estimate the reaction kinetics (Ander et al., 2004; Boulianne et al., 2008; Ridgway et al., 2008;
Slepchenko and Loew, 2010; Pastor et al., 2011), the effects of the formation of enzymatic complexes on the reaction rate (Amar et al., 2008), the time dependent rate constant (Berry, 2002), the anomalous diffusion (Vilaseca et al., 2011a; Vilaseca et al., 2011b), and the multifractality of the Michaelis-Menten constant ($K_M$) (Aranda et al., 2006), in homogeneous and heterogeneous media.

MC can be implemented on two or three dimensions of a lattice or lattice-off (continuous) space (potentially containing immobile obstacles), in which the diffusion of the particles is modelled by the random independent walks of each molecule in a discrete time step, if two molecules collide they may react with a certain probability.

In the case of (lattice-) kinetic Monte Carlo (kMC) (Reese et al., 2001, Hari et al., 2009; Fragkopoulous et al., 2012), a MC method, the probability that an event occurring, e.g. the diffusion of a molecule or one type of reaction, and the time (here the time step is variable) in which it happens depend on the rate constants of the processes (parameters that are required for the simulation).

The advantage of MC approaches is that they allow simulating complex systems. However, the high computational cost prevents obtaining kinetics for long time. On the other hand, if the space is represented by a grid, there are computational restrictions on the size of the domain that can be simulated (Pastor et al., 2011), and if it is a lattice-off space then the number of molecules simulated becomes a limiting factor in the efficiency of the search algorithm to find a nearby molecule (Boulianne et al., 2008).

In two-dimensional lattice approaches, the movement of the molecule is restricted to four possible directions. In reality, the direction may be chosen from a range of angles between 0 and 360 degrees. Moreover, lattice approaches ignore the real size of the molecules (considering that all the species has the same radius, or deforming the spherical shape of the molecules) and their exclusion volume, which alters the diffusion process (Vilaseca et al., 2011a) and therefore the reaction kinetics.

Several simulators are available to estimate the reaction dynamics (Table 2.1). These simulators are based on MC algorithm, ODE, PDE, SSA, some of them allow to
simulate the macromolecular crowding (i.e. when molecule’s density affect the
diffusion and reactivity), and others consider a well mixing solution. The limitations of
each one depend on the method used, as discussed above. However, none of them take
in account the difference between the size of the molecules (treating the molecules as
points or assuming the same size for them) and therefore its possible excluded volume’s
effect.

Table 2.1. Some simulators for the reaction dynamics

<table>
<thead>
<tr>
<th>Simulator</th>
<th>Method used</th>
<th>Macromolecular crowding</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>GridCell</td>
<td>MC</td>
<td>Yes</td>
<td>Boulianne et al., 2008</td>
</tr>
<tr>
<td>HSIM</td>
<td>MC</td>
<td>Yes</td>
<td>Amar et al., 2008</td>
</tr>
<tr>
<td>SmartCell</td>
<td>SSA(^a)</td>
<td>No</td>
<td>Ander et al., 2004</td>
</tr>
<tr>
<td>MesoRD</td>
<td>SSA(^a)</td>
<td>No</td>
<td>Hattne et al., 2005</td>
</tr>
<tr>
<td>MCell</td>
<td>MC</td>
<td>No</td>
<td>Boulianne et al., 2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ridgway et al., 2008</td>
</tr>
<tr>
<td>Cell++</td>
<td>MC</td>
<td>No</td>
<td>Boulianne et al., 2008</td>
</tr>
<tr>
<td>ChemCell</td>
<td>MC</td>
<td>No</td>
<td>Boulianne et al., 2008</td>
</tr>
<tr>
<td>Smoldyn</td>
<td>MC</td>
<td>No</td>
<td>Ander et al., 2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ridgway et al., 2008</td>
</tr>
<tr>
<td>VCell</td>
<td>PDE</td>
<td>No</td>
<td>Slepchenko and Loew, 2010</td>
</tr>
<tr>
<td></td>
<td>ODE</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SSA</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)The simulation space is divided in subvolumes in order to consider heterogeneities.

The HSIM proposed by Amar et al. (2008) is one of the most complete simulators; it is
based on MC algorithm on a lattice-off space. It tracks the position of each molecule,
and models the formation of temporary enzymatic complexes by changing the affinity
parameters among enzymes. The diffusion of a molecule is a function of the probability
to move a distance of 10nm, and although there are differences between the diffusion
coefficients of the species, it considers that all the molecules have the same size.
2.3. Concluding remarks

This Chapter has reviewed various models used for the simulation of metabolic processes. The selection of one method to represent the metabolism depends mainly on the purpose of the model, the available information of and the type of assumptions one can make about the analysed system.

The unstructured models are the simplest ones, they allow the estimation of the microbial growth and the production of metabolites of interest under certain environmental conditions, e.g. the substrate concentration. The formulation of these models requires the parameters estimation using data from several experiments testing different conditions, and do not provide information about the intracellular processes.

On the other hand, structured models give a deeper understating of the cellular metabolism, how the environmental/genetic perturbations could affect it, and the strategies for its effective manipulation.

The crowding conditions and the heterogeneity of the medium are among the characteristics of the intracellular medium that affect directly the activity of the metabolites, modifying the thermodynamic equilibrium, the rate of the reactions and the diffusion process. In an attempt to incorporate the effect of these characteristics on the simulations, several authors have proposed the use of time-dependent kinetic parameters in the models (whether they are stochastic or deterministic). However, the estimation of these parameters requires either experimental information (often not available) or the simulation results from other models such as the MC methods.

Even though the microscopic MC methods give an easy way to simulate the crowding effect, they have limitations due to the computational cost to represent large amount of molecules, lattice size, and/or long time simulations. Nevertheless, other more efficient methods to overcome these drawbacks such as LBM neglect the crowding effects assuming point-like molecules.

A methodology based in LBM and SPT for 2D crowded systems is presented in the
Chapter 5 of this thesis to reduce the computational time required for the simulation of the diffusion process in environments such as the cytoplasm.

Sometimes the kinetic parameters (reaction rate and/or the diffusion coefficient) of the analysed system are not available, but the fermentation conditions allow assuming that the cellular metabolism is at steady state. Under these circumstances, the stoichiometric models are a good alternative for the estimation of the intracellular fluxes. Several constraints can be used in addition to the mass conservation to restrict the optimisation problem given in the stoichiometric models, narrowing in this way the range of fluxes predicted by these models.

In the case of thermodynamically constrained stoichiometric models, a common assumption made in these methodologies is that the prediction of \( \nu_i = 0 \) indicates that the reaction \( i \) is blocked, therefore \( \Delta_r G_i \) can take any value. However, only the presence or absence of the enzyme can determine if the reaction is blocked or at equilibrium.

In this thesis, the effect of considering a reaction \( i \) blocked \((\Delta_r G_i \neq 0)\) or at equilibrium \((\Delta_r G_i = 0)\) on the range of fluxes of the metabolic network under different environmental conditions (ionic strength, temperature, and \( pH \)) is investigated and presented in Chapter 3.

Moreover, the effect of the crowding conditions on the prediction of the thermodynamic feasibility of the metabolic networks is investigated. A thermodynamically constrained stoichiometric methodology based on SPT is presented and discussed in Chapter 4.
Chapter 3

Steady State Metabolic Network Analysis: a Thermodynamically Constrained Stoichiometric Model

3.1. Preamble

Many industries employ biological agents in their processes because they have high reaction specificity, do not require extreme temperature and pressure conditions compared to the chemical processes, among others.

The objective of the metabolic models is to accurately represent the metabolism and the interactions among the cellular components. These models allow a deep understanding of the cell behaviour, and provide useful information for the design of strategies for the organism’s manipulation, which can potentially improve the profitability of the process, reduce the recovery cost of the product, help in the new product development, etc. A review of the models for metabolic systems is presented in Chapter 2.

In this sense, the stoichiometric models are valuable tools for the estimation of the current flux distribution or predicting the flux distribution under environmental or genetic perturbations. For this, they use optimisation techniques seeking maximise or minimise an objective function, e.g. the biomass production or the formation of a metabolite of interest, subject to a mass conservation constraint, but other restrictions can also be added. Particularly, the work presented in this chapter is based on the mass balance and thermodynamic constraints.

Previous methodologies have incorporated the thermodynamic analysis to the stoichiometric models (see Chapter 2), allowing the prediction of net fluxes in the
direction of the Gibbs free energy ($\Delta_G$) drop, where a $\Delta_G = 0$ means that the reaction is in equilibrium and therefore the net flux ($v$) through it is also zero.

A common assumption made in these methodologies is that the prediction of $v_i = 0$ indicates that the reaction $i$ is blocked, i.e. the enzyme that catalyses it is absent or inactive, therefore $\Delta_G_i$ can take any value. However, only the presence or absence of the enzyme can determine if the reaction is blocked or at equilibrium.

The aim of this chapter is to determine the effect of considering a reaction $i$ (with $v_i = 0$) blocked ($\Delta_G_i \neq 0$) or at equilibrium ($\Delta_G_i = 0$) on the range of fluxes attainable by the other metabolic reaction of system.

In order to do this, a methodology based on the TMFA idea (Henry et al., 2007) is proposed in the Publication 1 to compute thermodynamically feasible flux distribution, the $\Delta_G$ and intracellular metabolite concentration ranges at different environmental conditions. This methodology incorporates the information about the presence of the enzymes.

The case study used in this paper is the metabolic pathways of *Actinobacillus succinogenes* involved in the production of succinate from glycerol. *A. succinogenes* has drawn the attention of the researchers for its ability to synthesise large amounts of succinate and tolerate high acid products concentrations, what makes this microorganism interesting for industrial fermentation (see Chapter 1). This represents a potential alternative for the disposal and re-valuation of the by-product glycerol in the biodiesel manufacture.

The metabolic pathways (and reactions stoichiometry) involved in the synthesis of succinic acid from glycerol in the *A. succinogenes*’ metabolism were obtained from KEGG database. The metabolic network includes the reactions for formation of the typical by-products acetate, formate, ethanol, and lactate.

Although ethanol and lactate were not experimentally detected in the fermentation
medium for the growth on glycerol (Vlysidis et al., 2011a), the intracellular concentration of these metabolites, their intermediates, and the enzymatic activity were not measured and no information is available. However, since the presence of the enzymes related with ethanol and lactate has been detected in the *A. succinogenes*’ growth on substrates such as glucose and sorbitol, they were also considered in this study. Besides, since no experimentally data are available for the intracellular concentration of the metabolites a physiological range is used instead.

Moreover, the standard Gibbs free energy of formation of the metabolites and also the uncertainty $SE$ associated to the standard Gibbs free energy of the reactions are obtained from the group contribution method (GCM) estimations reported by Jankowski et al. (2008). While the standard enthalpy of formation of each metabolite is taken from database given by Alberty (2003), and if that value is not available, then a GCM (e.g. Gani method) is chosen for its estimation using Aspen Plus®. The thermodynamic data used here are presented in the Supplementary Material of Publication 1.

The results and discussion of this work are presented in following paper, where the influence of the environmental conditions: ionic strength, temperature and $pH$ on the estimation of the range of fluxes was also evaluated (the influence of crowding conditions are presented in Chapter 4). Here, it is assumed that the medium is homogeneous and the effects due to macromolecular crowding are negligible, which is equivalent to consider all the cellular components as volumeless molecules.

In this work, each metabolite is represented by its pseudoisomer group. In order to clarify the estimation procedure of the standard Gibbs free energy of formation of a pseudoisomer group at different ionic strength temperature and $pH$, a strategy is also suggested.
3.2. Publication 1

Estimation of flux distribution in metabolic networks accounting for thermodynamic constraints: The effect of equilibrium vs. blocked reactions

L. Angeles-Martinez\textsuperscript{a} and C. Theodoropoulos\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a}School of Chemical Engineering and Analytical Science, University of Manchester, Manchester, M13 9PL, UK

*Corresponding author at: School of Chemical Engineering and Analytical Science, University of Manchester, Manchester, M13 9PL, UK
Tel. +441613064386; Fax +44 1612367439
k.theodoropoulos@manchester.ac.uk
Abstract

Thermodynamically constrained stoichiometric-based models have been widely used for the estimation of feasible flux distributions of a metabolic network. The prediction of a zero net flux through a reaction indicates that this reaction is either blocked (the enzyme is absent) or at equilibrium (the enzyme is present but the Gibbs free energy change of the reaction is zero). The estimation of the thermodynamic equilibrium of a reaction requires the exact knowledge of the environmental conditions and the metabolites’ concentrations. This information, however, is not always available. Here, the potential effect of considering a reaction blocked instead of at equilibrium on the estimation of the fluxes of other reactions is analysed. The central carbon metabolism of Actinobacillus succinogenes for the production of succinic acid from glycerol has been used as case study, for the prediction of flux distributions based on results from experiments in 1.8 L batch bioreactors. The impact of changes in ionic strength (I), temperature (T) and intracellular pH and medium pH was also investigated, revealing that only I and T affect the prediction of the flux distributions compared with those obtained at standard biological conditions when zero flux reactions are considered either as blocked or at equilibrium. In general, the range of fluxes estimated for the equilibrium case is narrower than that for the block case.

Abbreviations: 13dpg, 3-phospho-D-glyceroyl phosphate; 2pg, glycerate 2-phosphate; 3pg, 3-phospho-D-glycerate; 6pgc, 6-phospho-D-gluconate; 6pgl, 6-phospho-D-glucono-1,5-lactone; ac, acetate; acald, acetaldehyde; acoa, acetyl-CoA; actp, acetyl phosphate; coa, coenzyme A; dhap, dihydroxyacetone phosphate; e4p, erythrose 4-phosphate; etoh, ethanol; f6p, fructose 6-phosphate; fdp, fructose 1,6-bisphosphate; for, formate; fum, fumarate; g3p, glyceraldehyde 3-phosphate; g6p, glucose 6-phosphate; glyc, glycerol; glyc3p, glycerol 3-phosphate; lac, lactate; mal, malate; nad, nicotinamide adenine dinucleotide; nadh, nicotinamide adenine dinucleotide — reduced; nadp, nicotinamide adenine dinucleotide phosphate — reduced; oaa, oxaloacetate; pep, phosphoenolpyruvate; pi, phosphate; pyr, pyruvate; r5p, ribose 5-phosphate; ru5p, ribulose 5-phosphate; s7p, sedoheptulose 7-phosphate; succ, succinate; xu5p, xylulose 5-phosphate; PGI, glucose-6-phosphate isomerase; PFK, 6-phosphofructokinase; FBP, fructose-1,6-bisphosphatase; FBAr, fructose-bisphosphate aldolase; TPI, triose-phosphate isomerase; GAPD, 3-phosphate glyceraldehyde dehydrogenase; PGK, 3-phosphoglycerate kinase; PGM, phosphoglycerate mutase; ENO, enolase; PYK, pyruvate kinase; LDH, lactate dehydrogenase; PFL, pyruvate formate lyase; PDH, pyruvate dehydrogenase complex; ME2, malic enzyme; OAD, Oxaloacetate decarboxylase; ALCD2x, alcohol dehydrogenase; PPCK, phosphoenolpyruvate carboxykinase; MDH, malate dehydrogenase; FUM, fumarase; FRD, fumarate reductase; G6PDH2r, glucose 6-phosphate dehydrogenase; PGL, 6-phosphogluconolactonase; GND, phosphogluconate dehydrogenase; RPI, ribose-5-phosphate isomerase; RPE, ribulose 5-phosphate 3-epimerase; TKT1, transketolase; TALA, transaldolase; TKT2, transketolase; GLYK, glycerol kinase; G3PD2, glycerol 3-phosphate dehydrogenase; THD2, nad(p) transhydrogenase; EX_pyr(e), pyruvate exchange; EX_coa(e), coenzyme A exchange; EX_co2(e), co2 exchange; EX_pyr(e), phosphate exchange; EX_lac(e), Lactate exchange; EX_for(e), formate exchange; EX_ac(e), acetate exchange; EX_etoh(e),...
Keywords Glycerol; Modelling; Bioconversion; Bioreactions; Thermodynamic flux analysis; Actinobacillus succinogenes.

1. Introduction

*Actinobacillus succinogenes* is a mesophilic, Gram negative, facultative anaerobic bacterium capable of growing in a broad range of substrates [1,2] among them glycerol. It can synthesise succinate at a high concentration as the main product, which makes this microorganism an interesting biocatalyst for the industrial production of succinate from biorefinery glycerol, the main by-product in the biodiesel manufacture [3,4].

The understanding of how the fluxes are distributed in the metabolic pathway of *A. succinogenes* or any other microorganism can help to determine the optimal environmental conditions and/or the genetic strategies for engineering the metabolism in order to maximise the production of a target metabolite and to reduce the formation of by-products.

Different methodologies such as Metabolic Flux Analysis, Flux Spectrum Approach, Flux Balance Analysis, among others, have been applied to model the metabolic network of a microorganism under certain environmental conditions. However, due to the lack of complete information on the system (metabolic network, thermodynamics of the reactions, regulation, etc.) the prediction of a unique solution for the flux distribution is not possible at present. See [5-8] for a review of the available stoichiometric-based models.

In this respect, the addition of thermodynamic constraints narrows the solution space because they only allow the estimation of net fluxes in the direction of the Gibbs free
energy change. The value of the change in Gibbs free energy \( \Delta_G' \) depends on intracellular conditions such as \( pH_c \), ionic strength \( I \), temperature \( T \), and concentration of the metabolites. Thermodynamic applications in biological systems are reviewed by Soh and Hatzimanikatis [9]. Among the methodologies proposed in the past for the integration the thermodynamic data to stoichiometric methods [10-17] are the Energy Balance Analysis (EBA) [18] and Thermodynamics-based Metabolic Flux Analysis (TMFA) [19]. EBA allows the prediction of feasible fluxes without previous knowledge of thermodynamic information, through the solution of a nonlinear optimisation problem. However, thermodynamic data are necessary if the calculation of the metabolite concentrations are required [20]. Without the use of thermodynamic properties, the impact of changes of \( pH_c \), \( I \), and \( T \) on the feasibility and directionality of reactions is not obvious. Conversely, TMFA allows the direct evaluation of the impact of these changes by solving a Mixed Integer Linear Problem calculating simultaneously the concentration of metabolites and the flux distribution.

Despite the fact that the cell is an open system and therefore it can only reach a nonequilibrium steady state [21], there is still the possibility to find some reactions at equilibrium within it. In most methodologies, including EBA and TFMA, a value of \( \Delta_G'=0 \) causes the prediction of a zero net flux \( (v) \), but a zero net flux does not necessarily result in \( \Delta_G'=0 \). Hence, within these methods a reaction is considered blocked when \( v=0 \) so \( \Delta_G' \) is practically allowed to take any value. Nevertheless, physically only the absence or presence of the enzyme that catalyses the reaction in question determines if this is blocked \( (v=0 \text{ and } \Delta_G'\neq0) \) or at equilibrium \( (v=0 \text{ and } \Delta_G'=0) \).

The prediction of \( \Delta_G'=0 \) requires a specific concentration ratio between products and reactants. When the concentration of a metabolite is unknown a range of physiological concentrations can be used. However, since there is not a direct relationship between the flux and the value of \( \Delta_G' \), when the concentration range used is too wide, a zero flux reaction can be computed (satisfying mass conservation) before thermodynamic constraints reach the corresponding \( \Delta_G'=0 \), thereby predicting that reaction as blocked instead of at equilibrium.
In other words, it is possible that during the optimisation of a variable $x$, another flux variable $v_k$ is intentionally assigned a zero value (provided that it fulfils the mass conservation constraints) only to “relax” the thermodynamic constraints (inasmuch as the reaction $k$ is then considered blocked and $\Delta_G$ can take any value), and thus achieve the maximum/minimum value of $x$.

Since a metabolite can be involved in several reactions, one of them being at the equilibrium could affect the range of flux of other metabolic reactions. The aim of this paper is to compare the effect of considering a reaction being at equilibrium or being blocked when its net flux is equal to zero, combined with the influence of the environmental conditions $I, T$ and $pH$ on the estimation of the corresponding flux distributions. Our case study will be the central metabolism of A. succinogenes for the production of succinic acid from glycerol.

For this, and based on the work of Henry et al. [19], we formulate the coupling of the mass balance and thermodynamic constraints as a nonlinear optimisation problem which explicitly evaluates the impact of $I, T$ and $pH$ on the estimation of thermodynamically feasible flux range as well as of intracellular metabolite concentrations.

2. Materials and methods

In the next section we summarise the methodology used for the calculation of the standard Gibbs free energy of formation, $\Delta_f G^0$, of each metabolite ahead of the description of the equations corresponding to the mass balance and thermodynamic constraints.

2.1. Calculation of $\Delta_f G^0$

When a metabolite, $i$, is in aqueous solution, it dissociates in different protonated species $j$ with either zero, positive or negative charge, the so called pseudoisomer group
The concentration of each species $j$ depends on the $pK_a$ (the logarithm of the acid dissociation constant) values and the $pH$ of the medium.

Here, we represent a metabolite $i$ by its pseudoisomer group, $i$. Therefore, the values of the standard Gibbs free energy of formation of the protonated species $j$ ($\Delta_f G_i^0$) are required for the estimation of the standard Gibbs free energy of formation of the metabolite $i$ ($\Delta_f G_i^{0*}$) at different values of $I$, $T$, and $pH$.

There exist databases with thermodynamic information for the species $j$ for several metabolites (e.g. [22,23]). When these data are not available then group contribution methods and other predictive methods can be used for their estimation, for example [24,25].

In this paper, we use the extended group contribution method proposed by Jankowski et al. [24] to illustrate the procedure for the calculation of $\Delta_f G_i^{0*}$ for pseudoisomer group $i$. The extended group contribution method [24] allows to estimate $\Delta_f G_j^0$ of the predominant protonated species $j$ at standard biological conditions $pH = 7$, $I = 0$ M and $T = 298.15$ K.

If $\Delta_f G_j^0$ of one species $j$ is known the Gibbs energy for the other protonated species can be calculated. For example, in the dissociation reaction $HA \xrightarrow{pK_a} A^-$, assuming that $\Delta_f G_{HA}^0$ of the uncharged species $HA$ is estimated by the Jankowski’s group contribution method, the change in standard Gibbs free energy for the negatively charged species $A^-$, $\Delta_f G_{A^-}^0$, can be calculated by Eq. (1) [26]:

\[
\Delta_f G_{A^-}^0 = \Delta_f G_{HA}^0 + pK_aRT \ln(10)
\]

Here, the $pK_a$ values of species with known molecular structures can be estimated by the Calculator Plugins, MarvinSketch ver. 5.5.1.0, 2011, ChemAxon (http://www.chemaxon.com).
Additional information such as the standard enthalpy of formation of metabolite \( i \) at 298.15 K and \( I = 0 \) M (\( \Delta_f H^0_i \)) is required to incorporate the effect of temperature in thermodynamic analysis. However, these values are not available for most biological compounds. Despite the existence of several group contribution methods for enthalpy estimation [27–31], none of them considers molecules with atoms of phosphorus, a common situation in metabolites. To address this problem approximate values are estimated as follows.

We assume a hypothetical reaction where only the enthalpy of the metabolite in question is unknown (while the enthalpies of all other reactants are reported in the literature or estimated by some existing method).

Assuming that the energy released during the breaking of chemical bonds is completely consumed in the formation of other bonds, i.e. the standard enthalpy change of this hypothetical reaction (\( \Delta_f H^0 \)) is equal to zero, then from \( \Delta_f H^0 = \sum_{i=1}^{m} \eta_i \Delta_f H^0_i \) the unknown \( \Delta_f H^0 \) can be calculated as a function of the standard enthalpy change of the other metabolites \( \Delta_f H^0_i \). Here \( \eta_i \) is the stoichiometric coefficient of the metabolite \( i \) (negative for substrates and positive for products).

For example a molecule of glucose 6-phosphate (g6p) can be formed from glucose (glc) and inorganic phosphate (pi), so

\[
\text{glc} + \text{pi} \rightarrow \text{g6p} + \text{H}_2\text{O}
\]

(2)

According to Alberty [22] \( \Delta_f H^0_{\text{glc}} = -301.67 \) kcal mol\(^{-1}\), for phosphate with charge -2 \( \Delta_f H^0_{\text{ph} \text{2}^-} = -310.46 \) kcal mol\(^{-1}\), and \( \Delta_f H^0_{\text{H}_2\text{O}} = -68.31 \) kcal mol\(^{-1}\). Hence, \( \Delta_f H^0_{\text{g6p}} \) can be calculated as \( \Delta_f H^0_{\text{g6p}} = \Delta_f H^0_{\text{glc}} + \Delta_f H^0_{\text{ph} \text{2}^-} - \Delta_f H^0_{\text{H}_2\text{O}} \), which is equal to -543.82 kcal mol\(^{-1}\), while Alberty reports \( \Delta_f H^0_{\text{g6p} \text{2}^-} = -544.08 \) kcal mol\(^{-1}\).
Since only small differences are found between the experimental values of $\Delta_f H^0$ for g6p$^{-1}$ and g6p$^{-2}$ as well as for pi$^{-1}$ and pi$^{-2}$ [22], in order to simplify the calculations, we use $\Delta_f H_{p}^{0}$ for the formation of all the phosphorylated metabolites. Furthermore we do not differentiate between the $\Delta_f H_j^0$ of protonated species $j$ of the same pseudoisomer group $i$. While this procedure is only an approximation and requires experimental data to validate the computed results, it gives us at least the qualitative effect of temperature on the system.

With the above thermodynamic data, the transformed Gibbs free energy change of the protonated species $j$ $\Delta_f G_j^{0}$ can be calculated as a function of the environmental conditions $I$, $T$, and $pH$ as follows [13,22]:

$$\Delta_f G_j^{0} = \left(\frac{T}{298.15}\right)\Delta_f G_j^{0} + \left(1 - \frac{T}{298.15}\right)\Delta_f H_j^{0} - \frac{A(T)z_j^i I_j^\frac{1}{2}}{1 + BI_j^\frac{1}{2}} \cdot RT \ln(10) - N_{H,j} \Delta_f G_j^{0}$$  \hspace{1cm} (3)

where

$$\Delta_f G_j^{0} = \left(\frac{T}{298.15}\right)\Delta_f G_j^{0} + \left(1 - \frac{T}{298.15}\right)\Delta_f H_j^{0} - \frac{A(T)I_j^\frac{1}{2}}{1 + BI_j^\frac{1}{2}} \cdot RT \ln(10) - RT \ln(10^{pH_{H^+}^{ref}})$$  \hspace{1cm} (4)

$$A(T) = \frac{1.10708 - 1.54508 \times 10^{-3} T + 5.95584 \times 10^{-6} T^2}{\ln(10)}$$  \hspace{1cm} (5)

$A(T)$ and $B = 1.6$ are the constants of the extended Debye-Hückel equation, $N_{H,j}$ and $z_j$ are the number of H atoms in the protonated species $j$, and their corresponding electric charge, respectively. The reference $pH$ is set as in the standard biological conditions $pH_{ref} = 7$.

The first two terms of the right-hand side of Eq. (3) represent the effect of $T$ and the third one the effect of $I$ (whose physiological range is 0.15 M to 0.2 M [19]). Note that $A(T)$ is also a function of $T$ [22]. The fourth term reflects the $pH$ influence. $pH$ is assumed to be constant, therefore the hydrogen ion $H^+$ is eliminated from the biochemical reactions [22].
Finally, in order to estimate $\Delta_f G_i^{0\ast}$ of metabolite $i$ (represented by its pseudoisomer group $i$), we assume that all species $j$ are in equilibrium. Then from [22]:

$$
\Delta_f G_i^{0\ast} = -RT \ln \left( \sum_{j=1}^{N_{\text{species}}} \exp \left( \frac{-\Delta_f G_j^{0\ast}}{RT} \right) \right)
$$

(6)

The methodology described above to calculate $\Delta_f G_i^{0\ast}$ is depicted in Fig. 1.

---

**Fig. 1.** Methodology for the calculation of $\Delta_f G_i^{0\ast}$ for pseudoisomer group $i$.

2.2. Coupling mass balances and thermodynamic constraints

The estimation of thermodynamically feasible flux distributions is based on the solution of an optimisation problem constrained by the mass balance and the thermodynamics of the system.

**Mass balance constraints.** The general assumption of the stoichiometric-based models is that there is no accumulation of intracellular metabolites, i.e. the biochemical reactions are at steady state.
\( \mathbf{S} \cdot \mathbf{v} = 0 \) \hspace{1cm} \text{(7)}

\( \mathbf{S} \) is a stoichiometric matrix with dimensions \( m \times (n+N) \), containing information of \( m \) intracellular metabolites (the hydrogen ion \( \text{H}^+ \) is not considered a reactant in biochemical reactions, and therefore not included either in the mass balance or in \( \mathbf{S} \)) involved in the \( n \) intracellular reactions and the \( N \) transport reactions form/to outside/inside the cell. Finally \( \mathbf{v} \) is the \((n+N) \times 1\) flux vector.

**Thermodynamic constraints.** The enzyme involved in the conversion of a reactant to a product alters the rate of the reaction but not its thermodynamic feasibility. Hence, when only the mass balance (Eq. (7)) is used to predict the flux distribution of a system, it can include fluxes that violate the second law of the Thermodynamics [19].

