The following section is an addendum that supplements the original thesis.

# Appendix E

## Analysis of simulation results for the model of the γ-butyrolactone system

### Defining the criteria for the comparison with the transcriptomics data

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Quantification formula</th>
<th>Value (μ)</th>
<th>Uncertainty (σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{\text{max}}$</td>
<td>Time at peak (minutes)</td>
<td>1320</td>
<td>150</td>
</tr>
<tr>
<td>Relative gain of $scbR$</td>
<td>$[R_{\text{peak}}] - [R_{\text{min}^a}]$</td>
<td>4.4</td>
<td>1.51</td>
</tr>
<tr>
<td>Relative gain of $scbA$</td>
<td>$[A_{\text{peak}}] - [A_{\text{min}^a}]$</td>
<td>5.27</td>
<td>1.678</td>
</tr>
<tr>
<td>Relative loss of $scbR$</td>
<td>$\frac{[R_{\text{min}^b}] - [R_{\text{min}^a}]}{[R_{\text{peak}}] - [R_{\text{min}^a}]}$</td>
<td>0.251</td>
<td>0.13</td>
</tr>
<tr>
<td>Relative loss of $scbA$</td>
<td>$\frac{[A_{\text{peak}}] - [A_{\text{min}^b}]}{[A_{\text{peak}}] - [A_{\text{min}^a}]}$</td>
<td>4.567</td>
<td>0.433</td>
</tr>
<tr>
<td>Slope of increase of $scbR$</td>
<td>$\frac{[R_{\text{peak}}] - [R_{\text{min}^a}]}{\text{Time}<em>{R</em>{\text{peak}}} - \text{Time}<em>{R</em>{\text{min}^a}}}$</td>
<td>2.51</td>
<td>1.565</td>
</tr>
<tr>
<td>Slope of increase of $scbA$</td>
<td>$\frac{[A_{\text{peak}}] - [A_{\text{min}^a}]}{\text{Time}<em>{A</em>{\text{peak}}} - \text{Time}<em>{A</em>{\text{min}^a}}}$</td>
<td>3.13</td>
<td>1.835</td>
</tr>
<tr>
<td>Slope of decrease of $scbR$</td>
<td>$\frac{[R_{\text{peak}}] + [R_{\text{min}^b}]}{2}$</td>
<td>1.144</td>
<td>0.653</td>
</tr>
<tr>
<td>Slope of decrease of $scbA$</td>
<td>$\frac{[A_{\text{min}^b}] - [A_{\text{peak}}]}{\text{Time}<em>{A</em>{\text{min}^b}} - \text{Time}<em>{A</em>{\text{peak}}}}$</td>
<td>-2.748</td>
<td>0.681</td>
</tr>
</tbody>
</table>

$R_{\text{min}^a}$: minimum concentration of $scbR$ before the peak  
$A_{\text{min}^a}$: minimum concentration of $scbA$ before the peak  
$R_{\text{min}^b}$: minimum concentration of $scbR$ after the peak  
$A_{\text{min}^b}$: minimum concentration of $scbA$ after the peak
Appendix E. Analysis of the GBL System Model

Complete log-likelihood profiles

Figure E.1: Complete log-likelihood profiles \( (scbA, scbR \text{ and total}) \) for the 4 scenarios of the GBL model, with promoters having equal strength \( (\chi = 1) \). Scenarios B (antisense RNA interaction) and D (combination of all scenarios) show a clear predominance over the other two.
Appendix E. Analysis of the GBL System Model

(a) *scbR* complete log-likelihood profiles for the 4 scenarios of the GBL model.

(b) *scbA* complete log-likelihood profiles for the 4 scenarios of the GBL model.

(c) Total complete log-likelihood profiles for the 4 scenarios of the GBL model.