The Gibbs free energy of reaction, \( \Delta_r G' \) gives information about feasibility, directionality and reversibility of a reaction in the metabolic network.

\[
\Delta_r G' = \Delta_r G_{br}' + \Delta_r G_v'
\]

where

\[
\Delta_r G_{br}' = \sum_{i=1}^{\text{Products}} \eta \left[ \Delta_f G_{i}^{\text{react}} + RT \ln(C_i) \right] - \sum_{i=1}^{\text{Reactants}} \eta \left[ \Delta_f G_{i}^{\text{prod}} + RT \ln(C_i) \right]
\]

\[
\Delta_r G_v' = \eta \left[ -RT \ln(10^{-pH_{i,max}} + pH_{i,min}) \right] + \eta F \Delta \phi
\]

\( \Delta_r G_{br}' \) (Eq. (9)) represents the contribution to the Gibbs free energy due to an intracellular biochemical reaction, \( \eta \) is the stoichiometric coefficient of the metabolite \( i \), and \( C_i \) the concentration of the metabolite \( i \).

Strictly speaking \( \Delta_r G_{br}' \) is a function of the activities of the metabolites \( a_i \) rather than their concentration \( (C_i) \). Assuming ideal solution, and the metabolites as volumeless molecules, the activity can be calculated as \( a_i = C_i / C_i^{\text{std}} \), where \( C_i^{\text{std}} \) is the standard concentration of the metabolites, which is equal to 1 M. Therefore, the value of \( a_i \)
is equal to the value of \( C_i \) (moles per liter). In this paper we use the term *concentration*. A brief discussion about \( a_i \) is presented in Section 3.1.

The concentration of water is considered much higher than any other metabolite, and its activity is set equal to 1 for the estimation of \( \Delta_r G_{br}^- \) [22].

Eq. (10) expresses the contribution to the Gibbs free energy from the hydrogen ion \( \text{H}^+ \) transport across the cellular membrane (\( \Delta_r G_{br}^- \)) when this is coupled to an intracellular reaction, e.g. in the synthesis of succinic acid (succ) from fumarate (fum) by the fumarate reductase [32]:

\[
\text{fum}_c + \text{NADH}_c + \text{H}^+_e \leftrightarrow \text{succ}_c + \text{NAD}_c \\
2\text{H}^+_c \leftrightarrow 2\text{H}^+_e 
\]

where the subscripts \( c \) and \( e \) refer to the intracellular and extracellular/medium, respectively.

The value of \( \Delta_r G_{br}^- \) depends on the \( \text{H}^+ \) concentration on both sides the membrane (first term of the right hand side of Eq. (10)) and on the electrical potential difference of the membrane (second term of the right hand side of Eq. (10)).

The symbol \( \eta_{\text{H}^+} \) is the number of \( \text{H}^+ \) cations crossing the membrane, \( p\text{H}_c \) and \( p\text{H}_e \) are the intracellular and extracellular/medium pH respectively, \( F \) is the Faraday constant equal to \( 2.306 \times 10^2 \text{ kcal mV}^{-1} \text{ mol}^{-1} \). Finally, \( \Delta\varphi \) is the difference between the internal potential of the membrane and the external one, given by [19]:

\[
\Delta\varphi = 33.33(p\text{H}_c - p\text{H}_e) - 143.33
\]

A non-linear optimisation (non-linear programming –NLP) formulation (Eq. (13–20) below) is proposed to integrate the mass balance and thermodynamic constraints for the estimation of flux distributions. The objective function \( \text{func} \) (Eq. (13)) can comprise the maximisation/minimisation of any variable of the system, i.e. either the net flux (\( v_k \)) or
the Gibbs free energy change ($\Delta_r G_k'$) of a reaction $k$, or the concentration of a metabolite $i$.

\[
\maximise/\minimise \quad \text{func} \\
\text{s.t.} \\
S \cdot v = 0 \quad \tag{13}
\]

\[
v_{\min,k} \leq v_k \leq v_{\max,k} \quad \tag{14}
\]

\[
C_{\min,i} \leq C_i \leq C_{\max,i} \quad \tag{15}
\]

\[
v_k \cdot \Delta_r G_k' \leq 0 \quad \tag{16}
\]

\[
\abs(v_k) + \frac{\Delta_r G_k'}{\tol} v_k \leq 0 \quad \tag{17}
\]

\[
d_k \left( \abs(\Delta_r G_k') + \frac{v_k}{\tol} \Delta_r G_k' \right) \leq 0 \quad \tag{18}
\]

\[
\Delta_r G_k' = \Delta_r G_{v,k}' + \Delta_r G_{w,k}' \quad \tag{19}
\]

Here, the thermodynamic analysis is focused on the intracellular reactions (constraints given by Eq. (17–20)). However, the mass balance (Eq. (14)) also includes the exchange flux to/from the extracellular medium. Eqs. (15) and (16) represent the lower/upper limits of the variables $v_k$ and $C_i$, respectively.

Since the net flux of a reaction $k$, $v_k$, goes into the direction of the Gibbs free energy drop, variables $v_k$ and $\Delta_r G_k'$ should have opposite sign. This condition is fulfilled by Eq. (17).

Eq. (18) guarantees that if $\Delta_r G_k'$ is predicted to be zero, i.e. $\abs(\Delta_r G_k') < \tol$, then $v_k$ will take a value close to zero. This is ensured through the term $\Delta_r G_k'/\tol$, which becomes small when $\abs(\Delta_r G_k') < \tol$, so that $v_k$ has to take also a small value to satisfy the constraint. This restriction is valid for both an equilibrium and a block state.

When $\Delta_r G_k'$ is not zero, then Eq. (18) is automatically satisfied with any value of $v_k$ (either zero or a value with opposite sign to that of $\Delta_r G_k'$). This is because the second
term of Eq. (18) \((\Delta_r G_k' v_k / tol)\) will be always negative with greater magnitude than that of the first term \((abs(v_k))\).

However for the *equilibrium* state an additional restriction is considered: if during the optimisation the flux through reaction \(k\) is estimated to be zero, i.e. \(abs(v_k) < tol\), then \(\Delta_r G_k'\) should be zero as well. This is covered by Eq. (19), which works similarly to Eq. (18).

The parameter \(d_k\) (Eq. (19)) is equal to zero if the reaction \(k\) is considered beforehand as blocked, or equal to 1 when it is considered at equilibrium, when the net flux for this reaction is predicted to be zero. Only when \(d_k = 1\) the constraint Eq. (19) is active, i.e. in the *equilibrium* case.

The selection of the \(d_k\) value depends on experimental observations, i.e. if the presence of the enzyme has been detected under the environmental conditions analysed, then just the *equilibrium* state of the reaction can explain a net flux equal to zero. For example, small amounts of ethanol were detected in the production of succinic acid from glucose by *A. succinogenes* [33], so it is likely that using glucose as substrate, the reaction ALCD2x (see Fig. 2 in Section 3) is at thermodynamic equilibrium if its net flux is estimated as zero.

The variables of the system, Eq. (13–20), are the net flux of the reaction \(k\) \(v_k\), the natural logarithm of the concentration of the metabolite \(i\) \((\ln(C_i))\), and the Gibbs free energy of the reaction \(k\) \((\Delta_r G_k')\) which is related with \(\ln(C_i)\) by Eq. (8) and (9). Finally, \(tol\) is the tolerance value for the flux and the Gibbs free energy change.

Note that a zero flux (or \(\Delta_r G_k'\)) is computationally equivalent to \(-tol < v_k < tol\). The \(tol\) value can be changed by the user, e.g. to the threshold of detection of lab sensors. In this paper, we use \(tol = 10^{-4}\), which is below the analytical detection limit.
The $\Delta_i G_{k}^\prime$ value given by Eq. (20) neglects the uncertainty $SE$ associated with the transformed standard Gibbs free energy of the reaction, $\Delta_i G_{m}^{\prime} = \sum_{i=1}^{\text{Products}} \eta_i \Delta_j G_{i}^{\prime} - \sum_{i=1}^{\text{Reactants}} \eta_i \Delta_j G_{i}^{\prime}$, estimated by the group contribution method [24].

The incorporation of the influence of the uncertainty can be accomplished by the addition of an error variable $E$, allowed to vary within the limits $\pm SE$, so that Eq. (20) becomes

$$\Delta_i G_{m}^{\prime} = \sum_{i=1}^{\text{Products}} \eta_i \left[ \Delta_j G_{i}^{\prime} + RT \ln(a_i) \right] - \sum_{i=1}^{\text{Reactants}} \eta_i \left[ \Delta_j G_{i}^{\prime} + RT \ln(a_i) \right] + E \quad (21)$$

It is worthwhile to note here that while the prediction of a negative net flux value indicates that a reaction goes in the opposite direction to that indicated in the stoichiometric matrix ($S$), no further corrections on the initial assumption of the reaction direction are needed in this methodology, since the optimisation algorithm described by Eq. (13–20) guarantees that fluxes and Gibbs free energy changes maintain the correct directionality. We just need to keep in mind the way the reactions were written in $S$ for the interpretation of the results.

It is well known that the non-linear problem described by Eq. (13–20) can have a number of local minima. Good initial guesses are therefore essential in order to compute global optimum. For this purpose, the following initial guesses (composed by all fluxes $v_k$, metabolite concentrations $\ln(C_i)$, and Gibbs free energy changes $\Delta_i G_{k}^\prime$) are proposed based on the type of the objective function ($func$) chosen.

(a) If $func = v_j$, $f \leq n$ (i.e. intracellular fluxes) for the NLP problem then initial guesses for fluxes $v_k$, $(k \neq f)$ metabolite concentrations $\ln(C_i)$, and Gibbs free energy changes $\Delta_i G_{k}^\prime$ can be provided from the solution of 2 linear optimisation problems (i) the corresponding Flux Variability Analysis (FVA) given by Eq. (13–15) where $func = v_j$ and (ii) the maximisation (and/or minimisation) of $\Delta_i G_{j}^\prime$ using only the linear thermodynamic constraints (Eq. (9, 10, 13, 16, 20)), which will provide a vector of $\Delta_i G^\prime$ and of $\ln(C)$ as solutions.

(b) If $func = v_j$, $n < f < N$ (i.e. extracellular fluxes) for the NLP problem, then initial guesses for the fluxes are obtained as in (a). Because in this paper thermodynamic
analysis does not include transport (uptake/secretion) reactions (i.e. extracellular fluxes), we simply use as an initial guess of the Gibbs free energy changes a unit vector where each of its elements has the opposite sign of the corresponding intracellular flux. The initial guesses for $\ln(C)$ are $(\ln(C_{\text{init}}))$ and $(\ln(C_{\text{max}}))$.

(c) If $func = \ln(C_i), \ 1 \leq nr \_of \_species$, then initial guesses for the fluxes are the upper limit ($v_{\text{max}}$) and the lower limit ($v_{\text{min}}$) while Gibbs free energy changes $\Delta_r G'_k$ are computed from the linear optimisation problem of maximising (and/or minimising) $func$ subject to Eq. (9, 10, 16, 20).

(d) If $func = \Delta_r G'_i, \ f \leq n$ then initial guesses are computed as in (c).

Although linear optimisation runs are required to compute these initial guesses, their execution time is short, and they allow the computation of solutions in the vicinity of global minimum (the results were compared to those from linear-methodology TMFA [19] showing that the values are similar, see Section 3.1) and significantly reduce the computational time required by the non-linear solver.

To summarise: our model assumes that 1) the intracellular temperature $T$ is constant and equal to that in the medium, 2) the protonated species $j$ of the metabolite $i$ are at equilibrium, 3) there is no local variation in the concentration of the metabolites meaning that the cytoplasm is completely homogeneous, and 4) the molecules are volumeless points over which molecular crowding inside the cell has no effect.

3. Results and discussion

In order to determine the effect of considering metabolic reactions either as \textit{blocked} or at \textit{equilibrium}, on the prediction of flux distributions, the central carbon pathways of \textit{A. succinogenes} (Fig. 2) was used as an illustrative metabolic system (comprising 53 reactions and 41 metabolites, see Supplementary Material for detailed information).

Based on the experimental data reported by Vlysidis et al. [3] for the batch production of succinic acid from glycerol by \textit{A. succinogenes} during the exponential phase, we
carried out Flux Variability Analysis (FVA) [34] to estimate the intracellular as well as the other unmeasured fluxes.

Here, the inputs and outputs to/from the cell were set to the experimentally measured fluxes for the consumption of glycerol (5.6524 mmol g_{DW}^{-1} h^{-1}), as well as the production of succinate (3.8711 mmol g_{DW}^{-1} h^{-1}), acetate (0.5965 mmol g_{DW}^{-1} h^{-1}), formate (0.4735 mmol g_{DW}^{-1} h^{-1}), and of biomass (1.8504 mmol g_{DW}^{-1} h^{-1}). The cell composition was assumed to be CH_{2}O_{0.5}N_{0.2} [35]). No ethanol, fumarate, lactate, oxaloacetate, or pyruvate were experimentally detected, thus their exchange fluxes to the medium were set to zero.
Fig. 2. Metabolic network of succinate production from glycerol by *Actinobacillus succinogenes*. Big arrows indicate how the reaction was written in the stoichiometric matrix, and if the reaction is considered irreversible (unidirectional arrow) or reversible (bidirectional arrow).
The upper and lower concentration bounds of all the metabolites in the network were set to physiologically reasonable values $C_{\text{max}} = 0.02 \text{ M}$ and $C_{\text{min}} = 10^{-5} \text{ M}$, respectively, while for the intracellular fluxes the corresponding limits were $\nu_{\text{max}} = 100 \text{ mmol g}_{\text{DW}}^{-1} \text{ h}^{-1}$ and $\nu_{\text{min}} = -100 \text{ mmol g}_{\text{DW}}^{-1} \text{ h}^{-1}$ [19]. The uptake/secretion of water was unconstrained.

Due to the nonlinear nature of the optimisation problem, the calculation of global minimum can be a computationally expensive task, requiring the use of appropriate deterministic methods [36] or of stochastic methods [37] (which probabilistically tend to converge to the vicinity of the global optimum, which is however not theoretically guaranteed). In this work we use the local NLP solver fmincon in MATLAB R2011a (The MathWorks, Natick, MA), with 4 different initial guesses for each variable (as discussed in Section 2.3).

The solution obtained by our methodology using the initial guesses described above, where all reactions with zero net flux were considered to be blocked, $d = 0$ were compared with the results given by TMFA [19] for validation purposes. The latter method was programmed in Matlab using the MILP solver BNB20 (Kuipers, 1998; available in www.mathworks.com/matlabcentral). The results show very good agreement between the two methods.

For example computing the range of fluxes of each of the 53 reactions (one-by-one) under condition 1 (i.e. the biological standard conditions), the 4 initial guesses above and the system given by Eq. (13–20) predicts the same minimum and maximum flux values as the ones given by TMFA for 49 of the reactions.

This agreement can be further improved if several variables (fluxes, $\Delta G^\prime$, and/or metabolite concentrations, each one with its own 4 initial guesses) are solved consecutively. The new results can be compared with those previously obtained in order to choose the best (maximum/minimum) value for each variable, which is similar to increasing the number of initial guesses per variable.
We next employ NLP using Eq. (13–20) to analyse the A. succinogenes central metabolism considering not only the block case, but also equilibrium one.

The fact that the net flux through a reaction is equal to zero can be due to 1) the reaction being blocked because the enzyme that catalyses the reaction is absent or 2) the reaction being at equilibrium. The absence of an enzyme could be interpreted as a consequence of genetic modification or regulation, irreversible inhibition of the enzyme, etc. Nevertheless, the regulation mechanisms are complex and not fully understood, and it is possible that the enzymatic activity of a reversible reaction could not be completely eliminated or delayed, so that exchange fluxes (in both reaction directions) can be present but with a zero net flux.

For example, $^{13}$C-experiments carried out on the glucose and fructose metabolism of Zymomonas mobilis [38] indicate that the 2 reactions catalysed by transketolase are at equilibrium (zero net flux with non-zero exchange fluxes were detected). Despite the fact that there were uncertainties in the reactions used in the model of the metabolism, which could change the flux distribution estimated the possibility to find a reaction at equilibrium cannot be completely ruled out.

As mentioned before, in the A. succinogenes case study the external fluxes for lactate, ethanol, fumarate, pyruvate, and oxaloacetate were set to zero. This could cause some internal fluxes to also be estimated as zero according to the mass balance of the system. In the case of ethanol, we expect that reactions ACALD (catalysed by acetaldehyde dehydrogenase) and ALCD2x (alcohol dehydrogenase) (see Fig. 2) to have zero net flux. Certainly the enzymatic activity levels depend on the microorganism capabilities and the environmental conditions, so the activity of acetaldehyde dehydrogenase and alcohol dehydrogenase were detected in A. succinogenes growing on glucose, sorbitol, and fumarate [33], but there is no data available for the growth on glycerol.

Also, PFK and FBP were considered as irreversible in the indicated direction, however the flux through FBP should be greater than that through PFK in order to allow the growth of the microorganism on glycerol (and vice versa if glucose is used as substrate). So if the flux through PFK (catalysed by 6-phosphofructokinase) were
predicted as zero in the \textit{equilibrium} case, this would mean that the reaction is at equilibrium. The presence of this enzyme could aid the quick adaptation of the microorganism to sudden changes in the environment e.g. the carbon source, making the system more robust.

In this paper, these zero flux reactions are used as an illustrative example to show the potential effect of considering a reaction at equilibrium instead of blocked, on the flux distribution.

Nevertheless, a number of studies has shown that some reactions cannot be feasible under certain environmental conditions ($I, T, pH_c, pH_e$) \cite{39}, which directly affect their $\Delta_r G^{0*}$. Furthermore, it is well known that the feasibility of a reaction is determined by its $\Delta_r G'$ which may differ from $\Delta_r G^{0*}$ due to the contribution of reactants and products’ concentrations involved in that reaction.

Hence, in order to compare the influence of the environmental conditions, five sets of parameters were tested (Table 1). Conditions 2-5 comprise changes in each of the variables ($pH_c, pH_e, I, T$), respectively while condition 1 represents standard biological conditions for all variables.

\begin{table}[h]
\centering
\begin{tabular}{cccccc}
\hline
 Condition 1 & Condition 2 & Condition 3 & Condition 4 & Condition 5 \\
 (biological standard conditions) & (change in $pH_c$) & (change in $pH_e$) & (change in $I$) & (change in $T^a$) \\
\hline
$pH_c = 7$ & $pH_c = 6.7$ & $pH_c = 7$ & $pH_c = 7$ & $pH_c = 7$ \\
$pH_e = 6.2$ & $pH_e = 6.2$ & $pH_e = 7.4$ & $pH_e = 6.2$ & $pH_e = 6.2$ \\
$T = 298.15$ K & $T = 298.15$ K & $T = 298.15$ K & $T = 298.15$ K & $T = 310.15$ K \\
$I = 0$ M & $I = 0$ M & $I = 0$ M & $I = 0.2$ M & $I = 0$ M \\
\hline
\end{tabular}
\caption{Sets of environmental conditions tested for the estimation of \textit{A. succinogenes}' flux distribution}
\end{table}

\textsuperscript{a} It is assumed that the temperature of the medium is equal to the intracellular temperature.
In the following sections the two extreme cases, where all the reactions are considered either blocked \((d=0)\) in Eq. (19)) or at equilibrium \((d=1)\) in Eq. (19), i.e. all the enzymes are actives) when their net flux is zero, will be analysed, as well as the impact of the environmental changes \((I, T, pH_c\) and \(pH_e\)) on the feasibility of all reactions and their corresponding flux distributions. For this the (group contribution-based) uncertainty in the calculation of \(\Delta_r G^{0r}\) is considered to be zero.

3.1. Effect of \(I, T, pH_c\), and \(pH_e\) changes on \(\Delta_r G^{0r}\)

The comparison of the values of \(\Delta_r G^{0r}\) calculated at different sets of conditions (see Table 1) reveals that 36.11\% of the 36 intracellular reactions reached their lowest \(\Delta_r G^{0r}\) value when a change in \(I\) to the physiological value \((I = 0.2)\) was made, while 36.11\% of reactions when \(T\) was set to 310.15 K. 16.66\% of reactions reached their lowest \(\Delta_r G^{0r}\) value through a change in intracellular \(pH_c\), and just 2.77\% through change in extracellular \(pH_e\). The latter percentage was expectedly small since only 1 reaction is coupled to the transport of \(H^+\) from/to the medium.

According to this, \(\Delta_r G^{0r}\) of the reactions are more sensitive to changes in \(I\) and \(T\). However, the thermodynamic analysis is more complex, because while \(\Delta_r G^{0r}\) values have decreased for some reactions (for changes in \(I\) or \(T\)), for others they have increased. Also in the real system, changes in \(I\) and \(T\) are taking place at the same time. Therefore only solving the problem as a whole (with mass balance and thermodynamic constraints) taking in account changes in \(\Delta_r G^*\) (not just \(\Delta_r G^{0r}\)) for each reaction will allow a better understanding of the system, and thus a better prediction for the flux and concentration distribution.

It should be noted here that in this system the differences between \(\Delta_r G^{0r}\) at standard biological conditions and under the other 4 tested conditions above were significant (greater than 1 kcal mol\(^{-1}\)) for 10 reactions, while these differences were found to be less than 1 kcal mol\(^{-1}\) for 26 reactions.
3.2. Prediction of thermodynamically feasible flux distributions

The classification of a reaction as reversible/irreversible depends on its predicted range of flux. If after performing FVA no sign change is predicted between the min and the max \( v_k \) of a reaction \( k \), this reaction is considered to be irreversible for the specific set of conditions tested, otherwise is considered as reversible.

For the metabolic pathways shown in Fig. 2 FVA using only the mass balance as a constraint (Eq. (13–15)) predict 12 reactions as reversible, while FVA considering both mass and thermodynamic constraints, here designated as tFVA (Eq. (13–20)) estimates that only 6 reactions in the block case (i.e. all reactions with zero flux considered blocked), and 5 in the equilibrium case (i.e. all reactions with zero flux considered at equilibrium) comprise flux in both directions (Fig. 3A), reducing in this way the size of the parameter space of the network.

![Fig. 3. Comparison of the number of reactions predicted as reversible (A) and substitutable (B) by FVA, and tFVA in both block and equilibrium cases under condition 1. No uncertainty in \( \Delta G^\circ \) is considered.](image)

As mentioned above, changes in \( I, T, pH_c \) and \( pH_e \) affect the \( \Delta G^\circ \) values at different extent causing differences in the fluxes predicted at these conditions as shown in Fig. 4 (coloured lines), where e.g. reaction PPCK is only computed as reversible for conditions 4 in both the equilibrium and the block case.
**Fig. 4.** Range of fluxes predicted for some reactions under different conditions. Colour lines indicate the flux range predicted when the uncertainty in $\Delta G^0r$ is neglected, and grey lines behind them are the ranges estimated when the uncertainty is considered. Reactions MDH and PPCK are reversible only for condition 4 in both *block* and *equilibrium* cases, and ME2 for all the five conditions, in the *block* case, while for the *equilibrium* case under condition 1 to 4. Besides, MDH is predicted as substitutable for condition 1,2,3,5 in the *block* case, but in the *equilibrium* case only under condition 1. No uncertainty is considered.

Furthermore, by classifying the reactions in the network as [19]:

1) essential in the direction indicated when their minimum and maximum flux have the same sign,

2) substitutable when they have either a zero minimum flux and a positive maximum flux or a zero maximum flux and a negative minimum flux,

3) blocked or at equilibrium (depending on which assumption was made) when both minimum and maximum flux value are zero,

it was found that the number of reactions classified as substitutable diminish from 7 for the *block* case to 2 for the *equilibrium* case as shown in Fig. 3B.

The results computed through tFVA show that for the same conditions, the range of fluxes predicted for the *equilibrium* case are narrower than those calculated for the *block* case as shown in Fig. 5A. This behaviour was expected since the *equilibrium* case involves more stringent restrictions.
Fig. 5. Range of thermodynamically feasible fluxes predicted for some reactions involved in the central carbon pathways of A. succinogenes. (A) No uncertainty in $\Delta_r G^{0\text{r}}$ is considered. (B) The uncertainty in $\Delta_r G^{0\text{r}}$ is taken into account. Coloured lines indicate the fluxes at different conditions of $I$, $T$, $pH_c$, and $pH_e$ for the equilibrium case, while grey lines behind them are the flux range predicted for the block case at the same conditions.

As shown in Fig. 4 and 5, often the range of fluxes predicted (for the same reaction) under conditions 4 and 5 differ from those estimated at standard conditions (condition 1). This is due to the $\Delta_r G^{0\text{r}}$ values of the reactions in the system being more sensitive to changes in $I$ and $T$ (conditions 4 and 5, respectively).

The results on the impact of $I$ and in less extent $T$ on the range of fluxes predicted for some reactions contrast with the results of Henry et al. [19] on the low influence of the parameter $I$. These differences come from the use of the predominant species in [19] (i.e. the most abundant protonated species) instead of the pseudoisomer group (i.e. all the protonated species formed during the dissociation of a metabolite in aqueous solutions, as was used in this work) to represent the thermodynamic properties of the metabolites involved in the network. This is corroborated by the flux distribution we obtained from tFVA – block case using the predominant species at conditions 1 and 4, which reveals that the flux range is not affected by the $I$ change.

Moreover, the impact of $pH_c$ and $pH_e$ (conditions 2 and 3 respectively) was not significant on the prediction of flux distribution compared with the standard biological condition (condition 1). The tested values did not drastically change because A.
succinogenes, as most aerobic and facultative anaerobic bacteria, regulates the intracellular pH within a narrow range [33], and the medium pH was controlled to be between 6.2 and 7.4 during the fermentation [3]. Nevertheless, the influence of these two parameters may be higher if the microorganism analysed is strictly anaerobic like Anaerobiospirillum succiniproducens, which decreases pH when pH is decreased [33].

When the uncertainty in the estimation of \( \Delta G^0^* \) is taken into account (Eq. 21), the thermodynamic constraints are relaxed, hence the range of fluxes predicted by tFVA (grey lines in Fig. 4) is wider than those estimated for zero uncertainties (colour lines in Fig. 4). In fact, when the uncertainty is considered the influence of the environmental conditions on the flux distribution is almost completely concealed for both the block and the equilibrium states (grey lines in Fig. 4 and grey and coloured lines in 5B). However, some differences were found between the cases where all the reactions are considered either blocked or at equilibrium when their net fluxes are zero (Fig. 5B).

Another important parameter is the intracellular concentration of the metabolites, which depends on many factors such as the composition of the medium, e.g. carbon source, the physiological state and type of the cell, etc. Often experimental values are not available for all the metabolites. In these cases the assumption of default physiological limits is the best option.

Fig. 6 depicts a comparison of the range of fluxes estimated by tFVA - block case (and zero uncertainty in \( \Delta G^0^* \)) when the concentration limits were set as \( C_{\text{max}} = 0.01 \text{ M} / C_{\text{min}} = 10^{-5} \text{ M} \) (colour lines) and \( C_{\text{max}} = 0.02 \text{ M} / C_{\text{min}} = 10^{-5} \text{ M} \) (grey lines) under conditions 1–5. Here, it is possible to see that when these limits are very wide (such as those used in this paper, i.e. \( 10^{-2} – 20 \text{ mM} \)), the influence of changes on environmental conditions could be damped and/or hidden and slight or no differences in the range of fluxes are found compared with those at biological standard conditions (i.e. a reaction is still computed as thermodynamically feasible under a wide range of conditions). While for \( 10^{-2} – 10 \text{ mM} \), differences among the fluxes predicted are more evident at different pH, I and T.
Fig. 6. Range of fluxes predicted for some reactions involved in the metabolic network of *A. succinogenes* using different concentration limits. Grey lines indicate the range of fluxes predicted by tFVA – *block case* for the limits $C_{\text{max}} = 0.02 \text{ M} / C_{\text{min}} = 10^{-5} \text{ M}$, while colors lines indicate the fluxes obtained using the limits $C_{\text{max}} = 0.01 \text{ M} / C_{\text{min}} = 10^{-5} \text{ M}$ at different conditions of $I$, $T$, $pH_c$, and $pH_e$. No uncertainty in $\Delta_r G^{0'}$ is considered.

Although the results obtained using $10^{-2} – 20 \text{ mM}$ are mathematically feasible, it is likely that the solution space (range of fluxes, concentration and $\Delta_r G'$) computed is greater than the actual one, hence it cannot capture the complex impact the environmental conditions can have on the system. Unfortunately, no information is available about the intracellular metabolite concentration for *A. succinogenes* growth on glycerol. Nevertheless, the uncertainty in the direction of a reaction and its corresponding flux range is reduced when the thermodynamic constraints are taken into account (tFVA) compared with only the mass balances (FVA) (see coloured lines in Fig. 4).

Another issue related to the concentration limits used in the model is the following. Vojinovic and von Stockar [39] tested different ranges of concentrations of the metabolites, including experimental values, involved in the lactic acid pathway. Using these data and the thermodynamic analysis proposed by Mavrovouniotis [10], they found that even for this well-known pathway there are problems in the correct prediction of the feasibility of the reactions. This is due to the uncertainties in the
experimental measurements of concentrations, Gibbs free energy, and enthalpies of the reactions.

While the influence of uncertainties (associated with the experimental measurements and/or the contribution method used) on the prediction of the feasibility of the reactions, and therefore on the flux distribution, is unquestionable, also the molecular crowding conditions within the cell could help to explain why even with a small number of reactive molecules, i.e. low metabolites’ concentrations \( C_i \), a reaction is feasible.

This is because, the metabolites and other cellular components are not volumeless molecules (as it is assumed in the stoichiometric models), and the presence of macromolecules, which could occupy up to 40% of the total cell volume \( \text{vol}_{\text{tot}} \) [40], reduces the available volume for the reactants inside the cell \( \text{vol}_{\text{av}} \), thus the activity of the metabolites \( a_i = \text{vol}_{\text{av}} C_i / \text{vol}_{\text{av}} \) is enhanced by several orders of magnitude.

This enhancement in the metabolite’s activity could affect the thermodynamic analysis and therefore alter the flux distribution predicted [42]. Besides, this reduction in the available volume of the molecules causes a decrease in their diffusion and the likely encounter-reaction between reactants [41].

3.3. Prediction of the range of metabolites concentrations and Gibbs free energy of the reactions

As mentioned above, the parameters \( I, T, pH \), affect to a different extent the \( \Delta_f G^{0'} \) values, therefore it is expected that the range of metabolites’ concentrations will change, in response to them, as well. For the equilibrium case the estimated range of concentrations of some metabolites is narrower than that obtained for the corresponding block case as shown in Fig. 7 (no uncertainty in \( \Delta_f G^{0'} \) is considered). Moreover, the number of metabolites whose range differs from the “default” range \( 10^{-2} – 20 \) mM for the equilibrium case, is higher than for the block case.
Fig. 7. Range of concentrations for some metabolites involved in the central carbon pathways of *A. succinogenes*. Coloured lines indicate the fluxes at different conditions of *I*, *T*, *pH<sub>c</sub>*, and *pH<sub>e</sub>* for tFVA – equilibrium case, while grey lines behind them represent the flux range predicted for the block case at the same conditions. No uncertainty in Δ<sub>r</sub>ΔG<sub>r</sub><sup>0</sup> is considered.

In order to determine the range of the Gibbs free energy attainable for the bounds of concentrations proposed, the variable Δ<sub>r</sub>ΔG<sub>r</sub> was maximised and minimised for each reaction assuming zero uncertainty.

The results of our thermodynamic analysis indicate that the number of reactions whose Δ<sub>r</sub>ΔG<sub>r</sub> range (Δ<sub>r</sub>ΔG<sub>r</sub><sub>min</sub> – Δ<sub>r</sub>ΔG<sub>r</sub><sub>max</sub>) is either <−1 kcal mol<sup>−1</sup> or >1 kcal mol<sup>−1</sup>, i.e. reactions that cannot reach the equilibrium and therefore could be candidates for regulation [19], varies with the environmental conditions and the equilibrium/block case assumed (Fig. 8). Only reactions ACYP and OAD were detected as potential regulatory points under all conditions tested. The results also indicate that more reactions were identified as candidates for regulation for the equilibrium case than for the block case. For example under condition 4 the block case predicts 3 reactions while the equilibrium case predicts 5 (Fig. 8).
Fig. 8. Comparison of the number of reactions predicted as candidates for regulation by tFVA – block case and tFVA – equilibrium case under condition 1 to 5. No uncertainty in $\Delta_r G^{0\circ}$ is considered.

It should be mentioned here that when just certain pathways are considered as case study, and not the complete network of the microorganism for either the block or the equilibrium case, in some way it is assumed that the flux through these “not considered” reactions is zero, and their $\Delta_r G^{0\circ}$ values are not taken in account. In other words it is assumed that they are blocked and therefore eliminated from the system.

3.4. Block and equilibrium states in Flux Balance Analysis predictions

Up to this point we have compared the block and equilibrium cases for a metabolic network experimentally constrained with inputs and outputs. In order to determine how these states affect the maximum attainable biomass production, thermodynamically constrained Flux Balance Analysis was carried out on the metabolic pathways of A. succinogenes shown in Fig. 1.

For this, the uptake glycerol flux was set to 5.6524 mmol g_{DW}^{-1} h^{-1}, CO$_2$ to 1.7602 mmol g_{DW}^{-1} h^{-1}, and phosphate to 0.1598 mmol g_{DW}^{-1} h^{-1}, while the extracellular fluxes were left unrestricted. The bounds of the metabolites’ concentration and intracellular flux were as in the previous example $v_{\text{max}} = 100$ mmol g_{DW}^{-1} h^{-1} / $v_{\text{min}} = -100$ mmol g_{DW}^{-1} h^{-1}, and $C_{\text{max}} = 0.02$ M / $C_{\text{min}} = 10^{-5}$ M respectively, and no uncertainty in $\Delta_r G^{0\circ}$ is considered.
Under condition 4, the maximum biomass flux was predicted as $1.8495 \text{ mmol g}_{\text{DW}}^{-1} \text{ h}^{-1}$ for both block and equilibrium states, which agrees with the experimental value. Since there are no other restrictions than the mass balance and thermodynamics, the production of ethanol (and other external metabolites not experimentally detected with the glycerol substrate) is allowed and feasible. Under these conditions only slight differences were found between the range of fluxes of the remaining reactions predicted for the block and equilibrium cases. However, these differences could be more significant if the system is more restricted (as in the previous example of FVA) and/or the number of active reactions considered in the network is larger (and the constraints in one reaction, e.g. the prediction of its irreversibility, affect more reactions at the same time).

4. Conclusions

The conditions $I$, $T$, $pH_c$, and $pH_e$ affect at different extents the prediction of the range of fluxes, concentration of metabolites and $\Delta G'$. The flux distribution for the A. succinogena metabolic network was found to be more sensitive to $I$, followed by $T$. This shows the importance of considering the intracellular environmental conditions on the estimation of thermodynamically feasible metabolic fluxes.

The potential effect of considering a reaction blocked or at equilibrium when the flux through it is zero was also evaluated. For this and unlike other thermodynamically constrained stoichiometric models that assume that a zero flux reaction as blocked, in this work the information about the presence of the enzymes is explicitly incorporated.

The problem of assuming that a reaction is blocked is that unless accurate thermodynamic data and metabolite concentrations are available, it is possible that during optimisation a flux is estimated to be zero (fulfilling the mass balance) only to relax the thermodynamic constraints (the corresponding $\Delta G'$ can take any value) and achieve a better maximum/minimum, despite the enzyme being active.
The presence of an enzyme even if the net flux is zero may be because it is part of a structural complex, e.g. a metabolon, and/or its presence helps to maintain specific crowding conditions required to convert an unfeasible pathway to a feasible one [42].

A reaction at thermodynamic equilibrium can modify the maximum/minimum values attainable for $v_k$ and $\Delta_G$, of other reactions and metabolite concentrations in the system. In this work, the equilibrium case (where all reactions are considered at equilibrium if their fluxes are zero) predicts a solution space that, in general, is a sub-region of that estimated for the block case.

The differences between both block and equilibrium states decrease, as more information becomes available about the system analysed. Thus, if the metabolite concentration is known only for some metabolites, but there is evidence of the presence of some enzymes (whose catalysed reaction $k$ will take the parameter value $d_k=1$, in place of the default $d_k=0$), then the combined information would help to reduce the solution space narrowing the range of fluxes and metabolite concentration of the unknown variables.

Acknowledgements

The financial support CONACYT-Mexico for the PhD of LAM is gratefully acknowledged.

References


Nomenclature

\( a \) Activity of a metabolite [dimensionless].

\( A \) Constant of the extended Debye-Hückel equation equivalent to 0.510651 [L\(^{-0.5}\) mol\(^{0.5}\)].

\( B \) Constant of the extended Debye-Hückel equation equivalent to 1.6 [L\(^{-0.5}\) mol\(^{0.5}\)].

\( c \) Subscript that indicates intracellular medium [dimensionless].

\( C \) Concentration of an intracellular metabolite [mol L\(^{-1}\)].

\( \mathbf{C} \) Concentration vector of all intracellular metabolites [mol L\(^{-1}\)].

\( C_{\text{max}} \) Maximum concentration of an intracellular metabolite [mol L\(^{-1}\)].

\( \mathbf{C}_{\text{max}} \) Maximum concentration vector of all intracellular metabolites [mol L\(^{-1}\)].

\( C_{\text{min}} \) Minimum concentration of an intracellular metabolite [mol L\(^{-1}\)].

\( \mathbf{C}_{\text{min}} \) Minimum concentration vector of all intracellular metabolites [mol L\(^{-1}\)].

\( C^{\text{std}} \) Standard concentration of an intracellular metabolite [mol L\(^{-1}\)].

\( d \) Parameter that indicates if a specific reaction is blocked (\( d=0 \)) or at equilibrium (\( d=1 \)) [dimensionless].

\( \mathbf{d} \) Parameter vector of all the reactions that indicates if a specific reaction is blocked (\( d=0 \)) or at equilibrium (\( d=1 \)) [dimensionless].

\( e \) Subscript that indicates extracellular medium [dimensionless].

\( E \) Error variable to consider the uncertainty in standard Gibbs free energy of the reaction estimated by the group contribution method [kcal mol\(^{-1}\)].

\( f \) Subscript that indicates the reaction used as objective function [dimensionless].

\( \text{func} \) Objective function to optimise in the NLP formulation.

\( F \) Faraday constant equivalent to 2.306 \( \times 10^{-2} \) [kcal mV\(^{-1}\) mol\(^{-1}\)].

\( i \) Subscript that indicates the pseudoisomer group of a metabolite \( i \) [dimensionless].

\( I \) Ionic strength [M].

\( j \) Subscript that indicates the protonated species \( j \) of a metabolite \( i \) [dimensionless].

\( k \) Subscript that indicates a specific reaction [dimensionless].
\( m \) Number of intracellular metabolites in the metabolic network analysed, the hydrogen ions \( H^+ \) is not included [dimensionless].

\( n \) Number of intracellular reactions in the metabolic network analysed [dimensionless].

\( N \) Number of reactions for the transport of metabolites outside/inside the cell [dimensionless].

\( N_H \) Number of hydrogen atoms in a metabolite [dimensionless].

\( pH \) Potential of hydrogen [dimensionless].

\( pH_{ref} \) Reference \( pH = 7 \) at standard biological conditions [dimensionless].

\( pK_a \) Acid dissociation constant [dimensionless].

\( R \) Gas constant equivalent to \( 1.9858 \times 10^{-3} \) [kcal mol\(^{-1}\) K\(^{-1}\)].

\( S \) \( m \times (n+N) \) stoichiometric matrix of the metabolic network [dimensionless].

\( SE \) Uncertainty associated to the standard Gibbs free energy of the reaction estimated by the group contribution method [kcal mol\(^{-1}\)].

\( tol \) Tolerance value of what is considered a non-zero Gibbs free energy [kcal mol\(^{-1}\)] or a non-zero flux [mmol L\(^{-1}\) g\(_{DW}\)^{-1}].

\( T \) Temperature of the medium [K].

\( v \) \( n \times 1 \) vector of the fluxes through the metabolic network [mmol L\(^{-1}\) g\(_{DW}\)^{-1}].

\( v_{max} \) Maximum flux through a reaction [mmol L\(^{-1}\) g\(_{DW}\)^{-1}].

\( v_{max} \) Maximum flux vector of all reactions [mmol L\(^{-1}\) g\(_{DW}\)^{-1}].

\( v_{min} \) Minimum flux through a reaction [mmol L\(^{-1}\) g\(_{DW}\)^{-1}].

\( v_{min} \) Minimum flux vector of all reactions [mmol L\(^{-1}\) g\(_{DW}\)^{-1}].

\( vol_{av} \) Available volume in the system [L].

\( vol_{tot} \) Total volume of the system [L].

\( z \) Electrical charge of the metabolite species \( j \) [dimensionless].

\( \Delta_j G_H \) Gibbs free energy of formation of the hydrogen [kcal mol\(^{-1}\)].

\( \Delta_j G^0 \) Standard Gibbs free energy of formation of a metabolite [kcal mol\(^{-1}\)].

\( \Delta_j G^{0i} \) Transformed Standard Gibbs free energy of formation of a metabolite [kcal mol\(^{-1}\)].

\( \Delta_j G^{0r} \) Transformed standard Gibbs free energy of a reaction formation of a metabolite [kcal mol\(^{-1}\)].

\( \Delta_j G' \) Transformed Gibbs free energy of a reaction [kcal mol\(^{-1}\)].
\( \Delta G' \) Transformed Gibbs free energy vector of all intracellular reactions [kcal mol\(^{-1}\)].

\( \Delta G'' \) Total transformed Gibbs free energy of a transport reaction of a metabolite [kcal mol\(^{-1}\)].

\( \Delta H^0 \) Standard enthalpy of formation of a metabolite [kcal mol\(^{-1}\)].

\( \Delta H^o \) Standard enthalpy of the reaction [kcal mol\(^{-1}\)].

\( \Delta \varphi \) Difference between the internal membrane potential and the external one [mV].

\( \gamma \) Activity coefficient of a metabolite [dimensionless].

\( \eta \) Stoichiometric coefficient of a metabolite in a reaction [dimensionless].
3.2.1. Supplementary Material for Publication 1

### METABOLITES

<table>
<thead>
<tr>
<th>No.</th>
<th>Abbreviation</th>
<th>Compound Name</th>
<th>Keeg compound</th>
<th>Formule</th>
<th>No. of H atoms</th>
<th>Charge at pH 7</th>
<th>ΔfG° (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13dpg</td>
<td>3-Phospho-D-glyceroyl phosphate</td>
<td>C00236</td>
<td>C3H8O10P2</td>
<td>8</td>
<td>-2</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>2pg</td>
<td>Glycerate 2-phosphate</td>
<td>C00631</td>
<td>C3H7O7P</td>
<td>7</td>
<td>-2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>3pg</td>
<td>3-Phospho-D-glycerate</td>
<td>C0197</td>
<td>C3H7O7P</td>
<td>7</td>
<td>-2</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>6pgc</td>
<td>6-Phospho-D-gluconate</td>
<td>C0345</td>
<td>C6H13O10P</td>
<td>13</td>
<td>-2</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>6pgl</td>
<td>6-phospho-D-glucono-1,5-lactone</td>
<td>C01236</td>
<td>C6H11O9P</td>
<td>11</td>
<td>-1</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>ac</td>
<td>Acetate</td>
<td>C0033</td>
<td>C2H4O2</td>
<td>4</td>
<td>-1</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>acald</td>
<td>Acetaldehyde</td>
<td>C0084</td>
<td>C2H4O</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>acoa</td>
<td>Acetyl-CoA</td>
<td>C00227</td>
<td>C2H5O5P</td>
<td>5</td>
<td>-1</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>adp</td>
<td>ADP</td>
<td>C00008</td>
<td>C10H15N5O10P2</td>
<td>15</td>
<td>-2</td>
<td>13</td>
</tr>
<tr>
<td>10</td>
<td>atp</td>
<td>ATP</td>
<td>C00002</td>
<td>C10H16N5O13P3</td>
<td>16</td>
<td>-3</td>
<td>13</td>
</tr>
<tr>
<td>11</td>
<td>co2</td>
<td>CO2</td>
<td>C00011</td>
<td>CO2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>coa</td>
<td>Coenzyme A</td>
<td>C00010</td>
<td>C2H3N6O7P</td>
<td>36</td>
<td>-3</td>
<td>33</td>
</tr>
<tr>
<td>13</td>
<td>dhap</td>
<td>Dihydroxyacetone phosphate</td>
<td>C0111</td>
<td>C3H7O6P</td>
<td>7</td>
<td>-1</td>
<td>6</td>
</tr>
<tr>
<td>14</td>
<td>e4p</td>
<td>Erythrose 4-phosphate</td>
<td>C0279</td>
<td>C4H9O7P</td>
<td>9</td>
<td>-1</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>etho</td>
<td>Ethanol</td>
<td>C00469</td>
<td>C2H6O</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>16</td>
<td>f6p</td>
<td>Fructose 6-phosphate</td>
<td>C0085</td>
<td>C6H13O9P</td>
<td>13</td>
<td>-1</td>
<td>12</td>
</tr>
<tr>
<td>17</td>
<td>fdp</td>
<td>Fructose 1,6-bisphosphate</td>
<td>C0354</td>
<td>C6H14O12P2</td>
<td>14</td>
<td>-2</td>
<td>12</td>
</tr>
<tr>
<td>18</td>
<td>for</td>
<td>Formate</td>
<td>C0058</td>
<td>C2H2O2</td>
<td>2</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>fum</td>
<td>Fumarate</td>
<td>C00122</td>
<td>C4H4O4</td>
<td>4</td>
<td>-2</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>g3p</td>
<td>Glycerate 3-phosphate</td>
<td>C00118</td>
<td>C3H7O6P</td>
<td>7</td>
<td>-1</td>
<td>6</td>
</tr>
<tr>
<td>21</td>
<td>g6p</td>
<td>Glucose 6-phosphate</td>
<td>C0092</td>
<td>C6H13O9P</td>
<td>13</td>
<td>-1</td>
<td>12</td>
</tr>
<tr>
<td>22</td>
<td>glyc</td>
<td>Glycerol</td>
<td>C00116</td>
<td>C3H8O3</td>
<td>8</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>23</td>
<td>glyc3p</td>
<td>sn-Glycerol 3-phosphate</td>
<td>C0093</td>
<td>C3H9O6P</td>
<td>9</td>
<td>-1</td>
<td>8</td>
</tr>
<tr>
<td>24</td>
<td>h</td>
<td>H+</td>
<td>C00080</td>
<td>H</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>h2o</td>
<td>H2O</td>
<td>C00256</td>
<td>C3H6O3</td>
<td>6</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>26</td>
<td>lac</td>
<td>Lactate</td>
<td>C00265</td>
<td>C3H6O3</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>mal</td>
<td>Malate</td>
<td>C00149</td>
<td>C4H6O5</td>
<td>6</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>28</td>
<td>nad</td>
<td>Nicotinamide adenine dinucleotide</td>
<td>C0003</td>
<td>C2H12N7O7P14F2</td>
<td>27</td>
<td>-1</td>
<td>26</td>
</tr>
<tr>
<td>29</td>
<td>nadh</td>
<td>Nicotinamide adenine dinucleotide - reduced</td>
<td>C0004</td>
<td>C2H12N7O7P14F2</td>
<td>27</td>
<td>-2</td>
<td>27</td>
</tr>
<tr>
<td>30</td>
<td>nadph</td>
<td>Nicotinamide adenine dinucleotide ph</td>
<td>C0006</td>
<td>C2H12N7O7P17P3</td>
<td>28</td>
<td>-2</td>
<td>26</td>
</tr>
<tr>
<td>31</td>
<td>oaa</td>
<td>Oxaloacetate</td>
<td>C0036</td>
<td>C4H4O5</td>
<td>4</td>
<td>-2</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>pep</td>
<td>Phosphoenolpyruvate</td>
<td>C0074</td>
<td>C3H6O6P</td>
<td>5</td>
<td>-2</td>
<td>3</td>
</tr>
<tr>
<td>33</td>
<td>pi</td>
<td>Phosphate (orthophosphate)</td>
<td>C0009</td>
<td>C3H6O4</td>
<td>3</td>
<td>-2</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>pyr</td>
<td>Pyruvate</td>
<td>C00222</td>
<td>C3H4O3</td>
<td>4</td>
<td>-1</td>
<td>3</td>
</tr>
<tr>
<td>35</td>
<td>r5p</td>
<td>Ribose 5-phosphate</td>
<td>C00127</td>
<td>C5H11O8P</td>
<td>11</td>
<td>-1</td>
<td>10</td>
</tr>
<tr>
<td>36</td>
<td>rul5p</td>
<td>Ribulose 5-phosphate</td>
<td>C0199</td>
<td>C5H11O8P</td>
<td>11</td>
<td>-1</td>
<td>10</td>
</tr>
<tr>
<td>37</td>
<td>succ</td>
<td>Succinate</td>
<td>C0042</td>
<td>C4H6O4</td>
<td>6</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>38</td>
<td>xul5p</td>
<td>Xylulose 5-phosphate</td>
<td>C0231</td>
<td>C5H11O8P</td>
<td>11</td>
<td>-1</td>
<td>10</td>
</tr>
</tbody>
</table>

### GLYCEROLYSIS

**R1**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Reaction</th>
<th>ΔG°/kcal/mol</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGI</td>
<td>glucose-6-phosphate isomerase</td>
<td>-5.13</td>
<td>[1]</td>
</tr>
<tr>
<td>PKF</td>
<td>6-phosphofructokinase</td>
<td>-2.11</td>
<td>[1]</td>
</tr>
<tr>
<td>FBP</td>
<td>fructose-1,6-bisphosphatase</td>
<td>-3.11</td>
<td>[1]</td>
</tr>
<tr>
<td>FBA</td>
<td>fructose-bisphosphatase aldolase</td>
<td>-1.12</td>
<td>[1]</td>
</tr>
<tr>
<td>TPI</td>
<td>triose-phosphate isomerase</td>
<td>-5.31</td>
<td>[1]</td>
</tr>
<tr>
<td>GAPD</td>
<td>3-phosphoglycerate dehydrogenase</td>
<td>-2.12</td>
<td>[1]</td>
</tr>
<tr>
<td>PKG</td>
<td>3-phosphoglycerate kinase</td>
<td>-2.73</td>
<td>[1]</td>
</tr>
<tr>
<td>R1518</td>
<td>phosphoglycerate mutase</td>
<td>-5.40</td>
<td>[1]</td>
</tr>
<tr>
<td>ENO</td>
<td>enolase</td>
<td>4.21</td>
<td>[1]</td>
</tr>
<tr>
<td>PYK</td>
<td>pyruvate kinase</td>
<td>2.71</td>
<td>[1]</td>
</tr>
</tbody>
</table>

### PYRUVATE METABOLISM

**R1**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Reaction</th>
<th>ΔG°/kcal/mol</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDH</td>
<td>D-lactate dehydrogenase</td>
<td>-1.12</td>
<td>[1]</td>
</tr>
<tr>
<td>PFL</td>
<td>pyruvate formate lyase</td>
<td>2.15</td>
<td>[1]</td>
</tr>
<tr>
<td>FHL</td>
<td>formate dehydrogenase</td>
<td>1.22</td>
<td>[1]</td>
</tr>
<tr>
<td>PSTV</td>
<td>acetyl-CoA synthetase</td>
<td>2.33</td>
<td>[1]</td>
</tr>
<tr>
<td>AKH</td>
<td>acetate kinase</td>
<td>1.71</td>
<td>[1]</td>
</tr>
<tr>
<td>ACY</td>
<td>acetyl-CoA synthetase</td>
<td>3.16</td>
<td>[1]</td>
</tr>
<tr>
<td>PCK</td>
<td>pyruvate kinase</td>
<td>1.23</td>
<td>[1]</td>
</tr>
<tr>
<td>ALD</td>
<td>acetate dehydrogenase</td>
<td>1.11</td>
<td>[1]</td>
</tr>
<tr>
<td>ALDO2A</td>
<td>alcohol dehydrogenase</td>
<td>1.11</td>
<td>[1]</td>
</tr>
</tbody>
</table>

### TCA CYCLE

**R1**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Reaction</th>
<th>ΔG°/kcal/mol</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP0X</td>
<td>phosphoenolpyruvate carboxylase</td>
<td>-4.14</td>
<td>[1]</td>
</tr>
<tr>
<td>MDH</td>
<td>malate dehydrogenase</td>
<td>-1.11</td>
<td>[1]</td>
</tr>
<tr>
<td>FUM</td>
<td>fumarase</td>
<td>4.12</td>
<td>[1]</td>
</tr>
<tr>
<td>RFD</td>
<td>fumarase dehydrogenase</td>
<td>1.90</td>
<td>[1]</td>
</tr>
</tbody>
</table>

### PENTOSE PHOSPHATE PATHWAY

**R2**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Reaction</th>
<th>ΔG°/kcal/mol</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>G6PDH2</td>
<td>glucose-6-phosphate dehydrogenase</td>
<td>-1.14</td>
<td>[1]</td>
</tr>
<tr>
<td>F6P</td>
<td>fructose-6-phosphate isomerase</td>
<td>3.16</td>
<td>[1]</td>
</tr>
<tr>
<td>RPI</td>
<td>ribose-5-phosphate isomerase</td>
<td>5.10</td>
<td>[1]</td>
</tr>
<tr>
<td>BPE</td>
<td>ribulose-5-phosphate 3-epimerase</td>
<td>5.13</td>
<td>[1]</td>
</tr>
<tr>
<td>TAT</td>
<td>transaldolase</td>
<td>2.22</td>
<td>[1]</td>
</tr>
<tr>
<td>TKT</td>
<td>transketolase</td>
<td>2.22</td>
<td>[1]</td>
</tr>
</tbody>
</table>

### GLYCEROL METABOLISM

**R3**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Reaction</th>
<th>ΔG°/kcal/mol</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>G3PD</td>
<td>glycerol kinase</td>
<td>-0.29</td>
<td>[1]</td>
</tr>
<tr>
<td>G6PD2</td>
<td>glucose-6-phosphate dehydrogenase (NADP)</td>
<td>-5.48</td>
<td>[1]</td>
</tr>
</tbody>
</table>

### NECTINATE AND NECTINOSEDAMIDE METABOLISM

**R4**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Reaction</th>
<th>ΔG°/kcal/mol</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>THD2</td>
<td>nectinolate dehydrogenase</td>
<td>7.89</td>
<td>[1]</td>
</tr>
</tbody>
</table>

### EXCHANGE

**R5**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Reaction</th>
<th>ΔG°/kcal/mol</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENK</td>
<td>pyruvate exchange</td>
<td>-0.29</td>
<td>[1]</td>
</tr>
<tr>
<td>ENK</td>
<td>pyruvate exchange</td>
<td>-0.29</td>
<td>[1]</td>
</tr>
<tr>
<td>ENK</td>
<td>pyruvate exchange</td>
<td>-0.29</td>
<td>[1]</td>
</tr>
<tr>
<td>ENK</td>
<td>pyruvate exchange</td>
<td>-0.29</td>
<td>[1]</td>
</tr>
<tr>
<td>ENK</td>
<td>pyruvate exchange</td>
<td>-0.29</td>
<td>[1]</td>
</tr>
</tbody>
</table>

### CONSUMPTION OF METABOLITES IN OTHER REACTIONS NOT CONSIDERED

**R6**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Reaction</th>
<th>ΔG°/kcal/mol</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALD</td>
<td>nectinolate dehydrogenase</td>
<td>0.00</td>
<td>[1]</td>
</tr>
<tr>
<td>ALD</td>
<td>nectinolate dehydrogenase</td>
<td>0.00</td>
<td>[1]</td>
</tr>
</tbody>
</table>