*Figure E.2:* Complete log-likelihood profiles (*scbA*, *scbR* and total) for the 4 scenarios of the GBL model, with *scbA* promoter being stronger ($\chi = 0.1 - 0.9$). Scenarios B (antisense RNA interaction) and D (combination of all scenarios) have been significantly reduced, due to the change in the promoter strength. Scenarios A and C did not seem affected by this change.
Appendix E. Analysis of the GBL System Model

(a) scbR complete log-likelihood profiles for the 4 scenarios of the GBL model.

(b) scbA complete log-likelihood profiles for the 4 scenarios of the GBL model.

(c) Total complete log-likelihood profiles for the 4 scenarios of the GBL model.

Figure E.3: Complete log-likelihood profiles (scbA, scbR and total) for the 4 scenarios of the GBL model, with scbR promoter being stronger ($\chi = 1 - 10$). Scenarios B (antisense RNA interaction) and D (combination of all scenarios) have been increased, due to the change in the promoter strength. Scenarios A and C were not affected.
## K-S test results

**Table E.2: Results from K-S test for the simulations with $\chi = 1 (k_{FR} = k_{FA})$**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d1}$</td>
<td></td>
<td>0</td>
<td>0.0983</td>
<td>0</td>
<td>0.5303</td>
<td>0.0222</td>
<td>0</td>
<td>0.0074</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{1-}$</td>
<td></td>
<td>0</td>
<td>0.2467</td>
<td>0</td>
<td>0.1007</td>
<td>0.2459</td>
<td>0</td>
<td>0.1262</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d2}$</td>
<td></td>
<td>0</td>
<td>0.0398</td>
<td>0</td>
<td>0.0553</td>
<td>0.0144</td>
<td>0</td>
<td>0.281</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{2-}$</td>
<td></td>
<td>0</td>
<td>0.0217</td>
<td>0</td>
<td>0.0527</td>
<td>0.1779</td>
<td>0</td>
<td>0.2200</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d3}^*$</td>
<td></td>
<td>0</td>
<td>0.0012</td>
<td>1</td>
<td>0.0002</td>
<td>0.0845</td>
<td>0</td>
<td>0.1112</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{3-}^*$</td>
<td></td>
<td>1</td>
<td>0.0003</td>
<td>1</td>
<td>0.0001</td>
<td>0.2370</td>
<td>0</td>
<td>0.2583</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d4}$</td>
<td></td>
<td>0</td>
<td>0.2252</td>
<td>0</td>
<td>0.4188</td>
<td>0.4337</td>
<td>0</td>
<td>0.7083</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{4-}$</td>
<td></td>
<td>0</td>
<td>0.1781</td>
<td>0</td>
<td>0.2206</td>
<td>0.9766</td>
<td>0</td>
<td>0.9132</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d5}^*$</td>
<td></td>
<td>0</td>
<td>0.9927</td>
<td>0</td>
<td>0.9522</td>
<td>0.0445</td>
<td>0</td>
<td>0.0435</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{5-}^*$</td>
<td></td>
<td>0</td>
<td>0.3688</td>
<td>0</td>
<td>0.3474</td>
<td>0.5182</td>
<td>0</td>
<td>0.5617</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d6}$</td>
<td></td>
<td>0</td>
<td>0.9290</td>
<td>0</td>
<td>0.4490</td>
<td>0.0230</td>
<td>0</td>
<td>0.146</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{6-}$</td>
<td></td>
<td>0</td>
<td>0.5102</td>
<td>0</td>
<td>0.5803</td>
<td>0.9796</td>
<td>0</td>
<td>0.9551</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d7}$</td>
<td></td>
<td>0</td>
<td>0.5472</td>
<td>0</td>
<td>0.2135</td>
<td>0.0549</td>
<td>0</td>
<td>0.0466</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{7-}$</td>