### BIOMASS

**R7**

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Reaction</th>
<th>ΔG°/kcal/mol</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.C.</td>
<td>biomass</td>
<td>-45.93</td>
<td>[1]</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>$\Delta fH^\circ$ (pi-2) (kJ/mol)</th>
<th>$\Delta fH^\circ$ (h2o) (kJ/mol)</th>
<th>Ref.</th>
<th>Method used</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>13dpg</td>
<td></td>
<td>-1299</td>
<td>5</td>
<td>Gani</td>
<td>$\Delta fH^\circ$ (glycerate) = -876.317 Gani $\Delta fH^\circ$ (13dpg) = -2902.657.</td>
</tr>
<tr>
<td>2pg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3pg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6pgc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6pgl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>accoa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>actp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dhap</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e4p</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f6p</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fdp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g3p</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glyc3p</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pep</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ru5p</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s7p</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>xu5p</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Compound</th>
<th>No. H</th>
<th>Charge</th>
<th>No. H / Specie</th>
<th>pKa (I=0)</th>
<th>ΔfH° (kJ/mol)</th>
<th>ΔfH° (kcal/mol)</th>
<th>ΔfG° (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13dpq-4</td>
<td>8</td>
<td>-4</td>
<td>4</td>
<td>6.64</td>
<td>-2922.657</td>
<td>-693.751673</td>
<td>-586.8</td>
</tr>
<tr>
<td>13dpq-3</td>
<td>-3</td>
<td>5</td>
<td>6</td>
<td>6.64</td>
<td>-2922.657</td>
<td>-693.751673</td>
<td>-586.8</td>
</tr>
<tr>
<td>13dpq-2</td>
<td>-2</td>
<td>6</td>
<td>6</td>
<td>6.64</td>
<td>-2922.657</td>
<td>-693.751673</td>
<td>-586.8</td>
</tr>
<tr>
<td>2dpq-3</td>
<td>7</td>
<td>-3</td>
<td>4</td>
<td>6</td>
<td>-1889.487</td>
<td>-453.598235</td>
<td>-371.9</td>
</tr>
<tr>
<td>2dpq-2</td>
<td>-2</td>
<td>5</td>
<td>6</td>
<td>6.64</td>
<td>-2922.657</td>
<td>-693.751673</td>
<td>-586.8</td>
</tr>
<tr>
<td>3dpq-3</td>
<td>7</td>
<td>-3</td>
<td>4</td>
<td>6.64</td>
<td>-1889.487</td>
<td>-453.598235</td>
<td>-371.9</td>
</tr>
<tr>
<td>3dpq-2</td>
<td>-2</td>
<td>5</td>
<td>6</td>
<td>6.64</td>
<td>-2922.657</td>
<td>-693.751673</td>
<td>-586.8</td>
</tr>
<tr>
<td>6pgc-3</td>
<td>13</td>
<td>-3</td>
<td>10</td>
<td>6.61</td>
<td>-2518.527</td>
<td>-601.9424</td>
<td>-481.16</td>
</tr>
<tr>
<td>6pgc-2</td>
<td>-2</td>
<td>11</td>
<td>9</td>
<td>6.61</td>
<td>-2518.527</td>
<td>-601.9424</td>
<td>-481.16</td>
</tr>
<tr>
<td>ac-4</td>
<td>4</td>
<td>-1</td>
<td>3</td>
<td>4.54</td>
<td>-486.01</td>
<td>-116.159178</td>
<td>-88.29</td>
</tr>
<tr>
<td>ac-3</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>4.54</td>
<td>-485.76</td>
<td>-116.099426</td>
<td>-88.29</td>
</tr>
<tr>
<td>6pgl-2</td>
<td>11</td>
<td>-2</td>
<td>9</td>
<td>6.33</td>
<td>-2225.133</td>
<td>-531.819551</td>
<td>-428.52</td>
</tr>
<tr>
<td>6pgl-1</td>
<td>-1</td>
<td>10</td>
<td>9</td>
<td>6.33</td>
<td>-2225.133</td>
<td>-531.819551</td>
<td>-428.52</td>
</tr>
<tr>
<td>actp-2</td>
<td>5</td>
<td>-2</td>
<td>1</td>
<td>6.28</td>
<td>-1784.76</td>
<td>-426.587878</td>
<td>-301.65</td>
</tr>
<tr>
<td>actp-1</td>
<td>-1</td>
<td>4</td>
<td>4</td>
<td>4.54</td>
<td>-1784.76</td>
<td>-426.587878</td>
<td>-301.65</td>
</tr>
<tr>
<td>adp-3</td>
<td>15</td>
<td>-3</td>
<td>12</td>
<td>7.1</td>
<td>-2620.94</td>
<td>-626.419694</td>
<td>-465.85</td>
</tr>
<tr>
<td>adp-2</td>
<td>-2</td>
<td>13</td>
<td>5.12</td>
<td>7.1</td>
<td>-2620.94</td>
<td>-626.419694</td>
<td>-465.85</td>
</tr>
<tr>
<td>adp-1</td>
<td>-1</td>
<td>14</td>
<td>5.12</td>
<td>7.1</td>
<td>-2620.94</td>
<td>-626.419694</td>
<td>-465.85</td>
</tr>
<tr>
<td>co2(tot)</td>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>6.05</td>
<td>-691.99</td>
<td>-165.389576</td>
<td>-140.26</td>
</tr>
<tr>
<td>co2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6.05</td>
<td>-691.99</td>
<td>-165.389576</td>
<td>-140.26</td>
</tr>
<tr>
<td>coa-4</td>
<td>36</td>
<td>-4</td>
<td>32</td>
<td>5.6</td>
<td>0</td>
<td>0</td>
<td>-751.994</td>
</tr>
<tr>
<td>coa-3</td>
<td>-3</td>
<td>33</td>
<td>5</td>
<td>5.6</td>
<td>0</td>
<td>0</td>
<td>-751.994</td>
</tr>
<tr>
<td>coa-2</td>
<td>-2</td>
<td>34</td>
<td>5</td>
<td>5.6</td>
<td>0</td>
<td>0</td>
<td>-751.994</td>
</tr>
<tr>
<td>No. H</td>
<td>charge</td>
<td>No. H specie</td>
<td>pKa</td>
<td>ΔfH° (kJ/mol)</td>
<td>ΔfH° (kcal/mol)</td>
<td>ΔfG° (kcal/mol)</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>-------------</td>
<td>-----</td>
<td>---------------</td>
<td>----------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>dhap-2</td>
<td>7</td>
<td>-2</td>
<td>5</td>
<td>6.22</td>
<td>-1620.93</td>
<td>-389.80</td>
<td></td>
</tr>
<tr>
<td>dhap-1</td>
<td>-1</td>
<td>6</td>
<td></td>
<td>-1620.93</td>
<td>-389.80</td>
<td>-320.66</td>
<td></td>
</tr>
<tr>
<td>ep-2</td>
<td>9</td>
<td>-2</td>
<td>7</td>
<td>6.51</td>
<td>-1881.045</td>
<td>-444.80</td>
<td></td>
</tr>
<tr>
<td>ep-1</td>
<td>-1</td>
<td>8</td>
<td></td>
<td>-1881.045</td>
<td>-444.80</td>
<td>-355.62</td>
<td></td>
</tr>
<tr>
<td>vdh</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td></td>
<td>-288.3</td>
<td>-68.91</td>
<td></td>
</tr>
<tr>
<td>f-2</td>
<td>13</td>
<td>-2</td>
<td>11</td>
<td>6.24</td>
<td>2272.55</td>
<td>543.67</td>
<td></td>
</tr>
<tr>
<td>f-1</td>
<td>-1</td>
<td>12</td>
<td></td>
<td>2272.55</td>
<td>543.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f2</td>
<td>14</td>
<td>-4</td>
<td>10</td>
<td>6.52</td>
<td>3295.57</td>
<td>785.30</td>
<td></td>
</tr>
<tr>
<td>f2-3</td>
<td>-3</td>
<td>11</td>
<td>5.92</td>
<td>3295.57</td>
<td>785.30</td>
<td>644.17</td>
<td></td>
</tr>
<tr>
<td>for-1</td>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>4.27</td>
<td>-425.55</td>
<td>103.70</td>
<td></td>
</tr>
<tr>
<td>for</td>
<td>0</td>
<td>2</td>
<td></td>
<td>-425.55</td>
<td>103.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fum-3</td>
<td>6</td>
<td>-2</td>
<td>7</td>
<td>6.41</td>
<td>2272.55</td>
<td>543.67</td>
<td></td>
</tr>
<tr>
<td>fum-1</td>
<td>-1</td>
<td>1</td>
<td></td>
<td>2272.55</td>
<td>543.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>g(3p)</td>
<td>7</td>
<td>-2</td>
<td>5</td>
<td>6.42</td>
<td>1925.48</td>
<td>460.15</td>
<td></td>
</tr>
<tr>
<td>g(3p)-1</td>
<td>-1</td>
<td>6</td>
<td></td>
<td>1925.48</td>
<td>460.15</td>
<td>-310.2</td>
<td></td>
</tr>
<tr>
<td>g(3p)</td>
<td>13</td>
<td>-2</td>
<td>11</td>
<td>6.25</td>
<td>2272.55</td>
<td>543.67</td>
<td></td>
</tr>
<tr>
<td>g(3p)-1</td>
<td>-1</td>
<td>12</td>
<td></td>
<td>2272.55</td>
<td>543.67</td>
<td>-430.78</td>
<td></td>
</tr>
<tr>
<td>g3p</td>
<td>8</td>
<td>0</td>
<td>8</td>
<td></td>
<td>776.55</td>
<td>185.10</td>
<td></td>
</tr>
<tr>
<td>g3p</td>
<td>9</td>
<td>-2</td>
<td>7</td>
<td>6.53</td>
<td>-1689.72</td>
<td>-403.85</td>
<td></td>
</tr>
<tr>
<td>g3p-1</td>
<td>-1</td>
<td>8</td>
<td></td>
<td>-1689.72</td>
<td>-403.85</td>
<td>-328.68</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-0.53</td>
<td></td>
</tr>
<tr>
<td>h20</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td></td>
<td>-233.81</td>
<td>56.68</td>
<td></td>
</tr>
<tr>
<td>lac</td>
<td>6</td>
<td>-1</td>
<td>5</td>
<td>0</td>
<td>-560.94</td>
<td>133.17</td>
<td></td>
</tr>
<tr>
<td>lac</td>
<td>7</td>
<td>-1</td>
<td>6</td>
<td>0</td>
<td>-560.94</td>
<td>133.17</td>
<td></td>
</tr>
<tr>
<td>No. H</td>
<td>charge</td>
<td>No. H specie</td>
<td>pKa</td>
<td>ΔfH° (kJ/mol)</td>
<td>ΔfH° (kcal/mol)</td>
<td>ΔfG° (kcal/mol)</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>--------------</td>
<td>-----</td>
<td>---------------</td>
<td>---------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>mal-2</td>
<td>6</td>
<td>-2</td>
<td>4</td>
<td>5.13</td>
<td>-1113.681</td>
<td>-266.17647</td>
<td></td>
</tr>
<tr>
<td>mal-1</td>
<td>-1</td>
<td>5</td>
<td></td>
<td></td>
<td>-1113.681</td>
<td>-266.17647</td>
<td></td>
</tr>
<tr>
<td>nad</td>
<td>27</td>
<td>-1</td>
<td>26</td>
<td>5.13</td>
<td></td>
<td>-529.53</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nadh</td>
<td>29</td>
<td>-2</td>
<td>27</td>
<td>5.13</td>
<td></td>
<td>-730.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nadp</td>
<td>28</td>
<td>-3</td>
<td>25</td>
<td>5.42</td>
<td></td>
<td>-742.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pep</td>
<td>3</td>
<td>-3</td>
<td>2</td>
<td>6.02</td>
<td>-1597.055</td>
<td>-381.705306</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pi</td>
<td>2</td>
<td>-2</td>
<td>1</td>
<td>6.95</td>
<td>-1299</td>
<td>-311.328872</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prr</td>
<td>4</td>
<td>-1</td>
<td>3</td>
<td>-12</td>
<td>-596.22</td>
<td>-152.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rfp</td>
<td>11</td>
<td>-2</td>
<td>9</td>
<td>6.25</td>
<td>-2041.48</td>
<td>-487.92543</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rufp</td>
<td>11</td>
<td>-2</td>
<td>9</td>
<td>6.5</td>
<td>-2041.11</td>
<td>-486.66109</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>succ</td>
<td>6</td>
<td>-2</td>
<td>4</td>
<td>5.89</td>
<td>-247.051</td>
<td>-61.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subp</td>
<td>11</td>
<td>-2</td>
<td>9</td>
<td>6.5</td>
<td>-2042.82</td>
<td>-488.245698</td>
<td></td>
</tr>
</tbody>
</table>
The values of ΔfH° were either estimated (see Calculation of ΔfH°) or obtained from:

The values of ΔfG° for the predominant species at pH=7 298.15K and I=0 (see compounds) are obtained from:
Supporting References

Chapter 4

Crowding Conditions in the Thermodynamic Analysis of a Metabolic Network

4.1. Preamble

The study and simulation of the cellular metabolism bring numerous advantages and applications, not only in the basic science research deepening the understanding of the interactions among the cellular components, but also in the bioengineering improving the yield of processes where the microorganisms or any other biological compound are involved.

The modelling of the cell behaviour allows us to be in a better position for planning strategies for the manipulation of the organism. For this, the first step is the understanding of how the fluxes are distributed within the cell and the factors that affect them. Among the most popular methods for the estimation of the flux distribution are the stoichiometric-based models.

The thermodynamic analysis has been incorporated to the stoichiometric methods in order to restrict the prediction to only fluxes that satisfy the second law of thermodynamics. As shown in Chapter 3, the intracellular environmental conditions ionic strength, temperature and $pH$ alter the Gibbs free energy of the reaction, which in turn affects the estimation of the range of fluxes.

Besides the intracellular conditions mentioned above, the cell is characterised by being a crowded medium whose components can occupy up to 40% of the total volume. These crowding conditions can change the thermodynamic activities of the metabolites and shift the thermodynamic equilibrium of the reactions.
Although the crowding conditions have been considered in the stoichiometric model FBAwMC (Beg et al., 2007), this method focuses on the space restriction that can be occupied by the enzymes within the cell, but not in the thermodynamic analysis.

Meanwhile, the crowding dependency of the thermodynamic feasibility of the reactions has not been explicitly taken into account by the thermodynamically constrained stoichiometric models available in the literature, which is equivalent to assume point-like molecules. Chapter 2 presents a review of the stoichiometric methods and other models used in the metabolic simulations.

The aim of this chapter is to show the effect of the crowding conditions on the feasibility of the metabolic pathways and also their influence on the range of fluxes predicted by the thermodynamically constrained stoichiometric model.

In this study, the methodology formulated in Publication 1 is taken and adapted to incorporate the crowding conditions into the thermodynamic constraints. For this the Scaled Particle Theory (SPT) is used to estimate the activity coefficient of the metabolites, which represents the non-ideal behaviour of the solution arising from the steric effects of the molecules, i.e. the prevention of the molecules’ overlapping.

The SPT has been used in the past for the estimation of the thermodynamic activity of macromolecules (Minton, 1981), the osmotic pressure (Minton, 1995), in kinetic models (Smolen and Keizer, 1990), among others. The derivation of SPT equation is shown in Appendix.

The main assumptions made in this work for the evaluation of the crowding effect can be summarised as: 1) the reactions are carried out in a well-mixed medium, 2) a reaction with zero flux is considered blocked, i.e. the enzyme is absent or inactive, 3) the volume of the molecules is proportional to their molecular weight, and 4) the molecules are considered hard-spheres interacting only by the excluded volume interactions.

The lactic acid pathway and, as in the Chapter 3, the central carbon metabolic network
of *A. succinogenes* for the production of succinate from glycerol are used as case studies to determine the effect of intracellular crowding conditions on the thermodynamic analysis of metabolic pathways. The results and discussion of this study is presented in the following paper.
4.2. Publication 2

The influence of crowding conditions on the thermodynamic feasibility of metabolic pathways

L. Angeles-Martinez and C. Theodoropoulos*

School of Chemical Engineering and Analytical Science, University of Manchester, Manchester, M13 9PL, UK

ABSTRACT Intracellular reactions are carried out in a crowded media where the macromolecules occupy about 40% of the total volume. This decrease in the available volume towards the centre of mass of reactants may affect their activity. Scaled Particle Theory (SPT) is used for the estimation of the activity coefficients of the metabolites, and thereby for the assessment of the impact of the presence of background molecules, on the estimation of the Gibbs free energy change ($\Delta G'$) of the reactions. The lactic acid pathway and the central carbon metabolism of Actinobacillus succinogenes for the production of succinic acid from glycerol have been used as illustrative case studies. Results suggest the importance of maintaining intracellular crowded regions to favour the feasibility of a pathway that in other circumstances would be infeasible. Moreover, the crowding conditions may change the directionality of reactions and can modify the feasible range of fluxes estimated for a metabolic system compared with those obtained at standard biological conditions.

KEYWORDS Thermodynamic constraints, Scaled Particle Theory, Flux analysis, Actinobacillus succinogenes, succinate production

INTRODUCTION

Stoichiometric-based models are valuable tools for the estimation of the likely range of fluxes in a metabolic network under certain environmental conditions (1–4).

The use of thermodynamic constraints allows the calculation of fluxes in the direction of the Gibbs free energy drop (5). Thermodynamic analysis indicates the feasibility, directionality and reversibility of the reactions involved in the network by the estimation of the corresponding change in Gibbs free energy $\Delta G'$, which depends on intracellular conditions such as pH, ionic strength ($I$), temperature ($T$), activity of the metabolites, and also the intracellular crowding conditions (6).

The cytoplasm contains several species including macromolecules (e.g. the enzymes) and solutes whose concentrations are not considered high but all together may occupy about 40% of the cellular space (7). Due to the impenetrability of the molecules, this reduces the available volume for the motion and reaction of the species (excluded volume effect). Hence, the macromolecular crowding may affect the equilibrium and rate of the reactions involving changes in the available volume (7).

* Correspondence: k.theodoropoulos@manchester.ac.uk
The thermodynamic activity of a metabolite $a$ (equivalent to the number of moles per available volume) is related to its concentration $C$ (moles per total volume) through the activity coefficient $\gamma$, $a = \gamma C$. Here $\gamma$ is the ratio between the total volume and the available volume.

The activity coefficient, $\gamma$, encompasses the deviations from the ideal behaviour of a mixture caused by all kinds of interactions among the molecules. Under crowding conditions such interactions are mainly due to steric repulsion.

Assuming that all the species in a solution are volumeless particles, then the available volume is equal to the total volume of the solution, and therefore $\gamma = 1$. However, in a more realistic scenario where the available volume is reduced due to the presence of inert molecules (called background molecules) the activity of the reactants is increased, which could shift an infeasible reaction ($\Delta_G > 0$) to a feasible one ($\Delta_G < 0$) even at low concentrations of reactants.

Several attempts have been made in order to integrate the thermodynamic information to Flux Balance Analysis (FBA) or to other stoichiometric methods (8–16), but none of them has developed an explicit relationship between the crowding conditions and the thermodynamic analysis.

Nevertheless, the crowding effect has been incorporated to FBA (17) by restricting the maximum number of macromolecules or the concentration of cytoplasmic enzymes. Through this restriction and the use of kinetic parameters, the maximum metabolic flux attainable for a reaction is estimated. However, the influence of crowding conditions on the thermodynamic feasibility of a pathway and its corresponding flux distribution has not been analysed.

Scaled Particle Theory (SPT) (18,19) provides the theoretical framework to calculate the activity coefficients of species in a hard-sphere mixture as a function of the concentration and the radii of the metabolites. SPT has been widely used to examine the influence of macromolecular crowding on the stability (20) and thermodynamic activity (21) of globular proteins, solvation (22), and the estimation of the osmotic pressure (23).

The aim of this paper is to illustrate the qualitative effect of the macromolecular crowding on thermodynamic analysis of metabolic pathways, and its further impact on the prediction of ranges of fluxes. To achieve this, we used SPT for the estimation of the activity coefficients and we subsequently coupled the crowding effect with thermodynamically constrained Flux Variability Analysis, formulated as a nonlinear optimisation problem.

Our illustrative case studies to demonstrate the excluded volume effects on the prediction of Gibbs free energies and ranges of fluxes of a metabolic system are the lactic acid metabolic pathway, and the central metabolism of Actinobacillus succinogenes for the production of succinic acid from glycerol, which is an industrially-relevant biochemical system with important implications in biorefinery engineering for the valorisation of the by-product crude glycerol of the biodiesel manufacture (24–26).
METHODS

The simultaneous solution of the mass conservation equation and the thermodynamic constrains is fundamental for the calculation of metabolic flux distributions whose components do not violate the second law of the Thermodynamics.

Mass balance constraints. As all stoichiometric models, this methodology is based on the assumption that the intracellular reactions are at steady state, and they are carried out in a homogeneous medium hence

\[ S \cdot \mathbf{v} = 0 \]  

where the \( m \times (n+N) \) stoichiometric matrix \( S \) contains the information of \( m \) intracellular metabolites (the hydrogen ion \( \text{H}^+ \) is not included in \( S \) since it not considered a metabolite), and the \( n+N \) reactions involved in the metabolic network (\( n \) intracellular reactions and \( N \) reactions for the transport of metabolites outside/inside the cell), while \( \mathbf{v} \) is the \( (n+N) \times 1 \) flux vector.

Incorporation of the Crowding Conditions on Thermodynamic Constraints. The feasibility of a reaction in a certain direction is assessed by its \( \Delta G' \) value. For a metabolic reaction \( \Delta_i G' \) can be split into the intracellular biochemical reaction contribution (\( \Delta_{i,i} G' \)) and the (coupled) transport reaction of some metabolites across the cell membrane (\( \Delta_{i,m} G' \))

\[ \Delta_i G' = \Delta_{i,i} G' + \Delta_{i,m} G' \]  

where

\[ \Delta_{i,i} G' = \sum_{i=1}^{\text{Products}} \eta_i [\Delta_{i,i} G_i^{\text{r}} + RT \ln(a_i)] - \sum_{i=1}^{\text{Reactants}} \eta_i [\Delta_{i,i} G_i^{\text{p}} + RT \ln(a_i)] \]  

\[ \Delta_{i,m} G' = \eta_{H^+} RT \ln(10^{\text{pH}_i - \text{pH}_m}) + \eta_{H^+} F \Delta \phi \]

\( \eta_i \) is the stoichiometric coefficient of the metabolite \( i \) involved in the intracellular reactions, and \( a_i \) its thermodynamic activity. \( \Delta_{i,i} G_i^{\text{r}} \) indicates the transformed Gibbs free energy of formation of the metabolite \( i \) (at a defined constant \( \text{pH} \) value, and ionic strength \( I \)). In this paper each metabolite is represented by its pseudoisomer group.

The pseudoisomer group \( i \) consists of all the protonated species \( j \) formed by the dissociation of the metabolite \( i \) in aqueous solutions (27). The equations used to estimate \( \Delta_{i,i} G_i^{\text{r}} \) are summarised in the Appendix.

In this paper, we only consider the transport of hydrogen ion \( \text{H}^+ \) across the membrane coupled to intracellular reactions. An example of this kind of reactions is the production of succinic acid (succ) from fumarate (fum) by the fumarate reductase (28):

\[ \text{fum} + \text{NAD} + \text{H}^+ \leftrightarrow \text{succ} + \text{NADH} \]  

\[ 2\text{H}^+ \leftrightarrow 2\text{H}_2 \]  

(5)
Eq. 4 only works for the $H^+$ transport from outside the cell into the cell, but a more general transport equation can be found in Jol et al. (29). Here, the same symbol $\eta$ is used to represent the stoichiometric coefficient or number of $H^+$ crossing the membrane, and $F$ is the Faraday constant (equivalent to $2.306 \times 10^2$ kcal mV$^{-1}$ mol$^{-1}$).

Although the hydrogen ion $H^+$ is not considered in biochemical reactions (27) (neither in $\Delta G_i$), this is not the case for $\Delta G^{\text{ion}}_i$, which is a function of the $H^+$ concentration difference (or $pH$) between both sides of the cell membrane (first terms of the right hand side of Eq. 4).

The second term of the right hand side of Eq. 4 is the electrical potential difference of the membrane. The difference between the internal potential of the membrane and the external one ($\Delta \varphi$) is calculated as (5):

$$\Delta \varphi = 33.33(pH_e - pH_i) - 143.33$$

where $pH_e$ and $pH_i$ are the intracellular and extracellular/medium $pH$ respectively.

As shown in Eq. 2–4, $\Delta G_i$ is a function of the metabolites’ activity $a_i$, defined as

$$a_i = \gamma_i \frac{C_i}{C_i^\text{st}}$$

where $C_i$ is the molar concentration of the metabolite $i$ (i.e. moles per total volume), and $C_i^\text{st}$ is its standard concentration equal to 1 M. For the solvent, water in this case $C^\text{water}$ is set equal to the molar concentration of pure water, i.e. 55.34 M.

The activity coefficient $\gamma_i$ represents the nonideal deviations raised by molecular interactions. For example when an electrolyte is dissociated into its component ions, $\gamma_i$ can be estimated using the Debye-Hückel equation (in fact this expression is used to calculate $\Delta G^0_i$ at different $I$, from its standard Gibbs free energy of formation $\Delta G^0_i$ at the biological standard conditions $pH = 7$, $I = 0$ M and $T = 298.15$ K).

Assuming hard-core molecules, and that no molecular interactions other than steric repulsion, $\gamma_i$ can be estimated using STP.

Scaled Particle Theory (SPT)

The mathematical derivation of SPT (18,19) is based on the calculation of the probability to find an empty cavity in an arbitrary location of the mixture or solution analysed, where a molecule of certain size and shape can fit without overlapping with other molecules.

SPT allows the analysis of mixtures of molecules of different size, but with the same shape. In this paper we assume that all the metabolites are hard spheres whose radii are
proportional to their molecular weight. Therefore, the activity coefficient of the metabolite \( i \) (\( \gamma_i \)) can be expressed as

\[
\ln \gamma_i = -\ln(1-S_i) + \frac{6S_i}{1-S_3} r_i + \left[ \frac{12S_i}{1-S_3} + \frac{18S_i^2}{(1-S_3)^2} \right] r_i^2 + \left[ \frac{8S_i}{1-S_3} + \frac{24S_i S_2}{(1-S_3)^2} + \frac{24S_i^3}{(1-S_3)^3} \right] r_i^3
\]  

(8)

where \( S_i \) (0 ≤ \( x \) ≤ 3) is given by

\[
S_i = \frac{\pi}{6} \sum_{i=1}^n \rho_i (2r_i)^i
\]  

(9)

\[
\rho_i = \frac{C_i N_A}{1000}
\]  

(10)

\[
r_i = \frac{M_i v_i}{4\pi} \frac{r}{3 N_A}
\]  

(11)

\( \rho_i \) represents the density number (molecules per cm\(^3\)), \( r_i \) the radius (cm\(^3\)), \( v_i \) the specific volume (cm\(^3\) g\(^{-1}\)), \( C_i \) the concentration (mol L\(^{-1}\)), and \( M_i \) is the molecular weight (g mol\(^{-1}\)) of metabolite \( i \), while \( N_A \) is Avogadro’s number.

As can be seen in Eq. 8-11, \( \gamma_i \) depends on the density number (or its equivalent concentration) of all species (metabolites and macromolecules) present in the cell or the system analysed, regardless if they are involved or not in a reaction. Hence, the impact of inert molecules (or crowders) can be easily simulated.

To do that, additional information related to the crowder molecules has to be provided such as the concentration (or its equivalent the density number), and radii (or the specific volume and molecular weight).

In this paper, the coupling of both constraints (mass balance and thermodynamics) is formulated as a non-linear optimisation (non-linear programming –NLP) problem. This methodology (expressed in Eq. 12-19 below) allows the prediction of thermodynamically feasible flux distributions, the range of intracellular metabolites’ concentration, and the Gibbs free energy change (\( \Delta G' \)) of the reactions involved in a metabolic network under crowding conditions.

\[
\text{maximise/minimise } \quad func
\]

\[
s.t.
\]

\[
S \cdot v = 0
\]  

(12)

\[
v_{\min,k} \leq v_k \leq v_{\max,k}
\]  

(13)

\[
C_i = 0 \quad \text{if} \quad v_k = 0 \quad \text{and} \quad \Delta_i G_k ' \neq 0
\]  

(14)

\[
C_{\min,i} \leq C_i \leq C_{\max,i} \quad \text{otherwise}
\]  

(15)

\[
v_i \cdot \Delta_i G_k ' \leq 0
\]  

(16)

\[
abs(v_k) + \frac{\Delta_i G_k '}{v_k} \leq 0
\]  

(17)
\[
\Delta G_k' = \Delta G_{f_{\text{var},k}} + \Delta G_{v_{\text{var},k}}
\]

\[
C_{\text{crowder}}, M_{\text{crowder}}, \nu_{\text{crowder}}
\]

In the above scheme, \textit{func} is the variable of the system we want to optimise (Eq. 12). It can be the net flux of the reaction \( k \) (\( v_k \)), the concentration of the metabolite \( i \) (\( C_i \)), or the Gibbs free energy of the reaction \( k \) (\( \Delta G_k' \)). Eq. 14 and 15 represent the lower/upper limits of the variables \( v_k \) and \( C_i \), respectively.

Here we assume that a reaction is blocked (i.e. its enzyme is absent or inactive) when \( v_k = 0 \) and \( \Delta G_k' \neq 0 \), hence the concentration of the metabolites produced from this reaction is equal to zero. Otherwise metabolite concentrations can take any value within the limits. This distinction avoids that the activity coefficient of the other metabolites (which is a function of the total volume occupied by the molecules) is altered by the false presence of product molecules.

Eq. 16 restricts the flux of reaction \( k \), \( v_k \) of the metabolic network to the direction of the corresponding Gibbs free energy drop, \( \Delta G_k' \), hence \( v_k \) and \( \Delta G_k' \) have opposite sign. The prediction of a negative flux value means that the reaction goes in the opposite direction to that indicated in \( S \).

If \( \Delta G_k' \) is estimated as zero (thermodynamic equilibrium), i.e. \( \text{abs}(\Delta G_k') < \text{tol} \), then the term \( \Delta G_k'/\text{tol} \) (Eq. 17) becomes a very small number, which forces \( v_k \) to take a value close to zero in order to satisfy the constraint given in Eq. 16.

Unlike previous thermodynamically based stoichiometric models, in this formulation there is an explicit relationship between the crowding conditions (macromolecular crowders and small molecules such as metabolites) and the thermodynamic constraint. This relationship is given by the calculation of \( \Delta G_k' \) (Eq. 18) which is a function of the activity of the metabolites (\( a_i \)) (Eq. 2–4, 7 and 8), i.e. their concentrations (\( C_i \)) and activity coefficients (\( \gamma_i \)). Here \( \gamma_i \) (Eq. 8) takes also into account the presence of background molecules. Eq. 19 indicates the parameters related with the properties of these molecules, i.e. concentration, molecular weight, and specific volume.

The parametric information required by this NL formulation is \( \Delta_l G^{0*} \) of all the metabolites at the \( \text{pH} \), and \( I \) analysed, the specific volume (\( \nu \)) and the molecular weight \( M \) of the molecules of the system (including \( \nu_{\text{crowder}} \) and \( M_{\text{crowder}} \)), and the tolerance value for the Gibbs free energy change (\( \text{tol} \)), that is the value beyond which \( \Delta G_k' \) is considered as non-zero, in this paper we use \( \text{tol} = 10^{-3} \).

In this paper, the thermodynamic analysis is focused on intracellular reactions and not on the transport of the metabolites across the membrane, except for those transport reactions coupled with biochemical ones, such as the transport of the hydrogen ion \( \text{H}^+ \) (Eq. 5).
Due to the non-linear nature of the problem expressed in Eq. 12–19 several local optima can be computed. In order to increase our chances to approximate the global optimum we use the following initial guesses according to the type of objective function (func) to optimise:

(a) $\text{func} = v_f$, $f \leq n$ (i.e. intracellular fluxes). The initial guesses composed by all the fluxes $v_k$ ($k \neq f$), metabolite concentrations $\ln(C_i)$, and Gibbs free energy changes $\Delta G'_r$ are given by the solution of the linear optimisation (LP) problems (i) the corresponding mass balance (Eq. 12-14), where $\text{func} = v_f$ and (ii) the maximisation/minimisation of $\Delta G'_r$ using the linear Eq. 2–4, 7, 12, 15, 18 assuming $\gamma_i = 1$, whose vector solution includes $\Delta_r G'$ and $\ln(C)$.

(b) $\text{func} = v_f$, $n < f < N$ (i.e. extracellular fluxes). The initial guesses for the fluxes are estimated as in (a). The initial guess for $\Delta_r G'$ is a unit vector whose elements have the opposite sign of the corresponding intracellular flux previously calculated. The initial guesses for $\ln(C)$ are $(\ln(C_{\min}))$ and $(\ln(C_{\max}))$.

(c) $\text{func} = \ln(C_i)$, $1 \leq nr \_ of \_ species$. The initial guesses for the fluxes are the upper limit ($v_{\text{max}}$) and the lower limit ($v_{\text{min}}$) while $\Delta_r G'$ is estimated from the solution of the LP of maximising/minimising $\text{func}$ constrained by Eq. 14, 18 with $\gamma_i = 1$.

(d) $\text{func} = \Delta_r G'_r$, $f \leq n$. The initial guesses are computed as in (c).

RESULTS AND DISCUSSION

Thermodynamic feasibility of lactic acid pathway and glycolysis

The importance of the environmental conditions $I$, $T$, $pH$, $pMg$ (the minus logarithm of Magnesium ions) and of the concentration of metabolites on $\Delta_r G'$ of the reactions and therefore their feasibility is unquestionable (27). Vojinovic and Stockar (6) have shown how these intracellular factors can turn an infeasible pathway into a feasible one. However, how important is the presence of background molecules on this type of thermodynamic analysis has not yet been clarified.

We take as an example the same lactic acid metabolic pathway and the $\Delta_r G'^{0\_r}$ values reported in (6) (Table 1) to analyse the impact of the crowding condition on the feasibility of the pathway. Here, the standard state is defined as $I = 0$ M, $pH = 7$, and $T = 298.15$ K. The solvent is considered as a continuum and therefore neglected.
TABLE 1 Lactic acid pathway, transformed standard reaction Gibbs free energy, and metabolites range concentration

<table>
<thead>
<tr>
<th>Reactions</th>
<th>$\Delta rG^\circ$ (6) [kcal mol$^{-1}$]</th>
<th>Metabolites range concentration (Range 2 (6)) [mM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Rx 1) glc + atp $\leftrightarrow$ g6p + adp</td>
<td>-4.00</td>
<td>13 dpg</td>
</tr>
<tr>
<td>(Rx 2) g6p $\leftrightarrow$ f6p</td>
<td>0.39</td>
<td>2pg</td>
</tr>
<tr>
<td>(Rx 3) f6p + atp $\leftrightarrow$ fdp + adp</td>
<td>-3.40</td>
<td>3pg</td>
</tr>
<tr>
<td>(Rx 4) fdp $\leftrightarrow$ dhap + g3p</td>
<td>5.70</td>
<td>adp</td>
</tr>
<tr>
<td>(Rx 5) dhap $\leftrightarrow$ g3p</td>
<td>1.79</td>
<td>atp</td>
</tr>
<tr>
<td>(Rx 6) g3p + nad + pi $\leftrightarrow$ 13dpg + nadh</td>
<td>1.50</td>
<td>dhap</td>
</tr>
<tr>
<td>(Rx 7) 13dpg + adp $\leftrightarrow$ 3pg + atp</td>
<td>-4.50</td>
<td>f6p</td>
</tr>
<tr>
<td>(Rx 8) 3pg $\leftrightarrow$ 2pg</td>
<td>1.09</td>
<td>fdp</td>
</tr>
<tr>
<td>(Rx 9) 2pg $\leftrightarrow$ pep</td>
<td>0.39</td>
<td>g3p</td>
</tr>
<tr>
<td>(Rx 10) pep + adp $\leftrightarrow$ pyr + atp</td>
<td>-7.50</td>
<td>g6p</td>
</tr>
<tr>
<td>(Rx 11) pyr + nadh $\leftrightarrow$ lac + nad</td>
<td>-5.99</td>
<td>glc</td>
</tr>
</tbody>
</table>

Because SPT considers no other particle interactions than the excluded volume, we use $\Delta rG^\circ$ at $I = 0$ M, instead of the physiological $I = 0.2$ M (which involves electrostatic interactions) to carry out the thermodynamic analysis.

The optimisation problem given by Eq. 2-4, 7, 8, 12, 14, 18, 19 was solved using the function fmincon in MATLAB R2011a (The MathWorks, Natick, MA), where the variable func (Eq. 12) can be any $\Delta rG^\circ$. If a solution with $\Delta rG^\circ < 0$ is found then the pathway is determined as thermodynamically feasible.

To evaluate the crowding effects on the feasibility of the lactic pathway, we set a fixed concentration lower limit of the all metabolites of $10^{-2}$ mM, we vary the upper limit from 2 to 3.5 mM, and estimate the $\Delta rG^\circ$ values of the reactions of Table 1 in the presence of different concentrations of one type of background molecules (or crowder) of 72 kDa (Minton).

Assuming the size of all species (metabolites and macromolecules) is proportional to the molecular weight of their neutral form with specific volume $\nu_i = 0.73 \text{ cm}^3 \text{ g}^{-1}$ (17), their radii can be estimated using Eq. 11 (Scheme $r \sim M$).

The $r$ values estimated for some solutes such as glucose, water, ethanol, and glycerol using $\nu = 0.73 \text{ cm}^3 \text{ g}^{-1}$ are within the experimental radii compiled by Tang and Bloomfield (22).
Fig. 1A shows the feasibility diagram where the line indicates the minimum crowder concentration \( C_{\text{crowder}} \) that makes the lactic pathway feasible \((\Delta G' < 0, \text{white area})\) for each concentration upper limit tested.

**FIGURE 1** Feasibility diagram of the lactic acid pathway using different upper limit of metabolites’ concentration (the lower limit was set as \(10^{-2} \text{ mM}\)) and crowder concentration \((r_{\text{crowder}} \approx 72 \text{ kDa})\). (A) Scheme \( r \sim M \), the radii of all metabolites is proportional to the molecular weight. (B) Scheme \( r = 0 \), the radii of all metabolites is neglected. The feasible region is the area above the line (white area).

In order to simplify the problem, we assume that small molecules with \( M < 72 \text{ kDa} \) (including all metabolites) are volumeless particles \((r = 0)\). Then Eq. 8 reduces to \( \ln \gamma_j = -\ln(1 - S_j) \) where \( S_j \) represents the volume fraction occupied by molecules, in this case only by macromolecules. Fig. 1B shows the feasibility diagram for metabolites with negligible radii.

For both schemes \( r = 0 \) and \( r \sim M \) the decrease of the available volume due to the crowder’s presence push the pathway toward the feasible region for a fixed upper limit.

However, when the radii of the metabolites are considered (scheme \( r \sim M \)) the pathway becomes feasible at lower macromolecules concentrations. This is because the available volume to the centre of mass of the metabolites decreases with the molecules size, which is taken in account by the second, third and fourth term of Eq. 8.

The importance of the parameter \( r \) on the feasibility of a pathway is also evident when a comparison is made using a vector \( r_{\text{random}} \) of random radii taken from the range \( r \pm 0.1r \) (where \( r \) is proportional to \( M \) and \( \nu = 0.73 \text{ cm}^3 \text{ g}^{-1} \) as mentioned before).

The feasibility diagram (Fig. 2) shows the general tendency to decrease the need of high \( C_{\text{crowder}} \) as the range of metabolite concentration is wider (i.e. increasing the upper limit). Nevertheless, the specific \( C_{\text{crowder}} \) value at which the lactic acid pathway turns feasible depends on the molecular radii of the metabolites involved used as shown in Fig. 2.
FIGURE 2 Feasibility diagram of the lactic acid pathway for 4 different random radii vectors $r_{\text{random}}$, at different metabolites concentration upper limit (lower limit was set as $10^{-2}$ mM) and crowder concentration ($r_{\text{crowder}} \sim 72$ kDa). The feasible region is the area above each line.

Although the intracellular crowding conditions may turn a positive $\Delta_f G^r$ to a negative value, they cannot overcome by themselves the influence of uncertainties in the measurements of $\Delta_f G^{0r}$ and of the concentration of the metabolites.

For example using the physiological concentration ranges reported as Range 2 in Vojinovic and Stockar (6) (see Table 1), the lactic pathway could not be predicted as feasible (scheme $r \sim M$) even at high background molecules concentration. In fact for this particular set of concentrations, Vojinovic and Stockar (6) found that the feasibility of the pathway could be reached only under unrealistic $I$, $pH$, and $pMg$.

On the other hand, slight variations were experimentally detected in the intracellular concentration of pep and dhap in Escherichia coli growing on either glycerol or acetate (30), suggesting that $\Delta_f G^{0r}$ of the reactions involved in lower glycolysis (from dhap to pep) must be near equilibrium, so that these small variations cause the activation of glycolysis in the direction required depending on the carbon source used, i.e. dhap $\rightarrow$ pep for glycerol, and pep $\rightarrow$ dhap for acetate.

Furthermore, the crowding conditions could help to change the reactions direction in glycolysis. For example, lumping the reactions Rx 8 and Rx 9 of Table 1 into the single reaction

$$3 \text{pg} \rightarrow \text{pep}$$

(20)

gives $\Delta_f G^{0r}$ ($I=0$ M, $pH=7$, $T = 298.15$ K) = 1.48 kcal mol$^{-1}$, with the concentration limits reported by Bennett et al. (30) for the growth on glycerol: $1.04$ mM $< C_{\text{pep}} < 1.73$ mM, and $3.36$ mM $< C_{3\text{pg}} < 4.95$ mM.
The $\Delta G'$ estimated for this reaction in absence of crowders ($C_{\text{crowder}} = 0 \text{ M}$) indicates the tendency towards the formation of 3pg ($\Delta G' > 0$), i.e. the likely glycolysis pathway is active in the direction pep $\rightarrow$ g6p, as was expected for growing on acetate. However, under the presence of 16.9 mM (or higher) of crowders of 72 kDa (scheme $r \sim M$), which occupy 89% of the total volume, the likely direction of reaction of Eq. 20 is (3pg $\rightarrow$ pep) as indicated by a computed negative $\Delta G'$.

While the estimated volume fraction occupied by molecules is more than twice that expected in the cell (about 40%) (7), this example shows qualitatively the change of reaction direction due to the excluded volume effects.

Nevertheless, other intracellular conditions like $I$, $pH$, $T$, and $pMg$ (not considered in this example) may affect the $\Delta G^0$, value making it to approach zero, as was suggested by Bennett et al. (30). Then the crowding condition needed to change the reaction direction could be within the physiological 40%.

Moreover, the cell consists of a heterogeneous medium where regions more crowded than others may occur, favouring a metabolic pathway that in other circumstances or even other regions would be infeasible.

Prediction of thermodynamically feasible flux distributions. *Actinobacillus succinogenes*’ metabolic network

In order to study the influence of crowding conditions on the flux distribution, thermodynamically constrained Flux Variability Analysis (31) was carried out on the central carbon pathways of *A. succinogenes* (Fig. 3, the detailed information of the 53 reactions and 41 metabolites is in the Supporting Material).

In this section we use the values of standard Gibbs free energy of formation of all metabolites ($\Delta G^0$) estimated by the expanded group contribution method proposed by Jankowski et al. (32).

Both input and output fluxes to/from the cell were fixed to the experimental data reported by Vlysidis et al. (24) for the batch production of succinic acid from glycerol by *A. succinogenes*. The aim here is to determine the intracellular and unmeasured fluxes.
FIGURE 3 Central carbon pathways of *Actinobacillus succinogenes* for the production of succinic acid from glycerol. Unidirectional arrow: the flux through the reaction is considered beforehand unidirectional. Bidirectional arrow: the reaction is considered beforehand as reversible, the bigger arrow indicates how the reaction was written in the stoichiometric matrix. Metabolites: 13dp, 3-phospho-D-glyceroyl phosphate; 2pg, glycerate 2-phosphate; 3pg, 3-phospho-D-glycerate; 6pgc, 6-phospho-D-gluconate; 6pgl, 6-phospho-D-glucono-1,5-lactone; ac, acetate; acald, acetaldehyde; accoa, acetyl-CoA; actp, acetyl phosphate; adp, ADP; atp, ATP; co2, CO₂; coa, coenzyme A; d hap,
dihydroxyacetone phosphate; e4p, erythrose 4-phosphate; etoh, ethanol; f6p, fructose 6-phosphate; fdp, fructose 1,6-bisphosphate; for, formate; fum, fumarate; g3p; glyceraldehyde 3-phosphate; g6p, glucose 6-phosphate; glyc, glycerol; glyc3p, glycerol 3-phosphate; h2o, water; lac, lactate; mal, malate; nad, nicotinamide adenine dinucleotide; nadh, nicotinamide adenine dinucleotide – reduced; nadph, nicotinamide adenine dinucleotide phosphate; nedph, nicotinamide adenine dinucleotide phosphate – reduced; oaa, oxaloacetate; pep, phosphoenolpyruvate; pi, phosphate (orthophosphate); pyr, pyruvate; r5p, ribose 5-phosphate; ru5p, ribulose 5-phosphate; s7p, sedoheptulose 7-phosphate; succ, succinate; xu5p, xylulose 5-phosphate. Reactions and their respective enzymes: PGI, glucose 6-phosphate isomerase; PFK, 6-phosphofructokinase; FBP, fructose-1,6-bisphosphatase; FBAr, fructose-bisphosphate aldolase; TPI, triose-phosphate isomerase; GAPD, 3-phosphoglycerate dehydrogenase; PGK, 3-phosphoglycerate kinase; PGM, phosphoglycerate mutase; ENO, enolase; PYK, pyruvate kinase; LDH, lactate dehydrogenase; PFL, pyruvate formate lyase; FDH, formate dehydrogenase; PTAr, acetate phosphotransferase; ACKr, acetate kinase; ACYP, acetyl phosphate phosphohydrolase; PDH, pyruvate dehydrogenase complex; ME2, malic enzyme (NADP); OAD, Oxaloacetate decarboxylase; ACALD, acetaldehyde dehydrogenase; ALCD2x, alcohol dehydrogenase; PPCK, phosphoenolpyruvate carboxykinase; MDH, malate dehydrogenase; FUM, fumarase; FRD, fumarate reductase; G6PDH2r, glucose 6-phosphate dehydrogenase; PGL, 6-phosphoglulonolactonase; GND, phosphogluconate dehydrogenase; RPI, ribose-5-phosphate isomerase; RPE, ribulose 5-phosphate 3-epimerase; TKT1, transketolase; TALA, transaldolase; TKT2, transketolase; GLYK, glycerol kinase; G3PD2, glycerol 3-phosphate dehydrogenase (NADP); THD2, nad(p) transhydrogenase; EX_pyr(e), pyruvate exchange; EX_coa(e), coenzyme A exchange; EX_co2(e), co2 exchange; EX_pi(e), phosphate exchange; EX_lac(e), Lactate exchange; EX_for(e), formate exchange; EX_ac(e), acetate exchange; EX_ etoh(e), ethanol exchange; EX_ succ(e), succinate exchange; EX_fum(e), fumarate exchange; EX_glyc(e), glyceraldehyde exchange; EX_oaa(e), oxaloacetate exchange; EX_h2o(e), water exchange; ATPm, ATP requirements for other reactions; NADPHm, NADPH requirements for other reactions; NADHm, NADH requirements for other reactions.

The uptake flux of glycerol was set to 5.6524 mmol g_W^{-1} h^{-1}, while the excreted products were set to the following fluxes: acetate to 0.5965, formate to 0.4735, and succinate to 3.8711 mmol g_W^{-1} h^{-1}. Biomass production was equal to 1.8504 mmol g_W^{-1} h^{-1} while the cell composition is assumed as CH_2O_0.5N_0.2 (33). Since lactate, ethanol, fumarate, pyruvate, and oxaloacetate were not experimentally detected in the medium (24) their production fluxes were set to zero.

The concentration limits of all the metabolites were set to C_{max} = 0.01 M and C_{min} = 10^{-5} M, while the crowding concentration was set to the fixed value C_{crowder1} = 6.5 mM for macromolecules of M_{crowder1} = 72 kDa and C_{crowder2} = 300 mM for other small molecules of M_{crowder2} = 200 Da. The limits of intracellular fluxes were v_{max} = 100 mmol g_W^{-1} h^{-1} and v_{min} = -100 mmol g_W^{-1} h^{-1} (5). The water flux exchange to/from the cell was unconstrained.
If water is assumed as a volumeless particle, then the number of molecules that can fit into the system (i.e., the water concentration) is proportional to the space not occupied by other molecules and the pure water concentration $C_{\text{water}} = (1 - S_3)C_{\text{water}}^{st}$, while the activity coefficient (Eq. 8) is reduced to $\gamma_{\text{water}} = 1/(1 - S_3)$. Therefore, the water activity ($a_{\text{water}} = \gamma_{\text{water}} C_{\text{water}} / C_{\text{water}}^{st}$) is always equal to unity.

The optimisation problem (constrained by thermodynamics and mass balances) given by Eq. 2–4, 7, 8, 12–19 was also solved using the function fmincon (MATLAB R2011a). In order to increase the possibility of finding a minimum/maximum value close to the global one 4 different initial guesses were tested for each variable $x$ (as discussed above). These values were then compared with the solution obtained when the objective function is another variable $y$ (with its own 4 initial guesses), which is equivalent to increasing the number of effective initial guesses per variable.

Table 2 shows the environmental conditions (I, $pH_c$, $pH_e$, and macromolecular crowding) and the molecules’ volume consideration ($r = 0$ for volumeless particles, and $r \sim M$ when the radii of the molecules is proportional to the molecular weight, the radii of molecules with $M < 20$Da is neglected) employed for the estimation of the maximum and minimum flux values predicted in this example.
<table>
<thead>
<tr>
<th>Condition 1</th>
<th>Condition 2</th>
<th>Condition 3</th>
<th>Condition 4</th>
<th>Condition 5</th>
<th>Condition 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>(biological standard conditions, volumless molecules)</td>
<td>(change in ( \text{pH}_c ), volumless molecules)</td>
<td>(change in ( \text{pH}_c ), volumless molecules)</td>
<td>(change in ( I ), volumless molecules)</td>
<td>(molecules’ volume)</td>
<td>(macromolecules’ volume)</td>
</tr>
<tr>
<td>( \text{pH}_c = 7 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
<td>( \text{pH}_c = 6.7 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
<td>( \text{pH}_c = 7 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
<td>( \text{pH}_c = 7 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
<td>( \text{pH}_c = 7 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
<td>( \text{pH}_c = 7 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
</tr>
<tr>
<td>( \text{pH}_c = 6.2 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
<td>( \text{pH}_c = 6.2 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
<td>( \text{pH}_c = 7.4 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
<td>( \text{pH}_c = 6.2 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
<td>( \text{pH}_c = 6.2 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
<td>( \text{pH}_c = 6.2 ) ( T = 298.15 \text{ K} ) ( I = 0 \text{ M} ) ( r = 0 )</td>
</tr>
<tr>
<td>( r &lt; 20 \text{ Da} )</td>
<td>( r = 0 ) for ( r &lt; 72 \text{ kDa} )</td>
<td>( r &lt; 20 \text{ Da} )</td>
<td>( r = 0 ) for ( r &lt; 72 \text{ kDa} )</td>
<td>( r &lt; 20 \text{ Da} )</td>
<td>( r = 0 ) for ( r &lt; 72 \text{ kDa} )</td>
</tr>
</tbody>
</table>
Conditions 2–4 express one change at a time in the biological standard values $I$, $pH_e$, or $pH_t$ towards physiological conditions. The aim of this is to compare the crowding effect (condition 5 and 6) on flux distribution regarding other environmental factors.

Although the crowders’ concentration is the same in all the conditions tested ($C_{\text{crowder}1} = 6.5$ mM for molecules of $M_{\text{crowder}1} = 72$ kDa and $C_{\text{crowder}2} = 300$ mM for molecules of $M_{\text{crowder}2} = 200$ Da), the assumption that the radii of all the molecules is zero (condition 1–4) causes their activity coefficient to be equal to one, therefore the presence of background molecules does not have an effect on the thermodynamics and flux distribution of the metabolic pathway.

As shown before, the crowding conditions can change the directionality of some reactions. For this metabolic network (Fig. 3), and neglecting the (group contribution-) uncertainty in $\Delta_rG^{\ominus}$, a greater number of reactions were predicted as reversible for condition 5 compared with those under conditions 1–4 (Fig. 4A).

![FIGURE 4](image)

(Fig. 4) (A) Comparison of the number of reactions predicted as reversible under conditions 1–6. (B) Number of reactions whose range of fluxes (under condition 2–6) differs from those predicted for condition 1. The group contribution uncertainty in $\Delta_rG^{\ominus}$ is not considered.

The fact that more reactions go in both backward and forward directions affects the attainable flux of other reactions as the system is theoretically more robust. FVA reveals that the range of fluxes predicted under crowding conditions (condition 5) is either equal or wider than those from conditions 1–4 (coloured lines in Fig. 5).
A comparison of the number of reactions (under conditions 2–5) whose range of fluxes differ from those predicted under biological standard conditions (condition 1) suggests that the flux distributions are more sensitive to crowding conditions primarily, and to ionic strength secondarily (Fig. 4B).

The impact of pH_c and pH_e was less significant on the flux ranges (Fig. 4B). That was expected since no drastic changes were tested inasmuch as A. succinogenes regulates the intracellular pH_c within a narrow range (34), and pH_e was maintained between 6.2 and 7.4 during the fermentation (24).

In this study we focus on the individual effect of I, pH_c, pH_e, and crowders’ concentration, but certainly the environmental conditions exert a combined effect on the feasibility of metabolic networks. Hence, an appropriate model which combines the effect of electrostatic and excluded volume interactions (the two parameters with more influence on the flux distributions), together with reliable thermodynamic data and metabolite concentrations, would give a deep insight in the complex intracellular processes.

In order to simplify the calculation of the activity coefficients, we assume that molecules with molecular weight less than 72 kDa, i.e. all the metabolites involved in the pathway analysed, have negligible volume (Condition 6).

Since the radii are equal to zero, all terms of the r.h.s. of Eq. 8 vanish except for the first one, which represents the volume fraction occupied by molecules. In this case, this is determined by a constant concentration of 72 kDa macromolecules equivalent to 34% of the total volume.

A comparison between the ranges of fluxes predicted using the full Eq. 8 (condition 5) and its simplified version (condition 6) shows no difference (coloured lines in Fig. 5).
This suggests that a good approximation of the crowding effect can be easily incorporated to any linear thermodynamically-stoichiometric model (e.g. Henry et al. (5) and Hoppe et al. (12)) by the addition of parameter \( RT \ln(\gamma_i) = RT / (1 - S_i) \) to \( \Delta_i G^0 \).

However there are slight differences in the range of \( \Delta_i G' \) estimated for both approximations (conditions 5 and 6) (coloured lines in Fig. 6). Although for this example these \( \Delta_i G' \) differences have no influence on the range of fluxes, they could be more important for reactions near equilibrium since small changes are needed to change the directionality or feasibility of a reaction.

![Diagram](image)

**FIGURE 6** Comparison of \( \Delta_i G' \) estimated for some reactions under conditions 1 to 6. Coloured lines indicate the range of \( \Delta_i G' \) when the uncertainty in \( \Delta_i G^0 \) is not taken into account, while grey lines behind them are the \( \Delta_i G' \) predicted considering the uncertainty.

Up to this point the uncertainty \( SE \) associated with the transformed standard Gibbs free energy of the reaction, i.e. \( \Delta_i G^{0t} = \sum_{\text{Products}} \eta_i \Delta_i G^0_i - \sum_{\text{Reactants}} \eta_i \Delta_i G^0_i \), estimated by the group contribution method was not taken into account.

In order to incorporate the influence of the uncertainty on the thermodynamic analysis, the error variable \( E \) is added to the Eq. 3, giving

\[
\Delta_i G'_{\text{err}} = \sum_{i=1}^{\text{Products}} \eta_i [\Delta_i G^{0t}_i + RT \ln(a_i)] - \sum_{i=1}^{\text{Reactants}} \eta_i [\Delta_i G^{0t}_i + RT \ln(a_i)] + E \tag{21}
\]

The variable \( E \) can take any value within the limits \( \pm SE \) of each reaction as calculated by the group contribution method (32).
The new optimisation problem that takes into account the uncertainty in $\Delta G^{\circ}$ (Eq. 2, 4, 7, 8, 12–19, 21) predicts a range of $\Delta G$ wider than those where this is not considered (or neglected) (grey lines in Fig. 6).

This causes some reactions to change their directionality, so that in general the range of fluxes is also estimated to be wider for all conditions tested (grey lines in Fig. 5). For this example, the uncertainty given by the group contribution method masks the effects of $I$, $pH_c$, $pH_e$, and crowding conditions on the estimation of the range of fluxes (Fig. 5). This reflects the importance of using more accurate thermodynamic data.

Finally, in order to determine the impact of crowding conditions on the maximum attainable biomass flux, the (thermodynamically-constrained) Flux Balance Analysis (Eq. 2–4, 7, 8, 12–19) was applied to the central carbon pathways of A. succinogenes (Fig. 3). The uncertainty in $\Delta G^{\circ}$ is not taken into account.

Experimental observations indicate the relation between the CO$_2$ availability in the medium and the growth and succinate production of A. succinogenes. The following uptake fluxes were set: glycerol 5.6524 mmol g$_{DW}^{-1}$ h$^{-1}$, CO$_2$ 1.7602 mmol g$_{DW}^{-1}$ h$^{-1}$, and phosphate 0.1598 mmol g$_{DW}^{-1}$ h$^{-1}$. The remaining extracellular fluxes were left unrestricted. The concentration and intracellular flux limits mentioned before were unchanged.

The maximum biomass production rate estimated was 1.8495 mmol g$_{DW}^{-1}$ h$^{-1}$ under both crowding conditions and the biological standard condition (condition 5 and 1 respectively, see Table 2), which agrees with the experimental results (24).

Only small differences were found for the maximum succinate flux: (condition 5) 4.100 mmol g$_{DW}^{-1}$ h$^{-1}$ and (condition 1) 3.5302 mmol g$_{DW}^{-1}$ h$^{-1}$. Both values are within the range $v_{\text{succinate,exp}} \pm 9\%$. This confirms that, for the system analysed, crowding conditions acting on the robustness of the metabolic network, aid in reaching the same output flux but with different flux distributions.

It is noteworthy that the concentration limits used here have an important role in this conclusion, since the use of other very restricted limits may turn complete pathways to be infeasible (as shown in the lactic acid pathway example). However, where the crowding conditions may favour the flux through them, this will affect the estimation of some output fluxes.

CONCLUSIONS

Along with other environmental conditions such as $I$, $T$, $pH_c$, and $pH_e$, the presence of background molecules may also play an important role in the feasibility and directionality of reactions.

In this paper, SPT was used to incorporate the crowding effects on the thermodynamic analysis. It was found that changes in the crowding conditions can push a pathway to one direction or another. This suggests the importance of heterogeneity, with more
crowded regions than others or the compartmentalisation (in eukaryotes) of the intracellular space, which could favour the feasibility of certain pathways.

Moreover, a crowded medium may affect the flux distribution in a metabolic system. In fact, FVA estimates a wider range of fluxes when crowding is taken into account compared with the standard conditions, indicating the potential contribution of crowding to the robustness of the system, which allows the microorganism to deal with environmental or genetic changes.

Nevertheless, the enhancement in the metabolite’s activity caused by the crowding conditions is accompanied by a decrease in the diffusion and the “degree of mixing” of the molecules so the encounter and further reaction between two reactants is less likely (35, L. Angeles-Martinez and C. Theodoropoulos, unpublished), which was not considered in this paper.

Due to the high influence of the excluded volume and electrostatic interactions (related with the crowding conditions and ionic strength) on the feasibility and flux distribution of a metabolic network, the use of a model that incorporates the combined effect of these two interactions would give a better understanding of environmental conditions’ control on the metabolism.

**APPENDIX. CALCULATION OF THE $\Delta_f G^0$.**

In this section the methodology for the estimation of transformed standard Gibbs free energy of formation of the metabolite $i$ ($\Delta_f G^{0_i}$) is presented, which for the purpose of this paper is represented by its pseudoisomer group $i$, at different values of $I$ and $pH$.

A pseudoisomer group $i$ is made up of all the protonated species $j$ formed by the dissociation of the metabolite $i$ in aqueous solutions (27), whose proportions depend on the logarithm of the acid dissociation constant $pK_a$. So for example for the reaction $\text{HA} \xrightarrow{pK_a} \text{A}^-$, if the change in standard Gibbs free energy of the species $\text{HA}$ ($\Delta_f G^{0_{\text{HA}}}$) is known then $\Delta_f G^{0_{\text{A}^-}}$ for the species $\text{A}^-$ can be calculated by Eq. A1

$$\Delta_f G^{0_{\text{A}^-}} = \Delta_f G^{0_{\text{HA}}} + pK_a RT \ln(10)$$

Eq. A1

The $\Delta_f G^{0_j}$ values of species $j$ (at standard biological conditions $pH = 7$, $I = 0$ M and $T = 298.15$ K) of several metabolites can be found in databases (e.g. Alberty (24) and Li et al. (36)) or estimated by any group contribution method and other predictive method, for example the one proposed by Jankowski et al. (32), or by Noor et al. (37).

In this paper we use the Calculator Plugins, MarvinSketch ver. 5.5.1.0, 2011, ChemAxon [http://www.chemaxon.com] for the estimation of the $pK_a$ values of all the metabolites (the molecular structure is required).
Once the $\Delta_f G_j^0$ values for all the species of a metabolite are found, then if the environmental conditions $I$ and $pH$ are different from the standard biological conditions, the transformed Gibbs free energy change $\Delta_f G_j^{0'}$ (when $pH$ is considered constant) can be calculated as (27):

$$\Delta_f G_j^{0'} = \Delta_f G_j^0 - \frac{A z_j^2 I^2}{1 + B I^2} \left[ RT \ln(10) - N_{H,j} \Delta_f G_H^0 \right]$$  \hspace{1cm} \text{Eq. A2}

where

$$\Delta_f G_H = \Delta_f G_H^0 - \frac{A I^2}{1 + B I^2} \left[ RT \ln(10) - RT \ln(10^{-pH_{\text{ref}}}) \right]$$  \hspace{1cm} \text{Eq. A3}

The constants of the extended Debye-Hückel equation are given as $A = 0.51$ and $B = 1.6$. $z_j$ is the electric charge of species $j$, $N_{H,j}$ the number of H atoms in species $j$ and $pH_{\text{ref}} = 7$ is the reference $pH$ at the standard biological conditions.

Finally, assuming that all species $j$ of the metabolite $i$ are in equilibrium, the $\Delta_f G_i^{0'}$ of the pseudoisomer group $i$ can be estimated as (27)

$$\Delta_f G_i^{0'} = -RT \cdot T \cdot \ln \left\{ \sum_{j=1}^{N_{\text{species}}} e^{-\left( \frac{\Delta_f G_j^{0'}}{RT} \right)} \right\}$$  \hspace{1cm} \text{Eq. A5}

**AUTHOR CONTRIBUTIONS**

LAM developed the methodology, performed all the computations and drafted the manuscript. CT supervised the method development and the work on the case studies, reviewed and edited the manuscript.

**ACKNOWLEDGEMENTS**

The financial support CONACYT-Mexico for the PhD of LAM is gratefully acknowledged.

**SUPPORTING CITATIONS**

Reference (38) appears in the Supporting Material.

**REFERENCES**


**NOMENCLATURE**

- **a** Activity of a metabolite (dimensionless).
- **A** Constant of the extended Debye-Hückel equation equivalent to 0.510651 (L^{-0.5} mol^{-1/2}).
- **B** Constant of the extended Debye-Hückel equation equivalent to 1.6 (L^{-0.5} mol^{-0.5}).
- **c** Subscript that indicates intracellular medium (dimensionless).
- **C** Concentration of an intracellular metabolite (mol L^{-1}).
- **C^\text{cmax}** Maximum concentration of an intracellular metabolite (mol L^{-1}).
- **C^\text{cmax}_\text{vector}** Maximum concentration vector of all intracellular metabolites (mol L^{-1}).
- **C^\text{cmin}** Minimum concentration of an intracellular metabolite (mol L^{-1}).
- **C^\text{cmin}_\text{vector}** Minimum concentration vector of all intracellular metabolites (mol L^{-1}).
- **C^\text{cst}** Standard concentration of an intracellular metabolite (mol L^{-1}).
- **e** Subscript that indicates extracellular medium (dimensionless).
- **E** Error variable to consider the uncertainty in standard Gibbs free energy of the reaction estimated by the group contribution method (kcal mol^{-1}).
- **f** Subscript that indicates the reaction used as objective function (dimensionless).
- **func** Objective function to optimise in the NLP formulation.
- **F** Faraday constant equivalent to 2.306 \times 10^{-2} (kcal mV^{-1} mol^{-1}).
- **i** Subscript that indicates the pseudoisomer group of a metabolite \text{i} (dimensionless).
- **I** Ionic strength (M).
- **j** Subscript that indicates the protonated species \text{j} of a metabolite \text{i} (dimensionless).
- **k** Subscript that indicates a specific reaction (dimensionless).
- **m** Number of intracellular metabolites in the metabolic network analysed, the hydrogen ions H^+ is not included (dimensionless).
- **M** Molecular weight of metabolite (g mol^{-1}).
- **M^\text{vector}** Molecular weight vector of all metabolites (g mol^{-1}).
- **n** Number of intracellular reactions in the metabolic network analysed (dimensionless).
- **N** Number of reactions for the transport of metabolites outside/inside the cell (dimensionless).
- **N_{\text{A}}** Avogadro’s number equivalent to 6.02214 \times 10^{23} (molecules mol^{-1}).
- **N_H** Number of hydrogen atoms in a metabolite (dimensionless).
- **pH** Potential of hydrogen (dimensionless).
- **pH^{\text{ref}}** Reference pH = 7 at standard biological conditions (dimensionless).
- **pK_a** Acid dissociation constant (dimensionless).
r Radium of a molecule (cm).
\( \mathbf{r} \) Radii vector of all molecules (cm).
\( \mathbf{r}_{\text{random}} \) Random radii vector (cm).
\( R \) Gas constant equivalent to \( 1.9858 \times 10^{-3} \) (kcal mol\(^{-1}\) K\(^{-1}\)).
\( S \) \( m \times (n+N) \) stoichiometric matrix of the metabolic network (dimensionless).
\( s_i \) Variable of the Scale Particle Theory equation, where \( 0 \leq x \leq 3 \).
\( SE \) Uncertainty associated to the standard Gibbs free energy of the reaction estimated by the group contribution method (kcal mol\(^{-1}\)).
\( tol \) Tolerance value of what is considered a non-zero Gibbs free energy (kcal mol\(^{-1}\)).
\( T \) Temperature of the medium (K).
\( v \) \( n \times 1 \) vector of the fluxes through the metabolic network (mmol L\(^{-1}\) g\(_{DW}\)^{-1}).
\( v_{\text{max}} \) Maximum flux through a reaction (mmol L\(^{-1}\) g\(_{DW}\)^{-1}).
\( v_{\text{max}} \) Maximum flux vector of all reactions (mmol L\(^{-1}\) g\(_{DW}\)^{-1}).
\( v_{\text{min}} \) Minimum flux vector of all reactions (mmol L\(^{-1}\) g\(_{DW}\)^{-1}).
\( v_{\text{min}} \) Minimum flux through a reaction (mmol L\(^{-1}\) g\(_{DW}\)^{-1}).
\( vol_{\text{av}} \) Available volume in the system (L).
\( vol_{\text{tot}} \) Total volume of the system (L).
\( z \) Electrical charge of the metabolite species \( j \) (dimensionless).
\( \Delta_f G_H \) Gibbs free energy of formation of the hydrogen (kcal mol\(^{-1}\)).
\( \Delta_f G'' \) Standard Gibbs free energy of formation of a metabolite (kcal mol\(^{-1}\)).
\( \Delta_f G''' \) Transformed standard Gibbs free energy of formation of a metabolite (kcal mol\(^{-1}\)).
\( \Delta_r G' \) Transformed Gibbs free energy of a reaction formation of a metabolite (kcal mol\(^{-1}\)).
\( \Delta_r G'' \) Total transformed Gibbs free energy of a reaction (kcal mol\(^{-1}\)).
\( \Delta_r G''' \) Total transformed Gibbs free energy vector of all intracellular reactions (kcal mol\(^{-1}\)).
\( \Delta_r G''_u \) Transformed Gibbs free energy of a biochemical reaction of a metabolite (kcal mol\(^{-1}\)).
\( \Delta_r G''_u \) Transformed Gibbs free energy of a transport reaction of a metabolite (kcal mol\(^{-1}\)).
\( \Delta \phi \) Difference between the internal membrane potential and the external one (mV).
\( \gamma \) Activity coefficient of a metabolite (dimensionless).
\( \eta \) Stoichiometric coefficient of a metabolite in a reaction (dimensionless).
\( \rho \) Density number (molecules cm\(^{-3}\)).
\( \nu \) Specific volume (cm\(^3\) g\(^{-1}\)).
4.2.1. Supplementary Material for Publication 2

<table>
<thead>
<tr>
<th>#</th>
<th>Abbreviation</th>
<th>Compound Name</th>
<th>Kegg compound</th>
<th>Formule</th>
<th>No. of H atoms</th>
<th>predominant species</th>
<th>Charge at pH 7</th>
<th>No. of H atoms</th>
<th>predominant species</th>
<th>ΔG° (kcal/mol)</th>
<th>Molecular Weight (g/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13dpd</td>
<td>3-Phospho-D-glycerol phosphate</td>
<td>CO0026</td>
<td>C3H8O10P2</td>
<td>8</td>
<td>-2</td>
<td>6</td>
<td>-586.8</td>
<td>266.0371</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2pg</td>
<td>Glycerate 2-phosphate</td>
<td>CO0031</td>
<td>C3H7O7P</td>
<td>7</td>
<td>-2</td>
<td>5</td>
<td>-371.9</td>
<td>186.0572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3pg</td>
<td>3-Phospho-D-glycerate</td>
<td>CO0097</td>
<td>C3H7O7P</td>
<td>7</td>
<td>-2</td>
<td>5</td>
<td>-371.9</td>
<td>186.0572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6gpc</td>
<td>6-Phospho-D-glucanate</td>
<td>CO0345</td>
<td>C6H13O10P</td>
<td>13</td>
<td>-2</td>
<td>11</td>
<td>-481.16</td>
<td>276.1352</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6gpl</td>
<td>6-phospho-D-glucan-1,5-lactone</td>
<td>CO1236</td>
<td>C6H11O9P</td>
<td>11</td>
<td>-1</td>
<td>10</td>
<td>-428.52</td>
<td>258.1199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ac</td>
<td>Acetate</td>
<td>CO0033</td>
<td>C2H4O2</td>
<td>4</td>
<td>-1</td>
<td>3</td>
<td>-88.29</td>
<td>60.052</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>acald</td>
<td>Acetaldehyde</td>
<td>CO0084</td>
<td>C2H4O</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>-33.4</td>
<td>44.0526</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>acto</td>
<td>Acetyl-CoA</td>
<td>CO0024</td>
<td>C2H38N7O12P3S</td>
<td>38</td>
<td>-3</td>
<td>35</td>
<td>-785.83</td>
<td>809.5708</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>acp</td>
<td>Acetylphosphate</td>
<td>CO0227</td>
<td>C2H5O5P</td>
<td>5</td>
<td>-1</td>
<td>4</td>
<td>-301.65</td>
<td>140.0319</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>atp</td>
<td>ATP</td>
<td>CO0002</td>
<td>C10H16N5O13P3</td>
<td>16</td>
<td>-3</td>
<td>13</td>
<td>-673.85</td>
<td>507.181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>c02</td>
<td>CO2</td>
<td>CO0011</td>
<td>C2O</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-92.26</td>
<td>62.0248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>coa</td>
<td>Coenzyme A</td>
<td>CO0010</td>
<td>C2H38N7O12P3S</td>
<td>36</td>
<td>-3</td>
<td>33</td>
<td>-751.996</td>
<td>767.5341</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>dhap</td>
<td>Dihydroxyacetone phosphate</td>
<td>CO0011</td>
<td>C3H7O6P</td>
<td>7</td>
<td>-1</td>
<td>6</td>
<td>-320.66</td>
<td>170.0578</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>e4p</td>
<td>Erythrose 4-phosphate</td>
<td>CO0279</td>
<td>C4H9O7P</td>
<td>9</td>
<td>-1</td>
<td>8</td>
<td>-355.62</td>
<td>200.0838</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>etoh</td>
<td>Ethanol</td>
<td>CO0469</td>
<td>C2H6O</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>-43.44</td>
<td>46.0684</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>f6p</td>
<td>Fructose 6-phosphate</td>
<td>CO0085</td>
<td>C6H13O9P</td>
<td>13</td>
<td>-1</td>
<td>12</td>
<td>-431.67</td>
<td>260.1358</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>flp</td>
<td>Fructose 1,6-bisphosphate</td>
<td>CO0354</td>
<td>C6H14O11P2</td>
<td>14</td>
<td>-2</td>
<td>12</td>
<td>-644.17</td>
<td>340.1157</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>for</td>
<td>Formate</td>
<td>CO0058</td>
<td>C2H2O2</td>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>-83.9</td>
<td>46.0254</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>fum</td>
<td>Fumarate</td>
<td>CO0122</td>
<td>C4H4O4</td>
<td>4</td>
<td>-2</td>
<td>2</td>
<td>-143.7</td>
<td>116.0722</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>g3p</td>
<td>Glyceraldelyde 3-phosphate</td>
<td>CO0018</td>
<td>C3H7O6P</td>
<td>7</td>
<td>-1</td>
<td>6</td>
<td>-319.2</td>
<td>170.0578</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>g6p</td>
<td>Glucose 6-phosphate</td>
<td>CO0092</td>
<td>C6H13O9P</td>
<td>13</td>
<td>-1</td>
<td>12</td>
<td>-430.78</td>
<td>260.1358</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>glyc</td>
<td>Glycerol</td>
<td>CO0016</td>
<td>C3H8O3</td>
<td>8</td>
<td>0</td>
<td>8</td>
<td>-116.9</td>
<td>92.0938</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>glyc3p</td>
<td>sn-Glycerol 3-phosphate</td>
<td>CO0093</td>
<td>C3H9O6P</td>
<td>9</td>
<td>-1</td>
<td>8</td>
<td>-328.68</td>
<td>172.0737</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>h</td>
<td>H+</td>
<td>CO0080</td>
<td>H</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-9.53</td>
<td>1.0079</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>h2o</td>
<td>H2O</td>
<td>CO0001</td>
<td>H2O</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>-56.087</td>
<td>18.0153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>lac</td>
<td>Lactate</td>
<td>CO0026</td>
<td>C3H6O3</td>
<td>6</td>
<td>-1</td>
<td>5</td>
<td>-123.17</td>
<td>90.0779</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>mal</td>
<td>Malate</td>
<td>CO0149</td>
<td>C4H6O5</td>
<td>6</td>
<td>-2</td>
<td>4</td>
<td>-201.0</td>
<td>134.0874</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>nad</td>
<td>Nicotinamide adenine dinucleotide</td>
<td>CO0003</td>
<td>C2H27N7O14P2</td>
<td>27</td>
<td>-1</td>
<td>26</td>
<td>-529.59</td>
<td>664.433</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>nadh</td>
<td>Nicotinamide adenine dinucleotide - oxidized</td>
<td>CO0004</td>
<td>C2H27N7O14P2</td>
<td>27</td>
<td>-1</td>
<td>26</td>
<td>-524.32</td>
<td>665.441</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>nadph</td>
<td>Nicotinamide adenine dinucleotide - reduced</td>
<td>CO0005</td>
<td>C2H27N7O14P2</td>
<td>28</td>
<td>-2</td>
<td>26</td>
<td>-742.09</td>
<td>744.4129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>nadph4</td>
<td>Nicotinamide adenine dinucleotide - reduced form</td>
<td>CO0005</td>
<td>C2H13N7O13P2</td>
<td>30</td>
<td>-3</td>
<td>27</td>
<td>-736.82</td>
<td>745.4209</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>oaa</td>
<td>Oxaloacetate</td>
<td>CO0036</td>
<td>C4H4O5</td>
<td>4</td>
<td>-2</td>
<td>2</td>
<td>-190.52</td>
<td>132.0716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>pep</td>
<td>Phosphoenolpyruvate</td>
<td>CO0074</td>
<td>C3H5O6P</td>
<td>5</td>
<td>-2</td>
<td>3</td>
<td>-316.08</td>
<td>168.042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>pi</td>
<td>Phosphate (orthophosphate)</td>
<td>CO0009</td>
<td>H3PO4</td>
<td>3</td>
<td>-2</td>
<td>1</td>
<td>-261.974</td>
<td>97.9952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>phy</td>
<td>Pyruvate</td>
<td>CO0022</td>
<td>C3H4O3</td>
<td>4</td>
<td>-1</td>
<td>3</td>
<td>-112.69</td>
<td>88.0621</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>r5p</td>
<td>Ribose 5-phosphate</td>
<td>CO0117</td>
<td>C5H11O8P</td>
<td>11</td>
<td>-1</td>
<td>10</td>
<td>-394.12</td>
<td>230.1098</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>r6p</td>
<td>Ribulose 5-phosphate</td>
<td>CO0199</td>
<td>C5H11O8P</td>
<td>11</td>
<td>-1</td>
<td>10</td>
<td>-393.5</td>
<td>230.1098</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>s7p</td>
<td>Sedoheptulose 7-phosphate</td>
<td>CO5382</td>
<td>C7H15O10P</td>
<td>15</td>
<td>-1</td>
<td>14</td>
<td>-466.34</td>
<td>290.1618</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>succ</td>
<td>Succinate</td>
<td>CO0042</td>
<td>C4H6O4</td>
<td>6</td>
<td>-2</td>
<td>4</td>
<td>-162.96</td>
<td>118.088</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>xul5p</td>
<td>Xylose 5-phosphate</td>
<td>CO0331</td>
<td>C5H11O8P</td>
<td>11</td>
<td>-1</td>
<td>10</td>
<td>-393.5</td>
<td>230.1098</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviation**

- pyruvate exchange
- 3-phosphoglycerate kinase
- pyruvate kinase

**Biochemical Reactions**

1. **Glycolysis**
   - **R1** PDH: glucose-6-phosphate dehydrogenase
   - **R2** PFK: 6-phosphofructokinase
   - **R3** FBP: fructose-1,6-bisphosphatase
   - **R4** FBA: fructose-bisphosphatase aldolase
   - **R5** TPI: triose-phosphate isomerase
   - **R6** GAPD: 3-phosphoglycerate dehydrogenase
   - **R7** PGK: 3-phosphoglycerate kinase
   - **R8** FGM: phosphorylase murate
   - **R9** ENO: enolase
   - **R10** PK: pyruvate kinase

2. **Pyruvate Metabolism**
   - **R11** LDH: D-lactate dehydrogenase
   - **R12** PPL: pyruvate formate lyase
   - **R13** FDH3: formate dehydrogenase
   - **R14** PTV: acetate phosphotransferase
   - **R15** ACX: acetate kinase
   - **R16** ACP: acylphosphatase-phosphatidylphosphatase
   - **R17** POF: pyruvate dehydrogenase complex
   - **R18** MD2: malic enzyme
   - **R19** OXAD: oxaloacetate decarboxylase
   - **R20** ACOZD: alcohol dehydrogenase

3. **TCA Cycle**
   - **R21** PKOX: phosphoenolpyruvate carboxylase
   - **R22** MDH: malate dehydrogenase
   - **R23** FUM: fumarase
   - **R24** RFP: fumarate reductase

4. **Pentose Phosphate Pathway**
   - **R25** GDPD: glucose-6-phosphate dehydrogenase
   - **R26** FGL: 6-phosphogluconolactonase
   - **R27** G6PD: glucose-6-phosphate dehydrogenase
   - **R28** RPI: ribose-5-phosphate isomerase
   - **R29** RBP: ribulose-5-phosphate 3-epimerase
   - **R30** BPE: ribulose-5-phosphate 3-epimerase
   - **R31** TKT: transketolase
   - **R32** TAL: transaldolase
   - **R33** TKT2: transketolase

5. **Glycolysis**
   - **R34** GCK: glycerol kinase
   - **R35** PGDO: glycerol-3-phosphate dehydrogenase

6. **NEOGLYCONE AND NEOGLYCONE METABOLISM**
   - **R36** THO2: nadi2:transhydrogenase

7. **Exchange**
   - **R37** EX:pyr(py): pyruvate exchange
   - **R38** EX:cel(py): pyruvate exchange
   - **R39** EX:cel(cyd): cytochrome c exchange
   - **R40** EX:sgl(py): glucose exchange
   - **R41** EX:sgl(cyd): cytochrome c exchange
   - **R42** EX:for(cyd): forrate exchange
   - **R43** EX:act(ac): acetate exchange
   - **R44** EX:act(eft): forrate exchange
   - **R45** EX:act(s): succinate exchange
   - **R46** EX:act(sum): fumarate exchange
   - **R47** EX:act(egl): glutamate exchange
   - **R48** EX:act(ech): oxaloacetate exchange
   - **R49** EX:act(h2o): water exchange

8. **Consumption of metabolites in other reactions not considered**
   - **R50** ATN: ATP requirement for other reactions
   - **R51** NADH: NADH requirement for other reactions
   - **R52** NADPH: NADPH requirement for other reactions

9. **Biomass**
   - **R53** BCKA: biomass

---

Supporting References

Chapter 5
Crowding Conditions in the Intracellular Diffusion Process

5.1. Preamble

The cell is a complex system, composed by macromolecules, organelles, ions, solutes, and other solid components immersed in a fluid phase, called cytoplasm. The interactions among the cellular components determine the behaviour of the cell.

The study of these interactions allows the understanding of the metabolism, and its further application in the industry as valuable tool to manipulate the microorganisms in order to increase the profitability of the process.

The prokaryotic cells, such as bacteria, have been widely used in fermentation processes due to its high population growth rate, and that the excreted products have to cross only one membrane instead of two or more, which is more favourable for large-scale production. For example, in the succinate producers Aspergillus spp., Fusarium spp., Penicillium simplicissimum and Saccharomyces cerevisae the succinic acid has to cross mitochondrial and cytoplasmic membrane in order be excreted (Beuprez et al., 2010).

In prokaryotes, the diffusion is the primary mean of molecules’ motion within the cell, so it could be decisive for the encounter of two or more reactants, and therefore affecting the possibility that a reaction is carried out. For this reason, the modelling and simulation of the diffusion process can help to the understanding of the metabolism.

As discussed in the previous chapters, the crowding conditions affect the thermodynamic activity of the molecules (metabolites, enzymes, etc.), which could
change the thermodynamic equilibrium of the reactions (as shown in Chapter 4), and also (with an opposite effect) increase reaction rate and decrease the diffusion of the molecules.

The intracellular diffusion process can be hindered by the presence of solid components, which decrease the probability to find a free space where a test molecule can move as higher is the cell volume occupied by crowders.

As was reviewed in Chapter 2, one of the most popular methods for the diffusion-simulation that satisfies the restriction of the molecules’ impenetrability is the Monte Carlo algorithm. However, these simulations can become computationally expensive for systems of large domain size, long time simulations, and/or large number of molecules, e.g. when all the intracellular molecules (metabolites and enzymes) are considered.

Other alternative is the use of Lattice Boltzmann Method (LBM) that allows faster simulations, but considering point-like molecules, so that the crowding effect is not considered and it could over-predict the diffusion.

The objective of this chapter is to propose an alternative for quick simulations of the molecules’ diffusion in 2D crowded media. For this, a methodology that couples LBM and SPT is presented in Publication 3. As in Chapter 4, SPT is used to estimate the activity coefficient, which is also defined as the inverse of the probability to find free space in the system analysed. The derivation of the SPT equation for 2D is analogous to that shown in the Appendix of this thesis for 3D.

Selected examples are used to show the potential of this crowding-adaptation of LBM for the simulation of the diffusion in crowded systems. The diffusion coefficients, parameters required for the simulation were chosen to be of the order of magnitude those reported for proteins in the cytoplasm (Elowitz et al., 1998). As in the previous chapters, the results and discussion of this work, as well as the validation of the results of this crowding-adaptation of LBM with kinetic Monte Carlo algorithm are presented in the following publication.
5.2. Publication 3

A Lattice-Boltzmann scheme for the simulation of diffusion in intracellular crowded systems

Liliana Angeles-Martinez\textsuperscript{1}

Email: liliana.angelesmartinez@postgrad.manchester.ac.uk

Constantinos Theodoropoulos\textsuperscript{1*}

*Corresponding author

Email: k.theodoropoulos@manchester.ac.uk

\textsuperscript{1}School of Chemical Engineering and Analytical Science, University of Manchester, Manchester, M13 9PL, UK
Abstract

Background

The intracellular environment is a complex medium whose solid components are molecules whose volume can occupy up to 40% of the total volume of the cell. In such crowding conditions the diffusion of metabolites, enzymes, and other molecules in the medium can be decreased. One of the most popular methodologies for the simulation of diffusion in crowding systems is the Monte Carlo algorithm (MC) which tracks the movement of each particle. This can, however, be computationally expensive for a system comprising a large number of molecules. On the other hand, the Lattice Boltzmann Method (LBM) tracks the movement of collections of molecules, which represents significant savings in computational time. Nevertheless in the classical manifestation of such scheme the crowding conditions are neglected.

Results

In this paper we use Scaled Particle Theory (SPT) to approximate the probability to find free space for the displacement of hard-disk molecules and in this way to incorporate the crowding effect to LBM. The results indicate that LBM over-predicts the diffusion in 2D crowded systems, while the proposed coupled SPT-LBM predicts the same behaviour as an in-house developed kinetic Monte Carlo (kMC) algorithm (which here is used as our “computational experiment”, for validation purposes) but with a significantly reduced computational effort. Despite the fact that small deviations between the two methods were observed, in part due to the mesoscopic and microscopic nature of each method, respectively, the agreement was satisfactory both from a qualitative and a quantitative point of view.
Conclusions

A crowding-adaptation to LBM has been developed using SPT, allowing fast simulations of diffusion-systems of different size hard-disk molecules in two-dimensional space. This methodology takes into account crowding conditions; not only the volume fraction occupied by the crowder molecules but also the influence of the size of the crowder which can affect the displacement of molecules across the lattice system.

Keywords

Lattice Boltzmann Method, Scaled Particle Theory, crowding conditions, diffusion systems.

Background

Several microorganisms are used in the conversion of raw materials to added-value products, for example Actinobacillus succinogenes has been used for the synthesis of succinic acid from crude bio refinery glycerol [1,2]. The analysis and simulation of the factors affecting the metabolism of these organisms allow the further identification of the strategies needed for its manipulation in order to increase the formation of the metabolite of interest over other by-products.

As it is known the environmental conditions and the properties of the medium play an important role in the metabolism. The intracellular reactions are carried out in a complex, heterogeneous, and crowded medium composed by solid components.
(macromolecules, enzymes, etc.) in a fluid phase called cytoplasm, where for prokaryotes the diffusion is the primary mean of intracellular motion.

According to a drawing proposed by Goodsell [3], if the cytoplasm of *Escherichia coli* is divided into 600 cubes of (100 nm)$^3$, an average of 130 glycolytic enzymes and 100 from the Krebbs cycle are present in each cube in addition to the metabolites and other compounds, which all together comprise a very large number molecules for the simulation of the intracellular environment. Henceforth, we will use the terms *molecule* and *particle* interchangeably to refer to the intracellular enzymes and/or the metabolites.

The solid components of the cell occupy about 40% of the total volume [4], and due to the impossibility that two molecules occupy the same space at the same time (steric effects), these *crowding* conditions affect the intracellular process in two opposite ways: 1) decreasing the diffusion of the molecules [5], and 2) increasing the thermodynamic activity of the metabolites [4], and therefore enhancing the reaction rate, and modifying the thermodynamic feasibility of some reactions. A review of the crowding effects can be found in Chebotareva et al. [6].

Several methodologies at different levels of detail are available for the simulation of molecular diffusion systems. One of the most popular is the Monte Carlo (MC) algorithm [7–13] (microscopic level) where each molecule is tracked during its journey through the cell so the restriction of the impenetrability of the molecules is satisfied in a straightforward way.
MC is a powerful technique and easy to implement, however it is limited to short simulation times, restricted lattice/domain sizes, and/or reduced number of molecules because of the large computational costs. Besides, due to the stochastic nature of MC, it requires several simulations to smoothen the noise of the results by computing average quantities. Moreover, in most cases the molecules are considered to be of the same size, so the size effect could be hidden [7,8].

Alternatively, Lattice Boltzmann (LBM) [14] is a mesoscopic method which allows the efficient simulation of the dynamics of collections of across a defined lattice according to expressions that conserve mass and momentum [15]. Here, the solute transport is simulated either 1) treating the solute as another fluid and solving a multicomponent problem (active solute component) or 2) assuming that the molecules do not have velocity of their own so they are advected by the fluid (passive solute component). In both cases the volume of the solute is neglected. See Sukop and Thorne [16] for a review.

Alternatively, in particle suspensions simulations the motion of each molecule is described by a hard sphere model (with the drawback of being computationally expensive for large numbers of particles) while the fluid is described by LBM [17,18]. This is similar to other hybrid methods used for example to follow the enzymes’ motion with MC, or the tumour growth with Cellular Automata [19], while the passive transport of the metabolites and the fluid is simulated by LBM.
Since LBM computes the evolution of the average molecules’ density, it represents a
good alternative to simulate the diffusion of large number of intracellular metabolites.
However, since classical LBM does not take into account the volume of the molecules,
and therefore the effect of obstacles on the molecules’ diffusion, it may overestimate
the degree of mixing of the system analysed.

The displacement of a molecule depends on the probability $P$ to find enough empty
space to move at every step of its journey. Scaled Particle Theory (SPT) is a method
that allows the estimation of this probability $P$ which is a function of the radii and
concentration/number of molecules present in the system. SPT also has been used to
investigate the effect of macromolecular crowding on solvation [20], thermodynamic
activity of proteins [21] and of metabolites [Angeles-Martinez and Theodoropoulos,
unpublished].

The aim of this paper is to incorporate the crowding effect on the LBM simulation of
the particles’ diffusion. For this, a methodology is proposed for coupling SPT and
LBM, allowing in this way faster simulations for systems with a large number of
molecules of different size under crowded conditions, such as the intracellular
environment. Here, we consider the metabolites/enzymes as passive solute components
assuming that the fluid (cytoplasm) is at rest within the cell. For validation purposes the
results are compared with those obtained from an in-house developed kinetic Monte
Carlo (kMC) algorithm [11-13].
Methods

In the classical LBM, the system is represented by a regular lattice, where the molecules located at the same site or node at time $t$ may interact with each other (collision step), and then according to a set of rules, some particles move to one of their neighbouring lattice sites (known as the streaming step), where they will interact with molecules from other nodes at time $t + \Delta t$ and so on.

The methodology we propose here corrects the average number of molecules that enter a neighbouring lattice site, taking into account crowding effects and it can be summarised as follows:

1. Solve the classical LBM to find the number of molecules that will try to move into the $d$ direction at time $t + \Delta t$ ($F_d^{LB}$).

2. Use $F_d^{LB}$ to estimate the corrected distribution values $F_d$ by solving the explicit formulation (see below) constrained by the size and number of the molecules, as well as the size of voxels or sites in which the lattice is divided.

3. Use the $F_d$ values obtained in (2) for the streaming step (in the same way as in the classical LBM), and go back to point (1).

Here we use the term crowding-Lattice Boltzmann Method (cLBM) to distinguish this proposed methodology from the classical LBM, which in principle considers volumeless molecules.
crowding-Lattice Boltzmann Method (cLBM)

1. Lattice Boltzmann Method

For comparison purposes with the kMC algorithm, in this paper the D2Q5 scheme [19] (Figure 1B) is implemented, consisting of 5 possible directions that the molecules can move, in a 2D lattice. The lattice is divided in squares of $\Delta x$ [nm] side, called “voxels”, whose position is identified by the index $(i, j)$ (Figure 1A).

The evolution of the distribution function of the species $sp$, $F_{d,sp}$, in the diffusion system is given by the discrete Boltzmann equation [15]

$$F_{d,sp}(i_{new}, j_{new}, t + \Delta t) = F_{d,sp}(i, j, t) + \Omega_{d,sp}^{diff}(i, j, t)$$

(1)

The superscript LB is used to distinguish $F_{d,sp}^{LB}$ that is calculated from the classical LBM from the crowding-corrected value $F_{d,sp}$. Both $F_{d,sp}^{LB}$ and $F_{d,sp}$ represent the concentration or number of molecules of the species $sp$ in a voxel that try to move to a neighbouring site, so they are given in [molecules per voxel]. Other concentration units e.g. [mol per voxel] can be used, but dimensional changes in other variables are required for consistency.

We use the BGK approximation to estimate the non-reactive collision term $\Omega_{d,sp}^{diff}$, which is given by [15]

$$\Omega_{d,sp}^{diff}(i, j, t) = \omega_{sp} \left[ F_{d,sp}^{eq}(i, j, t) - F_{d,sp}(i, j, t) \right]$$

(2)
Assuming that the fluid within the cell is at rest, the equilibrium distribution function $F_{d,sp}^{eq}$ takes the form [19]

$$F_{d,sp}^{eq}(i, j, t) = \frac{\rho_{sp}(i, j, t)}{m_{sp}}w_d$$  \hspace{1cm} (3)

where the weight factor $w_d$ is 0 for $d=0$, and 1/4 for $d=1, 2, 3, 4$, while $m_{sp}$ is the mass of one single molecule of type $sp$. The macroscopic density of species $sp$, $\rho_{sp}$, is expressed as:

$$\rho_{sp}(i, j, t) = m_{sp}\sum_{d} F_{d,sp}(i, j, t)$$  \hspace{1cm} (4)

The expression for the relaxation parameter $\omega_{sp}$ (indicated in Eq. (2)) can be deduced using the Enskog-Chapman procedure, and is given by [22]

$$\omega_{sp} = \frac{2}{1 + 4D_{sp}^0 \frac{\Delta t}{\Delta x^2}}$$  \hspace{1cm} (5)

Due to the fact the BGK model was formulated for non-crowded systems [15] $D_{sp}^0$ takes the value of the diffusion coefficient for diluted solutions. Here, $D_{sp}^0$ is considered independent of the position $(i,j)$.

According to Fick’s first law the diffusive flux of volumeless molecules $(J)$ from one region to another is proportional to the gradient of the concentration. However, when
the volume of the molecules is important and/or the solution is not considered diluted, 

\[ J \] should be proportional to gradient of the activities [23].

Since the molecules in a defined system occupy a volume in space, not all the system’s volume is available to the centre of mass of a test molecule [4].

The activity \( a \) is a term that describes the number of molecules per available volume, while the concentration \( C \) is the number of molecules per total volume [4]. Both variables are related by Eq. (6), where the activity coefficient \( \gamma \) indicates deviations from an ideal solution.

\[ a_{sp} = \gamma_{sp} \frac{C_{sp}}{C_{sp}^{\alpha}} \]  \hspace{1cm} (6)

Since \( a_{sp} \) is a dimensionless variable, for the purpose of dimensional consistency in this paper the standard concentration of the species \( sp \) (\( C_{sp}^{\alpha} \)) is considered equal to 1 molecule per voxel, i.e. diluted solutions where there is no crowding influence. Hence, the \( a_{sp} \) value is not affected by \( C_{sp}^{\alpha} \), therefore Eq. (6) is simplified to \( a_{sp} = \gamma_{sp} C_{sp} \).

\( \gamma \) is defined as the ratio of the total volume (or area in 2D systems) divided by the available volume (or area in 2D), in other words, it is inversely proportional to the probability to find free space in the system.
From the above and for the lattice-system analysed here, the flux of the molecules \( J_{sp} \) of the species \( sp \) in the \( x \) direction from voxel \((i,j)\) to \((i,j+1)\) is defined in Eq. (7). The reverse flux is \( J_{sp(i,j+1\rightarrow i,j)} = -J_{sp(i,j\rightarrow i,j+1)} \).

\[
J_{sp(i,j\rightarrow i,j+1)} = \frac{D^{sp}}{\Delta x} \left[ a_{sp(i,j\rightarrow i,j+1)}(i, j, t) - a_{sp(i,j+1\rightarrow i,j)}(i, j + 1, t) \right]
\]  

(7)

where the concentration involved in the term \( a_{sp(i,j\rightarrow i,j+1)} \) is proportional to the number of molecules (per voxel) trying to move from \((i,j)\) to \((i,j+1)\) at time \( t \), i.e. \( F^{LB}_{L,sp}(i, j, t) \), while \( a_{sp(i,j+1\rightarrow i,j)} \) is proportional to \( F^{LB}_{3,sp}(i, j + 1, t) \).

Substituting Eq. (1) in Eq. (2), and using as concentration the values \( F^{LB}_{L,sp}(i, j, t) \) and \( F^{LB}_{3,sp}(i, j + 1, t) \) yields:

\[
J_{sp(i,j+1\rightarrow i,j)} = \frac{D^{sp}}{\Delta x} \left[ \gamma_{sp}(i, j, t) F^{LB}_{L,sp}(i, j, t) - \gamma_{sp}(i, j + 1, t) F^{LB}_{3,sp}(i, j + 1, t) \right]
\]  

(8)

The use of a constant \( D^{sp} \) independent of the concentration could be questionable for crowded and inhomogeneous systems since the presence of background/crowder molecules hinders the movement of a test molecule.

Due to the steric effects, the presence of background molecules leads to a reduction of the available space where the test molecule can move due to Brownian motion. Therefore the probability to find free space next to the test molecule also decreases.
Considering that the rate of Brownian displacement (or diffusion) is a function of the work required to free the target space from background molecules ($\Delta W$), Muramatsu and Minton [24] proposed the relation

\[
\ln \left( \frac{D_{sp}}{D_{sp}^0} \right) = -\frac{\Delta W}{k_B T} \tag{9}
\]

where $D_{sp}^0$ is the diffusion coefficient of species $sp$ in diluted solutions. $\Delta W$ depends on the size and shape of the space required. The probability of observing the spontaneous formation of such free space as result of a fluctuation is [25, 26]

\[
P_{sp} = \exp \frac{\Delta W}{k_B T} \tag{10}
\]

where $k_B$ is the Boltzmann constant, and $T$ is the temperature. Assuming a well-mixed system (or subsystem), the probability $P_{sp}$ is defined as the available volume divided by the total volume, i.e. the inverse of the activity coefficient

\[
P_{sp} = \frac{1}{\gamma_{sp}} \tag{11}
\]

Because the diffusion coefficient $D_{sp}$ is a function of the available volume in the target voxel where the test molecule is trying to move in (and therefore dependent of the position $(i, j)$), it cannot be factored as in Eq. (8), so

\[
J_{sp(i, j \rightarrow i+1, j)} = \frac{1}{\Delta t} \left[ \gamma_{sp}(i, j, t)F^L_{1,sp}(i, j, t)D_{sp}(i, j + 1, t) - \gamma_{sp}(i, j + 1, t)F^L_{3,sp}(i, j + 1, t)D_{sp}(i, j, t) \right] \tag{12}
\]
Substituting Eq. (9) and (10) in Eq. (12), the flux \( J_{sp(i \rightarrow j \rightarrow i+1)} \) can be rewritten as:

\[
J_{sp(i \rightarrow j \rightarrow i+1)} = \frac{D^0}{\Delta x} \left[ \gamma_{qp}(i, j, t) F_{1,qp}^{LB}(i, j, t) P_{sp}(i, j+1, t) - \gamma_{qp}(i, j+1, t) F_{3,qp}^{LB}(i, j+1, t) P_{sp}(i, j, t) \right]
\]  

(13)

where \( P_{sp} \) is the probability of species \( sp \) to find available space in the target voxel. Eq. (13) looks like the Theorell formula for non-perfect systems [23].

In order to conserve the driving force, the corrected values \( F_{1,sp}(i, j, t) \) and \( F_{3,sp}(i, j+1, t) \) should also be related to the flux \( J_{sp(i \rightarrow j \rightarrow i+1)} \), so that

\[
J_{sp(i \rightarrow j \rightarrow i+1)} = \frac{D^0}{\Delta x} \left[ \gamma_{qp}(i, j, t) F_{1,sp}(i, j, t) - \gamma_{qp}(i, j+1, t) F_{3,sp}(i, j+1, t) \right]
\]  

(14)

Comparing Eq. (13) and Eq. (14), the new \( F_{1,sp}(i, j, t) \) and \( F_{3,sp}(i, j+1, t) \) values are found as

\[
F_{1,sp}(i, j, t) = F_{1,sp}^{LB}(i, j, t) P_{sp}(i, j+1, t)
\]  

(15)

\[
F_{3,sp}(i, j+1, t) = F_{3,sp}^{LB}(i, j+1, t) P_{sp}(i, j, t)
\]  

(16)

Generalizing, the corrected values \( F_{d,sp} \) \((d=1,2,3,4)\) are calculated as

\[
F_{d,sp}(i, j, t) = F_{d,sp}^{LB}(i, j, t) P_{sp}(i_{next}, j_{next}, t)
\]  

when \( d = 1,2,3,4 \)

Or their equivalent:
The molecules that could not move into their corresponding target voxel \((i_{\text{next}},j_{\text{next}})\), i.e. the voxel next to \((i,j)\) in the direction \(d\), will remain in the current \((i,j)\). Therefore in order to conserve mass \(F_{0,sp}\) becomes

\[
F_{0,sp}(i,j,t) = \frac{F_{\text{int}}^{\Delta}(i,j,t)}{\gamma_{sp}(i_{\text{next}},j_{\text{next}},t)}
\]

(18)

The activity coefficient \(\gamma_{sp}(i,j,t)\) is a function of the size and shape of the molecules present in the voxel/site \((i,j,t)\).

In this paper, the Scale Particle Theory \([25,26]\) is used to approximate \(\gamma_{sp}\) in a mixture of (non-overlapping) hard disk-molecules (Eq. (20) and (21)) which could be of different radii \(r\). For this we assume that the volume and temperature of the subsystems/voxels are held constant. In order to simplify the notation, the position-time dependence \((i,j,t)\) of the variables \(\gamma_{sp}\) and \(S_{s}\) has been dropped from Eq. (20) and (21).

\[
\ln \gamma_{sp} = -\ln(1-S_{s}) + \left[ \frac{2S_{s}}{1-S_{s}} \right] r_{sp} + \left[ \frac{S_{s}}{1-S_{s}} + \frac{S_{s}^2}{(1-S_{s})} \right] r_{sp}^2
\]

(20)

where \(S_{s} (0 \leq x \leq 2)\) is given by

\[
S_{s} = \frac{\pi}{\Delta x^2} \sum_{i=1}^{l} \rho_{i}(r_{i})\]

(21)
Note that if all the particles simulated are considered point-like molecules \((r \rightarrow 0, S_2 \rightarrow 0)\), and/or they are in a non-crowded system \((S \rightarrow 0)\), then \(\ln(\gamma_{sp}) \rightarrow -\ln(1)\) in Eq. (20) so that the probability \(P_{sp}\) (Eq. (11)) would be equal to 1, therefore \(F_{d,op}\) becomes equal to \(F_{d,op}^{LB}\) (Eq. (17)).

The use of \(P_{sp}\) (or its equivalent \(\gamma_{sp}^{-1}\)) in Eq. (17) restricts the maximum number of molecules in a voxel, and also avoids the exchange of positions between molecules. When two molecules moving in opposite directions are adjacent and have volume (or area in 2D systems), one acts as an obstacle for the other, so exchanging positions between them (Figure 2B) should be prohibited. In other words the displacement of a molecule is limited by the probability to find empty space \(P_{sp}\).

3. Streaming step

The new \(F_{d,op}\) values are used in the streaming step in the same way as in LBM (Eq. (22)–(26) below). Then the scheme goes back to step 1 and the procedure is repeated until the simulation end time is reached.

\[
\begin{align*}
F_{1,op}(i, j+1,t+\Delta t) &= F_{1,op}(i, j,t) \\
F_{2,op}(i-1, j,t+\Delta t) &= F_{2,op}(i, j,t) \\
F_{3,op}(i, j-1,t+\Delta t) &= F_{3,op}(i, j,t) \\
F_{4,op}(i+1, j,t+\Delta t) &= F_{4,op}(i, j,t) \\
F_{d,op}(i, j,t+\Delta t) &= F_{d,op}(i, j,t)
\end{align*}
\]
In order to validate the crowding-adaptation of LBM presented above, the simulation of a diffusion system was carried out by an in-house developed kMC algorithm on a lattice [11–13], whose results are used as our computational experiments. A brief description of the kMC algorithm is given below.

**Lattice Kinetic Monte Carlo (kMC) algorithm**

In the lattice kMC [11] algorithm each site of the lattice can be occupied by at most one molecule. Each molecule can move to the one of the 4 neighbouring sites (top, bottom, right, or left), as long as they are free, i.e. there is no other molecule in the target site.

The basic idea of the procedure is the following:

1) Identify the classes of species, i.e. the combinations of adjacent species that can react or diffuse (if there is an empty site next to a molecule).

2) The rates (including the diffusion rate) of all the possible events or processes are listed and their cumulative value is computed.

3) An event is probabilistically chosen, e.g. the next molecule to move and the site where it will take place, using a random number and the cumulative value of the rates listed in point 2.

4) The diffusion (or reaction) of the chosen molecule(s) takes place.

5) A variable time step value is also probabilistically estimated using another random number and the cumulative value of the rates.

6) The number of species and classes in a region and/or in the whole lattice is updated as well as the time.
7) Go back to point 1 until the end simulation time is reached.

See [11–13] for detailed information of the algorithm and the corresponding equations for each step of the process.

Results and discussion

For the diffusion examples presented in this section the following assumptions have been made:

1) The system analysed consists of a square lattice of $(1000 \text{ nm})^2$, divided in equal-sized voxels or sites of $\Delta x$ per side whose value is indicated in each example.

2) Each voxel is well-mixed.

3) The fluid is considered as a continuum and it is at rest.

4) The system has periodic boundaries.

5) The mass of all species are assumed equal to $1 \text{ g}$ per molecule, therefore the value of $\rho_{sp}$ is equal to the number of molecules, i.e. $\sum_{d} F_{d,sp}$. Notice that $m_{sp}$ can take other values if it is required.

The methodologies LBM and cLBM were programed in MATLAB R2011a (The MathWorks, Natick, MA), while the lattice kMC was implemented in Fortran 90.
Validation of cLBM: a lattice-model

For validation purposes, we consider a lattice model (Figure 2A) for the diffusion of molecules here represented by non-rotating squares of \((10 \text{ nm})^2\) which have a square uniform packing order. Taking into account that the system’s size is \((1000 \text{ nm})^2\) the maximum number of molecules that can fit inside is 10 000.

This type of system is consistent with the ones commonly used in kMC simulations where a molecule can move to one of its neighbouring sites at every time step.

According to the \(\Delta x\) value used in the cLBM simulations, for comparison purposes, the average results obtained (after 1000 repetitions) at every time step \(t\) by kMC were coarse grained in lattice regions equivalent to voxels of side length equal to \(\Delta x\).

Due to the square uniform packing order of the same size molecules in this lattice model the activity coefficient \(\gamma_{sp}\) (which is required in cLBM) given by Eq. (20) and (21) can be simplified to

\[
\ln \gamma_{sp} = -\ln(1 - S_2) \tag{27}
\]

\[
S_2 = \frac{1}{\Delta x^2} \sum_{i=1}^{m} \rho_{i} A_{i} \tag{28}
\]

where \(A_{i}\) is the area of a molecule of species \(i\).

The diffusion of two types of molecules has been simulated using the parameters shown in Example 1 of Table 1. For this, the lattice is divided into 10 vertical regions, where
the voxels (of $\Delta x = 100$ nm) of the first column of the lattice have been filled with 100 molecules of type A and B as indicated in the Figure 3A.

Once the molecules begin to diffuse, it is possible to compare their movement summing the number of molecules A (used as tracer molecule) in each vertical region of the lattice at different times. Figure 3B shows that LBM and cLBM predict the same diffusion profile, which in turn is close to the results estimated by kMC. This is because there are no obstacles in the horizontal direction and therefore cLBM results are the same as the ones from LBM. The same happens when the volume (or area in 2D) of the molecules are zero (data not shown).

The relative error of the results obtained by cLBM and/or LBM compared with kMC is calculated as the Frobenius norm of the difference of the matrixes containing the number of molecules $sp$ in each voxel predicted by kMC ($\rho_{kMC}^{sp}$) and cLBM ($\rho_{cLBM}^{sp}$), or by kMC and LBM ($\rho_{LBM}^{sp}$), divided by the total number of molecules $sp$ simulated, i.e.

$$\text{error}_{sp} = \frac{\text{norm}(\rho_{kMC}^{sp} - \rho_{cLBM}^{sp})}{\sum_i \sum_j \rho_{kMC}^{sp}(i, j)} \times 100 = \frac{\sum_i \sum_j (\rho_{kMC}^{sp}(i, j) - \rho_{cLBM}^{sp}(i, j))^2}{\sum_i \sum_j \rho_{kMC}^{sp}(i, j)} \times 100$$ (29)

Since $m_{sp}$ of all $sp$ is assumed to be equal to 1 g per molecule, we keep the term $\rho_{sp}$ to represent the number of molecules in the system.
A comparison of the $\text{norm}(\rho_{kMC} - \rho_{cLBM})$ and $\text{norm}(\rho_{kMC} - \rho_{LBM})$ for the diffusion of species A in the Example 1 indicates that cLBM predictions are more accurate than those of LBM (Figure 3C). For example, the relative error estimated by cLBM at time 30ms is $\text{error}_A = 0.93\%$ (which indicates that from 100 molecules A simulated, cLBM predicts that 0.93 molecules are allocated in different voxel’s positions than those predicted by kMC, in other words there are deviations in the distribution of 0.93 molecules A from the 100 simulated in this example), while the error from LBM amounts to $\text{error}_A = 16.39\%$.

This is because LBM considers point-like molecules so that they will always find enough space to fit in the target voxel, in other words there are no restrictions about the molecules’ exchange positions (Figure 2B), causing LBM to over-predict diffusion in the vertical direction.

Nevertheless, if we assume that both molecules A and B are of the same type then the relative error between kMC and LBM decreases, becoming equal to that between kMC and cLBM (Figure 3D). This means that LBM is a good alternative for quick simulations of one type of metabolites/enzymes, but when two or more species are present then it can over predict the system’s degree of mixing.

As was pointed out by Li et al. [28], the relaxation parameter value affects the accuracy of LBM. According to Eq. (5) and the chosen parameters $\Delta x$ and $\Delta t$ in Example 1, both $\omega_A$ and $\omega_B$ were estimated equal to one. In order to show the dependency
between the error and \( \omega_A \), we tested different \( \Delta t \) values, maintaining constant \( \Delta x = 50 \) nm, which corresponds to \( \omega_A \) ranging from 0.5 to 1. The results indicate that \( \omega_A = 1 \) gives the lowest error \( \text{error}_A \) (Figure 4A), therefore in the following we use \( \Delta x \) and \( \Delta t \) values that allow setting \( \omega_{sp} \) close to one.

Moreover, since cLBM is formulated as an explicit method that requires information of the neighbouring voxels at time \( t \) to estimate \( \rho_{sp} \) of a voxel \((i,j)\) for the next time \( t + \Delta t \), decreasing \( \Delta t \) also diminishes the error predicted between kMC and cLBM for Example 1 as shown in Figure 4B, where 3 different sets of coefficient \( \Delta t \) and the corresponding \( \Delta x \) that keeps constant \( \omega_{sp} = 1 \) were tested.

In all cases, the LB methods required \( n \) iterations, i.e. \( n\Delta t \) steps, before reaching a quasi-stable error. Therefore the smaller the chosen \( \Delta t \) (and the corresponding \( \Delta x \)) the faster the system reached that state. Since good results were obtained with \( \Delta t = 1.25 \) ms and \( \Delta x = 50 \) nm (where a voxel allows to fit a maximum of 25 molecules) compared with the more accurate but slower \( \Delta t = 0.2 \) ms and \( \Delta x = 20 \) nm, in the following we use \( \Delta x = 50 \) nm.

In the previous example the crowding conditions are in the vertical direction so that reduced displacement of species A is expected in this direction. In order to prove this, a second example is proposed for the diffusion of three types of molecules initially allocated in the region indicated in Figure 5A (the parameters are shown in Example 2 of Table 1).
For the purpose of this example, a region consists of two columns of 20 voxels with \( \Delta x = 50 \text{ nm} \), where each voxel of region 4 (Figure 5A) is filled with 25 molecules of species A. Region 5 and 6 are filled with equal numbers of species B and C, respectively. The results of the diffusion of species A show that cLBM predicts the same behaviour as kMC as can be seen in Figure 5B.

On the other hand, LBM predicts a symmetric movement of A despite the fact that species B acts as an obstacle in its way (Figure 5B). The same symmetric profile is obtained for species B and C (Figure 5C and D).

A comparison of the relative error between kMC–cLBM and kMC–LBM is shown in Figure 6. As was also observed in Example 1, the error computed with the cLBM results is smaller than that computed through the LBM results. In fact, at time 30 ms the errors given by cLBM (\( error_A = 1.103\% \), \( error_B = 0.833\% \), \( error_C = 1.172\% \)) represent deviations in no more than 2\% of the total number of molecules simulated for each species.

Additionally, a comparison of the mean squared displacement of a tracer molecule, i.e. the displacement from one voxel to another, was carried out under different crowding conditions (the parameters are given in Example 3 of Table 1). The total number of tracer molecules represents 1\% of the lattice area of the 2D system simulated by cLBM and LBM.
The results show (Figure 7) that only cLBM is sensitive to increments in the concentration of the background/crowder molecules, which is reflected in a reduction of the displacement of the tracer particle.

This latter comparison cannot be extended to kMC due to the fact that LB methods are unable to identify and quantify the motion of the molecules that remain in the same voxel at time $t + \Delta t$, i.e. only the “effective” displacement made when the molecules pass from one to another voxel, but not how many movements have to be executed before entering to the next voxel as is the case with lattice kMC.

For this reason, the mean squared displacement predicted by kMC could be higher than the LBM methods, and therefore cLBM cannot be used directly in the parameter estimation of anomalous diffusion as done with the results from Monte Carlo algorithm by Vilaseca et al. [7,8]. However, this loss in information is balanced by a significant reduction of the execution time required for the simulation.

Despite the fact that kMC and cLBM were implemented in different computing languages (Fortran and Matlab, respectively) a comparison of the execution time for the diffusion Example 1 ($t_{kMC} = 103.05$ hr vs $t_{cLBM} = 7.13$ s), and Example 2 ($t_{kMC} = 139.72$ hr vs $t_{cLBM} = 18.31$ s) reveals the potential use of cLBM for faster simulations of larger systems. Notice that the execution time $t_{kMC}$ covers 1000 simulation repetitions required to obtain the average values reported for $\rho$. 
Different size molecules in cLBM

Up to this point we have analysed a lattice-model were only molecules of the same size can be considered, however as was pointed out by Vilaseca et al. [7,8], the size (and also the shape) of the molecules could be an important parameter in their movement or diffusion.

In order to study the influence of the size of the molecules on the diffusion process we simulate the motion of a tracer molecule in a 2D crowding system composed by 5 types of background particles (all of them of circular shape with different radii and concentrations) which are randomly located and all together occupy 30% of the total lattice space (the corresponding parameters are given in Example 4 of Table 1).

According to the SPT assumption of a well-mixed voxel, here the molecules can be anywhere unlike the previous lattice model where only a square uniform packing order is allowed.

Figure 8A shows the displacement across the lattice predicted by cLBM for a tracer molecule of different radius magnitude whose total concentration (i.e. the total number of molecules in the lattice) is 796 molecules. As it can be seen (Figure 8) more movements from one voxel to another were detected when a small radius is assumed. This is because the probability of the tracer species to move to the target voxel increases inasmuch as the second and third term of Eq. (20) disappear when \( r_{\text{tracer}} = 0 \) nm. In other words, the available space for point-like molecules equals the free space in
the target voxel, i.e. the total area not occupied by other molecules represented by the term $1 - S_2$ in Eq. (20).

Moreover, the comparison of the displacement of the tracer molecule with $r_{tracerr} = 0$ nm obtained by cLBM (broken pink line in Figure 8A) and that estimated by the classical LBM for the same Example 4 (broken pink line in Figure 8B) indicates that even if the metabolites are considered as point-like particles they are still affected by the molecular crowding conditions, unless all the molecules simulated in the system are also point-like molecules.

Finally, the effect of the size of the crowder molecules on diffusion was investigated. For this, 3 different diffusion simulations of a system composed of tracer molecules of $r_{tracerr} = 1.5$ nm which occupies 1% of the lattice area, and the crowder molecules having different radii: 2, 1.5, and 1nm, representing 30% of the total lattice space were performed.

Since the area covered by the crowder is kept constant despite the fact that the radius is modified, more particles of $r_{crowder} = 1$ nm are simulated compared with those used if $r_{crowder}$ is 1.5 or 2 nm. A comparison of mean squared displacement of a tracer molecule (Figure 9) under such conditions indicates that the particles’ motion decreases when the number of crowder molecules increases.
This suggests that not only the area-fraction occupied by crowders is important but also the way it is covered, i.e. the size and number of molecules. Similar behaviour was found by Vilaseca et al. [7] for the diffusion simulation of molecules having different sizes using a Monte Carlo algorithm.

**Conclusions**

The LBM predicts with great accuracy the diffusion of particles under ideal conditions, i.e. considering point-like molecules, and/or non-crowding systems, and/or when only one type of molecules is simulated. However, if conditions change, for example, a system involving more than two species at crowding conditions, LBM predictions increasingly deviate from our lattice kMC-based computational experiment.

Although small discrepancies were found between the cLBM and kMC results (differences that were expected due to the level of detail inherent in each method), the proposed crowding adaptation of LBM is able to predict the same behaviour in the species diffusion profile. This suggests that the coupled SPT-LBM can be considered as a computational alternative for fast simulations of diffusion systems with a large number of molecules of different size and/or for long times, under crowding conditions at a fraction of the computational cost compared to a molecular (microscopic) method such as kMC.

Nevertheless, as in other mesoscopic methods, the saving in the execution time is accompanied by a reduction in the information that cLBM can provide compared with that obtained from a microscopic method. For example, the quantification of the total
displacement of the molecules on time for the parameters estimation of anomalous
diffusion [7,8].

The accuracy of cLBM is influenced by the chosen voxel size $\Delta x$ and time increment
$\Delta t$, both related by the relaxation parameter $\omega_{sp}$. It was found that the use of $\omega_{sp}$
values close to one gives better results than any other in the range $0 \leq \omega_{sp} \leq 1$.
Moreover, cLBM being an explicit method, small values of $\Delta t$ (maintaining $\omega_{sp} = 1$,
which involves reducing $\Delta x$) also reduce the error between the proposed methodology
(cLBM) and kMC.

Regarding the influence of the size of the particles on the diffusion process, a reduction
in the mean squared displacement of a tracer molecule when its size is increased was
observed, as well as when the size of the crowders is decreased (but maintaining
constant the lattice fraction occupied by them). Hence, the incorporation of small
molecules, e.g. metabolites, in the simulation system can affect the diffusion profile
predicted for macromolecules.

Even though cLBM requires the species’ concentration of the neighbour voxels at time
$t$ to compute the results at $t + \Delta t$, therefore the LBM’s local feature is lost, the
correction for the crowding effects is external to the estimation of LBM distributions,
i.e. $F_{a,sp}^{LB}$, so that alternative LBM schemes can potentially be implemented within
cLBM, e.g. for the simulation of reaction–diffusion systems.
Abbreviations

LBM, Lattice Boltzmann Method; cLBM, crowding adaptation to Lattice Boltzmann Method; kMC, kinetic Monte Carlo algorithm; MC, Monte Carlo algorithm.

Nomenclature

$\alpha_{sp}$ Activity of the species $sp$ [dimensionless].

$A_{sp}$ Area of a molecule of the species $sp$ [nm$^2$].

$C_{sp}$ Concentration of the species $sp$ [molecules nm$^{-2}$].

$C^*_sp$ Standard concentration of the species $sp$ [molecules nm$^{-2}$].

d Direction chosen by the molecules to jump to neighboring voxel [dimensionless].

$D_{sp}$ Diffusion coefficient [nm$^2$ ms$^{-1}$].

$D^0_{sp}$ Diffusion coefficient in dilute solutions [nm$^2$ ms$^{-1}$].

$error_{sp}$ Relative error of molecules’ distribution predicted by kMC–cLBM or kMC–cLBM of the species $sp$ [%].

$F_{d,sp}$ Distribution function of the species $sp$ in the direction $d$ predicted by cLBM [molecules per voxel].

$F_{eq}^{sp}_{d,sp}$ Equilibrium distribution function of the species $sp$ in the direction $d$ [molecules per voxel].

$F_{LB}^{sp}_{d,sp}$ Distribution function of the species $sp$ in the direction $d$ predicted by LBM [molecules per voxel].

$i$ Index that identify the position of a voxel [dimensionless].
Index that identify the position of a voxel [dimensionless].

$J_{sp}$ Diffusive flux of molecules $sp$ in 2D [molecules nm$^{-1}$ ms$^{-1}$].

$m_{sp}$ Mass of a molecule $sp$ [g molecule$^{-1}$].

$next$ Subscript that indicates the target voxel where the molecules will move in the $t+\Delta t$ [dimensionless].

$P_{sp}$ Probability to find available space for species $sp$ in the target voxel [dimensionless].

$r_{sp}$ Radii of the species $sp$ [nm].

$k_B$ Boltzmann constant equivalent to $1.3806 \times 10^{-23}$ [J K$^{-1}$].

$sp$ Index that identify the molecule species $sp$ [dimensionless].

$t$ Time [s].

$T$ Temperature of the medium [K].

$w_{ld}$ Weight factor for the calculation of the equilibrium function [dimensionless].

$\Delta t$ Time increment [s].

$\Delta W$ Work required to free the target space from background molecules [J molecule$^{-1}$].

$\Delta x$ Size of the voxel in which the lattice is divided [nm].

$\gamma_{sp}$ Activity coefficient of a molecule $sp$ [dimensionless].

$\Omega_{s,sp}$ Non-reactive collision term [molecules per voxel].

$\omega_{sp}$ Relaxation parameter [dimensionless].

$\rho_{sp}$ Macroscopic density of species $sp$ in a voxel [g per voxel].

$\rho_{sp}$ Matrix with the macroscopic density of species $sp$ in all voxels of the lattice [g voxel].
Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

LAM developed the crowding-adaptation to LBM, performed the diffusion-simulations, and drafted the manuscript. CT supervised the development of the methodology, reviewed and edited the manuscript.

Acknowledgements

The financial support CONACYT-Mexico for the PhD of LAM is gratefully acknowledged.

LAM and CT would like to thank Prof. Alexander Gorban for useful discussions as well as his suggestions in the development of the crowding- LBM methodology.

References


Figures

Figure 1. (A) Lattice scheme and (B) D2Q5 scheme.

Figure 2 Lattice model and space restrictions. (A) Lattice model of square uniform packing order used for the validation of cLBM, molecules are represented by dashed squares. (B) Volume- (or area- in 2D) restriction on the movement of the molecules associated to the impediment of the position exchange between them.

Figure 3 Diffusion example 1. (A) Initial location of a system composed by of two types of molecules. (B) Diffusion profile of the test molecules A predicted by LBM, cLBM, and kMC at times 10, 20, 30, 40 and 50 ms. (C) Relative error of the distribution of molecules A estimated by cLBM–kMC and LBM–kMC. (D) Relative error predicted assuming that all the molecules (A and B) are of the same type. The parameters used in the simulation are indicated in Table 1.

Figure 4 Relative error of the molecules distribution estimated by cLBM–kMC for Example 1. (A) Different relaxation parameter $\omega_A$ were tested but maintaining constant $\Delta x = 50$ nm. (B) Different $\Delta x$ and $\Delta t$ values were tested maintaining $\omega_A = 1$.

Figure 5 Diffusion example 2. (A) Initial conditions of the diffusion of three types of molecules. Diffusion profile of the species (B) A, (C) B, and (D) C predicted by LBM, cLBM, and kMC at times 10, 20, 30, 40, 50 ms.
Figure 6 Relative error of the molecules distribution estimated by kMC–cLBM and kMC–LBM for Example 2.

Figure 7 Diffusion example 3. Mean squared displacement of tracer molecules representing 1% of the lattice area at different concentration of crowder molecules: 10, 20, 30, and 40% of the lattice area, predicted by (A) cLBM and (B) LBM. The parameters used in the simulation are indicated in Table 1.

Figure 8 Diffusion example 4. Mean squared displacement estimated by (A) cLBM and (B) LBM for hard-disk tracer molecules of different radii: 2, or 1.5, or 1 or 0 nm, with total concentration equal to 796 molecules per (1000 nm)$^2$, in a crowded media composed by 4 types of molecules which together occupied 30% of the lattice area. The parameters used in the simulation are indicated in Table 1.

Figure 9 Diffusion example 5. Mean squared displacement of a tracer molecule in 3 different systems where the 30% of the total lattice space is occupied by the same type of crowder molecules, but having different radii: 23 873 molecules of $r_{\text{crowder}} = 2$ nm, or 42 441 molecules of $r_{\text{crowder}} = 1.5$ nm, or 95 493 molecules of $r_{\text{crowder}} = 1$ nm). The radii and concentration of the tracer are 1.5 nm and 1 414 molecules, respectively. The parameters used in the simulation are indicated in Table 1.
### Table 1 Parameters used in the diffusion examples

<table>
<thead>
<tr>
<th>Example 1</th>
<th>Example 2</th>
<th>Example 3</th>
<th>Example 4</th>
<th>Example 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta x = 100$</td>
<td>$\Delta x = 50$</td>
<td>$\Delta x = 50$</td>
<td>$\Delta x = 50$</td>
<td>$\Delta x = 50$</td>
</tr>
<tr>
<td>$\Delta t = 5$</td>
<td>$\Delta t = 0.625$</td>
<td>$\Delta t = 0.625$</td>
<td>$\Delta t = 0.625$</td>
<td>$\Delta t = 0.625$</td>
</tr>
<tr>
<td>$D_A = 500$</td>
<td>$D_A = 1\ 000$</td>
<td>$D_{tracer} = 1\ 000$</td>
<td>$D_{tracer} = 1\ 000$</td>
<td>$D_{tracer} = 1\ 000$</td>
</tr>
<tr>
<td>$A_A = 100$</td>
<td>$A_A = 100$</td>
<td>$A_{tracer} = 100$</td>
<td>$r_{tracer} = 2, \text{ or } 1.5, \text{ or } r_{tracer} = 1.5$</td>
<td>$C_{tracer} = 1\ 414$</td>
</tr>
<tr>
<td>$C_A = 100$</td>
<td>$C_A = 1\ 000$</td>
<td>$C_{tracer} = 100$</td>
<td>$C_{tracer} = 796$</td>
<td>$D_{crowder} = 1\ 000$</td>
</tr>
<tr>
<td>$D_B = 500$</td>
<td>$D_B = 1\ 050$</td>
<td>$D_{crowder} = 1\ 000$</td>
<td>$D_{crowder1} = 1\ 050$</td>
<td>$r_{crowder} = 2$</td>
</tr>
<tr>
<td>$A_B = 100$</td>
<td>$A_B = 100$</td>
<td>$A_{crowder} = 100$</td>
<td>$r_{crowder1} = 1.9$</td>
<td>$C_{crowder} = 23\ 873$</td>
</tr>
<tr>
<td>$C_B = 900$</td>
<td>$C_B = 1\ 000$</td>
<td>$C_{crowder} = 1\ 000$</td>
<td>$D_{crowder2} = 1\ 100$</td>
<td>$r_{crowder} = 1.5$</td>
</tr>
<tr>
<td>$D_C = 1\ 100$</td>
<td>or</td>
<td>$C_{crowder} = 6\ 172$</td>
<td>$D_{crowder2} = 1\ 100$</td>
<td>$r_{crowder2} = 1.8$</td>
</tr>
<tr>
<td>$A_C = 100$</td>
<td>$C_{crowder} = 2\ 000$</td>
<td>$r_{crowder2} = 1.8$</td>
<td>$C_{crowder} = 42\ 441$</td>
<td>$C_{crowder} = 42\ 441$</td>
</tr>
<tr>
<td>$C_C = 1\ 000$</td>
<td>or</td>
<td>$C_{crowder} = 9\ 824$</td>
<td>or</td>
<td>$C_{crowder} = 42\ 441$</td>
</tr>
<tr>
<td>$C_{crowder} = 3\ 000$</td>
<td>or</td>
<td>$D_{crowder3} = 1\ 200$</td>
<td>$r_{crowder3} = 1.6$</td>
<td>$C_{crowder} = 95\ 493$</td>
</tr>
<tr>
<td>$C_{crowder} = 4\ 000$</td>
<td></td>
<td>$r_{crowder3} = 1.6$</td>
<td>$C_{crowder} = 123\ 174$</td>
<td></td>
</tr>
</tbody>
</table>

The parameters indicated above are expressed in units: $\Delta x [=\ \text{nm}]$, $\Delta t [=\ \text{ms}]$, $D_{sp} [=\ \text{nm}^2\ \text{ms}^{-1}]$, $A_{sp} [=\ \text{nm}^2]$, $C_{sp} [=\ \text{molecules per (1 000 nm)}^2]$. 
The order of magnitude of diffusion coefficients used agrees with that reported by Elowitz et al. [27].
Figure 1. (A) Lattice scheme and (B) D2Q5 scheme.
Figure 2 Lattice model and space restrictions. (A) Lattice model of square uniform packing order used for the validation of cLBM, molecules are represented by dashed squares. (B) Volume- (or area- in 2D) restriction on the movement of the molecules associated to the impediment of the position exchange between them.
Figure 3 Diffusion example 1. (A) Initial location of a system composed by of two types of molecules. (B) Diffusion profile of the test molecules A predicted by LBM, cLBM, and kMC at times 10, 20, 30, 40 and 50 ms. (C) Relative error of the distribution of molecules A estimated by cLBM–kMC and LBM–kMC. (D) Relative error predicted assuming that all the molecules (A and B) are of the same type. The parameters used in the simulation are indicated in Table 1.
Figure 4 Relative error of the molecules distribution estimated by cLBM–kMC for Example 1. (A) Different relaxation parameter \( \omega_A \) were tested but maintaining constant \( \Delta x = 50 \text{ nm} \). (B) Different \( \Delta x \) and \( \Delta t \) values were tested maintaining \( \omega_A = 1 \).
Figure 5 Diffusion example 2. (A) Initial conditions of the diffusion of three types of molecules. Diffusion profile of the species (B) A, (C) B, and (D) C predicted by LBM, cLBM, and kMC at times 10, 20, 30, 40, 50 ms.
Figure 6 Relative error of the molecules distribution estimated by kMC–cLBM and kMC–LBM for Example 2.
Figure 7 Diffusion example 3. Mean squared displacement of tracer molecules representing 1% of the lattice area at different concentration of crowder molecules: 10, 20, 30, and 40% of the lattice area, predicted by (A) cLBM and (B) LBM. The parameters used in the simulation are indicated in Table 1.
Figure 8 Diffusion example 4. Mean squared displacement estimated by (A) cLBM and (B) LBM for hard-disk tracer molecules of different radii: 2, or 1.5, or 1 or 0 nm, with total concentration equal to 796 molecules per (1000 nm)$^2$, in a crowded media composed by 4 types of molecules which together occupied 30% of the lattice area. The parameters used in the simulation are indicated in Table 1.
Figure 9 Diffusion example 5. Mean squared displacement of a tracer molecule in 3 different systems where the 30% of the total lattice space is occupied by the same type of crowder molecules, but having different radii: 23 873 molecules of $r_{\text{crowder}} = 2 \text{ nm}$, or 42 441 molecules of $r_{\text{crowder}} = 1.5 \text{ nm}$, or 95 493 molecules of $r_{\text{crowder}} = 1 \text{ nm}$). The radii and concentration of the tracer are 1.5 nm and 1 414 molecules, respectively. The parameters used in the simulation are indicated in Table 1.
Chapter 6
Conclusions and Future Work

6.1. Conclusions

The use of living organisms for the conversion of raw materials into goods for the benefit of mankind dates back to ancient times, e.g. in the preparation of fermented beverages. Early industrial applications of microorganisms were based on the empirical knowledge of the process accumulated over generations. Nowadays, many tools are available for the improvement and control of processes, or even the design of new ones. One of these tools is the modelling and simulation of the cellular metabolism.

The metabolism consists of a series of reactions connected by the metabolites, i.e. the products of reaction 1 are the reactants of reaction 2, and so on. These intracellular reactions, responsible for the chemical transformation of a substrate to a final product, are carried out in a crowded and heterogeneous medium composed by solid components immersed in a fluid phase (called cytoplasm). The interactions among these cellular components determine the cell behaviour.

The modelling and simulation of the metabolism, the interactions among the intracellular components, and the factors affecting positively or negatively the production of a metabolite of interest, provide the background for the successful optimisation of the processes from the biological point of view, which could be reflected in a higher production yield, lower recovery cost, revaluation of by-products, alternatives for reduction and/or disposal of industrial waste, etc.

Due to the complexity of intracellular processes and the uncertainty in the biological information, e.g. kinetic parameters, the full mathematical representation of the cell behaviour is not possible at this moment. The current models focus only on some
aspects of the cellular processes depending on the information available for the system to be analysed. Chapter 2 presents a review of the type of models available in the literature.

As mentioned in Chapter 1, the aim of this thesis is to develop computational tools for the study and simulation of intracellular processes, specifically for the estimation of metabolic fluxes and the diffusion in crowded systems.

The measurement and/or estimation of the flux distribution of the metabolic network is the first step to determine an effective strategy (genetic, type of culture, environmental conditions, type of substrate, etc.) to maximise the production of the metabolite of interest and minimise the by-products ones. The calculation of metabolic fluxes allows the identification of the reactions with more positive/negative influence in the synthesis of a metabolite of interest.

Stoichiometric-based models have become one of the most popular tools for the estimation of the flux distribution due to their simplicity, and because kinetic parameters, such as reaction rate constants, is not essential for these models. Other restrictions related to the thermodynamics, experimental fluxes, regulation, among others, can be added to the system in order to increase the accuracy of the predictions.

In Chapter 3 of this thesis, a thermodynamically-constrained stoichiometric method was presented for the estimation of the range of fluxes, metabolites’ concentration and Gibbs free energy of the reactions under different environmental conditions. The methodology is formulated as a nonlinear optimisation problem that simultaneously solves the mass balance and the thermodynamic constraints.

A common assumption made in previous methods available in the literature (e.g. Beard et al., 2005; Henry et al., 2007; Hoppe et al., 2007) is that a zero flux trough a reaction indicates that this reaction is blocked so that its corresponding \( \Delta_G \) can take any value. The \( \Delta_G \) values of the metabolic reactions are dependent on each other, inasmuch as they are all linked by the metabolites’ concentration. The fact that \( \Delta_G \) of one or more
reactions can take any value “relaxes” the thermodynamic constraints, so the $\Delta_r G$ value (and the corresponding flux) attainable for the other metabolic reactions is affected.

Unlike these methods, the methodology presented in Chapter 3 is based on the premise that if the enzyme has been experimentally detected, then the estimation of a zero flux indicates that the reaction is at equilibrium ($\Delta_r G = 0$). For this reason, this stoichiometric model restricts the attainable flux incorporating the parameters related to the presence of the enzymes that catalyse the reactions.

The combined use of the experimental detection of the enzymes and the metabolites’ concentration under specific growth conditions can help to solve the uncertainty of whether a reaction $i$ is blocked or at equilibrium when the flux through it is predicted to be zero.

An interesting metabolic system for the biodiesel industry, namely the carbon central metabolic pathways of *Actinobacillus succinogenes* for the production of succinate from glycerol (a by-product in the manufacture of biodiesel) was used as case study.

The comparison of the range of fluxes estimated for the metabolic network analysed assuming that all the enzymes in the network are detected actives (*equilibrium* case) and those obtained considering that a zero flux indicates the absence of the enzyme (*block* case) shows that the *equilibrium* case predicts narrower range of fluxes than the *block* one. The results obtained in this case study highlight the importance of the use of complementary knowledge to overcome the arising uncertainties (e.g. in the thermodynamic data) and lack of information in order to compute more accurate predictions. In other words, the enzymes’ presence information is not essential but complementary for the calculation of thermodynamically feasible flux distributions.

Perhaps the presence of an enzyme, even if the net flux of the reaction that catalyses is zero, is because it helps to maintain a particular crowding condition (as will be discussed below) or the enzyme is related to the formation of a complex such as a
metabolon.

The impact of the environmental conditions $I$, $pH$, and $T$ on the estimation of the flux distributions was also evaluated. Results suggest that the ionic strength $I$, followed by the temperature $T$, have more influence on the range of fluxes predicted for this metabolic system under the conditions tested (Chapter 3).

However, the use of a very wide default range of metabolites’ concentration when the exact values are not available (as in this case study), can hide the actual impact of the environmental conditions.

Note that the simplifying assumption of using the protonated predominant species to represent the thermodynamic properties of the metabolites can also hide the effect of the environmental conditions (particularly $I$). The consequence of this assumption is that different range of fluxes, compared with those estimated using the pseudoisomer group properties (as was done in this work), can be predicted. Therefore, results from considering this assumption should be treated carefully.

The effects of the intracellular crowding conditions on the thermodynamic feasibility of the reactions require a special mention. As in previous thermodynamically-constrained stoichiometric methods available in literature, the model presented in Chapter 3 assumes that all the cellular components are volumeless molecules. However, they can occupy up to 40% of the total cell volume, which causes a reduction in the available space for the motion of molecules and the reactions among them. This reduction in the available space increases the activity of the components, i.e. metabolites, proteins, enzymes, nucleic acids, etc. The activity of a species is defined as the amount of molecules of that species divided by the available space.

Since $\Delta_G$ is a function of the activity of the metabolites (only when the crowding effects are negligible the activity equals the value of the concentration), changes in the volume fraction occupied by the solid components also alter the thermodynamic feasibility of the pathways. Extending the work of Chapter 3, a new methodology based on the Scaled Particle Theory is proposed in Chapter 4 to incorporate the crowding
effect on the thermodynamic analysis of metabolic networks, and the estimation of the (thermodynamically-) feasible fluxes. To the best of my knowledge, there is no a similar thermodynamically-constrained stoichiometric method in the literature. The lactic acid pathway and the central metabolism of *A. succinogenes* were used as case studies.

The studies on the thermodynamic feasibility of the lactic acid pathway indicate that the crowding conditions given by the presence of background molecules may determine the direction of a reaction. Moreover, these crowding conditions may convert an infeasible pathway to a feasible one. These results suggest that it is important to maintain regions more crowded than others within the cell, in order to favour certain pathways that could be predicted infeasible for the available metabolite concentration (i.e. number of molecules per total cell volume) and/or at different crowding condition.

A comparison of the effect of the environmental conditions on the range of fluxes predicted by the thermodynamically-constrained stoichiometric model (for the metabolic network of *A. succinogenes*) indicates that the flux is more sensitive to the crowding conditions than to *I* and *pH*. In fact, the range of fluxes estimated is wider and more reactions were predicted as reversible when the intracellular crowding conditions are taken into account. These results suggest that the crowded medium may contribute to the robustness of the network to face genetic or environmental changes.

Notwithstanding, the uncertainties in the thermodynamic data used in the thermodynamic constraints may mask the influence of the environmental conditions (crowding, *pH*, *I* and *T*) on the rage of fluxes, predicting wider and similar ranges for the conditions tested.

While the intracellular crowding conditions enhance the activity of the metabolites, with the corresponding effect on the thermodynamic feasibility of pathways and the potential increment in the reaction rate (according to the law of mass action), the crowding also affects the diffusion process. The presence of background molecules hinders the motion of a test molecule due to the reduction of the available space.
Monte Carlo algorithm has been widely used to simulate the diffusion in crowded media. However, it could become computationally expensive to simulate every single intracellular molecule (including metabolites and enzymes) and/or for long time periods (considering, for example, that the doubling time by binary fission in bacteria range from 30 min to days).

In order to overcome these drawbacks, an alternative methodology for fast diffusion-simulation of different size hard-disk molecules in 2D crowded systems was proposed in Chapter 5. This methodology consists of a crowding-adaptation of the Lattice Boltzmann Method, which in principle considers point-like molecules. For this, the system is divided in regular size voxels, where the hard-disk molecules of voxel \((i,j)\) will try jump to one of the nearest voxels. However, the number of molecules that effectively enter in the target voxel \((i_{\text{next}},j_{\text{next}})\) depends on the probability to find available space to fit these incoming molecules. SPT was used to estimate this probability.

The crowding-adaptation of the Lattice Boltzmann Method, designed here cLBM to distinguish form the classical LBM, was validated with a kinetic Monte Carlo algorithm (kMC). Results of the diffusion examples simulated indicate that the classical LBM over-predicts the diffusion of the molecules in crowded media, while cLBM predicts the same behaviour than kMC but in a significantly shorter execution time. Small discrepancies were found between cLBM and kMC, which are related to the mesoscopic and microscopic nature of both methods, respectively. Nonetheless, the results from cLBM and kMC agree satisfactory (both qualitative and quantitative).

However, the considerable saving in the execution time is accompanied by a reduction in the information provided by cLBM compared with from kMC. For example, cLBM is incapable of identifying and quantifying the movements of the molecules within a voxel, only the “effective jumps” from one voxel to another, so that the cLBM results of the displacement of the molecules cannot be used for the parameter estimation of anomalous.

One of the advantages of cLBM is that it allows the simulation of different size of
molecules. The cLBM results indicate a dependency between the mean squared displacement (from one voxel to another) of the molecules and their size. Specifically, it was observed 1) an increase of the displacement of a hard-disk test molecule when its radius is decreased, and 2) a reduction of its displacement when, keeping constant the volume fraction occupied by the crowders, the size of the background molecules is decreased, this indicates the potential effect of the metabolites’ presence (or other small molecules) on the diffusion of macromolecules.

6.2. Future work

The complete modelling and simulation of the intracellular processes and the complex interactions among cellular components the processes are far from being achieved at present. This thesis focuses on 3 methodologies that incorporate specific aspects of the intracellular environmental conditions that could affect the intracellular processes. However, further studies are required to understand and predict the cell behaviour with improved accuracy. Some recommendations for future work are given bellow.

The thermodynamic data are essential for the thermodynamic analysis (and its coupling with the stoichiometric models, see Chapter 3 and 4) of a metabolic network. Therefore, the use of more accurate information (from experiments or estimated by a Group Contribution Method) such as the standard Gibbs free energy, the enthalpy of formation of the metabolites, and the intracellular metabolites’ concentration, would allow predicting narrower range of fluxes and $\Delta G$ for the reactions involved in network analysed under certain environmental conditions.

In this thesis, it was observed that the range of fluxes predicted by the thermodynamically stoichiometric method were more sensitive to changes in the intracellular crowding conditions and the ionic strength (Chapter 4), so that the appropriate model that couples the effect of electrostatic and excluded volume interactions would help to achieve a better understanding of the intracellular processes. For this, one option is to vary the radii of the molecules used for the SPT, so that the repulsion/attraction among the molecules (given by the electrostatic interactions) is
taken into account in the maximum distance between two centers of mass, and therefore in the available volume for the particular molecule. This idea can also be applied to the diffusion-simulation of hard-spheres (3D) or hard-disk (2D) molecules in crowded media.

Regarding the crowding-adaptation of LBM (Chapter 5), this methodology was formulated for the diffusion in 2D systems. The methodology can be easily extended to 3D, and to include the reaction term allowing fast reaction-diffusion simulations in crowded media.

cLBM can be modified to take into account the molecules’ movement to specific regions or voxels, where they have more affinity for the molecules present in those regions. This would enable the simulation of the formation of hyperstructures inside the cell.


Energy Information Administration (US).
Available in: http://www.eia.gov/
Date accessed: August 1st, 2014.


European Biodiesel Board.
Date accessed: August 4th, 2014.

Available in:
Date accessed: October 24th, 2014.


Minton, A.P. 2001a. The influence of macromolecular crowding and macromolecular


Provost, A., and Bastin, G. 2004. Dynamic metabolic modeling under the balanced


The Scaled Particle Theory (Reiss et al., 1959; Lebowitz et al., 1965) allows the estimation of the work $\Delta W(r)$ required to insert a test hard-core sphere of radius $r$ in a mixture (or a fluid) of background hard-core spheres of radius $r_s$. In this section the mathematical formulation of the SPT is presented for systems of one type of molecule species in 3D, but its derivation is analogous for mixtures of molecules and/or in 2D.

The probability of observing the spontaneous formation of a spherical cavity of radius $r$ (where the test molecule can fit in) at an arbitrary place within the (one component-) mixture as a result of a fluctuation is

$$P_o(r) = \exp \left( \frac{\Delta W(r)}{k_BT} \right)$$  \hspace{1cm}  \text{Eq. A1}$$

where $k_B$ is the Boltzmann constant and $T$ is the temperature. This cavity has to be free of any molecule (or part of it), i.e. a spherical region of radius $r + r_s$ should be free of the centre of mass of any background molecule. The probability to find a background molecule in a sphere of radius $r + r_s$ is given by

$$P_b(r) = \frac{4}{3} \rho_s \pi (r + r_s)^3 = 1 - P_o(r)$$  \hspace{1cm}  \text{Eq. A2}$$

where $\rho_s$ is the density of the background molecules, i.e. number of molecules per liter. Substituting Eq. A1 in Eq. A2 gives

$$\Delta W(r) = -k_B T \ln \left( 1 - \frac{4}{3} \rho_s \pi (r + r_s)^3 \right)$$  \hspace{1cm}  \text{Eq. A3}$$

The work required to insert a small molecule ($r$ is close to zero) can be approximated
by the Taylor expansion series of Eq. A3 up to second order\(^1\), while for large molecules is the work for the expansion of a point-like molecules up to a volume \(\frac{4}{3} \pi r^3\) at constant pressure \(P\). Therefore the work is given by

\[
\Delta W(r) = \Delta W(0) + \left( \frac{d\Delta W}{dr} \right)_{r=0} r + \frac{1}{2} \left( \frac{d^2\Delta W}{dr^2} \right)_{r=0} r^2 + P \frac{4}{3} \pi r^3
\]

Eq. A4

After solving and evaluating the derivatives, Eq. A4 becomes

\[
\frac{1}{k_s T} \Delta W(r) = -\ln(1 - S) + \frac{4 \pi \rho_s r^2}{1 - S} r + \frac{1}{2} \left( \frac{8 \pi \rho_s r^2}{1 - S} + \frac{(4 \pi \rho_s r^2)^2}{(1 - S)^2} \right) r^2 + \frac{P}{k_s T} \frac{4}{3} \pi r^3
\]

Eq. A5

where \(S\) is

\[
S = \frac{4}{3} \pi \rho_s r^3
\]

Eq. A6

Assuming that \(r = r_s\) then Eq. A5 is reduced to

\[
\frac{1}{k_s T} \Delta W(r_s) = -\ln(1 - S) + \frac{6 S}{1 - S} + \frac{9 S^2}{2(1 - S)^2} + \frac{P}{k_s T} S
\]

Eq. A7

In order to determine the pressure \(P\) (of Eq. A7) the following thermodynamic relation is used

\[
\frac{\partial P}{\partial \rho_s} = \rho_s \frac{\partial \mu_s}{\partial \rho_s}
\]

Eq. A8

where the chemical potential of the background molecules \(\mu_s\) (or any other species) is defined as

\[
\mu_s = k_s T \ln \left( \frac{\rho_s h^3}{(2\pi m k_s T)^{3/2}} \right) + W(r_s)
\]

Eq. A9

\(^1\) For the derivation of the SPT equation in 2D systems, the equation of the work (analogous to Eq. A3) was expanded only up to first order.
\( h \) is the Planck’s constant, and \( m_b \) is the mass of a background molecule. Substituting Eq. A9 and A7 in Eq. A8 and integrating \( P \) with respect to \( \rho \), gives

\[
\frac{P}{k_B T \rho} = \frac{1 + S_3^3 + S_3^2}{(1 - S_3)^3} \tag{Eq. A10}
\]

Finally, the probability to find available space to insert a test molecule of \( r = r_o \) is obtained by substituting Eq. A7 and A10 in Eq. A1

\[
\ln(P_o(r_o)) = -\ln(1 - S_3) + \frac{6S_3^3}{1 - S_3} + \frac{9S_3^2}{2(1 - S_3)^2} + \frac{S_3 + S_3^2 + S_3^3}{(1 - S_3)^3} \tag{Eq. A11}
\]

Since \( P_o(r_o) \) is the inverse of the activity coefficient \( \gamma_{r_o} \), then Eq. A11 also can be expressed as

\[
\ln(\gamma_{r_o}) = -\ln(1 - S_3) + \frac{6S_3^3}{1 - S_3} + \frac{9S_3^2}{2(1 - S_3)^2} + \frac{S_3 + S_3^2 + S_3^3}{(1 - S_3)^3} \tag{Eq. A12}
\]