<td></td>
<td>0</td>
<td>0.0136</td>
<td>0</td>
<td>0.0023</td>
<td>0.4161</td>
<td>0</td>
<td>0.3302</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d8}$</td>
<td></td>
<td>0</td>
<td>0.0346</td>
<td>0</td>
<td>0.1959</td>
<td>0.0995</td>
<td>0</td>
<td>0.0654</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{8-}$</td>
<td></td>
<td>0</td>
<td>0.0390</td>
<td>0</td>
<td>0.0200</td>
<td>0.7220</td>
<td>0</td>
<td>0.5718</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d9}^*$</td>
<td></td>
<td>0</td>
<td>0.6622</td>
<td>0</td>
<td>0.2777</td>
<td>0.1109</td>
<td>0</td>
<td>0.2411</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{9-}^*$</td>
<td></td>
<td>0</td>
<td>0.9678</td>
<td>0</td>
<td>0.7088</td>
<td>0.9206</td>
<td>0</td>
<td>0.9551</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d10}^*$</td>
<td></td>
<td>0</td>
<td>0.9667</td>
<td>0</td>
<td>0.8983</td>
<td>0.2572</td>
<td>0</td>
<td>0.4751</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{10-}^*$</td>
<td></td>
<td>0</td>
<td>0.5879</td>
<td>0</td>
<td>0.8361</td>
<td>0.4519</td>
<td>0</td>
<td>0.4946</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d11}^*$</td>
<td></td>
<td>0</td>
<td>0.6762</td>
<td>0</td>
<td>0.4314</td>
<td>0.6569</td>
<td>0</td>
<td>0.6135</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{11-}^*$</td>
<td></td>
<td>0</td>
<td>0.9170</td>
<td>0</td>
<td>0.7654</td>
<td>0.4840</td>
<td>0</td>
<td>0.5095</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d12}^*$</td>
<td></td>
<td>0</td>
<td>0.7347</td>
<td>0</td>
<td>0.2616</td>
<td>0.7353</td>
<td>0</td>
<td>0.7797</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k^{12-}^*$</td>
<td></td>
<td>0</td>
<td>0.4308</td>
<td>0</td>
<td>0.0191</td>
<td>0.4524</td>
<td>0</td>
<td>0.4067</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{c}$</td>
<td></td>
<td>0</td>
<td>0.6591</td>
<td>0</td>
<td>0.9513</td>
<td>0.4417</td>
<td>0</td>
<td>0.2921</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{on,R}$</td>
<td></td>
<td>0</td>
<td>0.1463</td>
<td>0</td>
<td>0.3433</td>
<td>0.3853</td>
<td>0</td>
<td>0.3625</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Appendix E. Analysis of the GBL System Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>K-S</th>
<th>p value</th>
<th>K-S</th>
<th>p value</th>
<th>K-S</th>
<th>p value</th>
<th>K-S</th>
<th>p value</th>
<th>K-S</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{on,A}$</td>
<td>0</td>
<td>0.0963</td>
<td>0</td>
<td>0.9071</td>
<td>0</td>
<td>0.2335</td>
<td>0</td>
<td>0.3893</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{ar}^+$</td>
<td>0</td>
<td>0.8860</td>
<td>0</td>
<td>0.5463</td>
<td>0</td>
<td>0.0017</td>
<td>0</td>
<td>0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{ar}^-$</td>
<td>0</td>
<td>0.9928</td>
<td>0</td>
<td>0.7182</td>
<td>0</td>
<td>0.2796</td>
<td>0</td>
<td>0.1591</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_R$</td>
<td>0</td>
<td>0.1636</td>
<td>0</td>
<td>0.4825</td>
<td>0</td>
<td>0.0038</td>
<td>0</td>
<td>0.0154</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_A$</td>
<td>0</td>
<td>0.5470</td>
<td>0</td>
<td>0.8323</td>
<td>0</td>
<td>0.5179</td>
<td>0</td>
<td>0.5316</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{mA}$</td>
<td>0</td>
<td>0.0411</td>
<td>0</td>
<td>0.0553</td>
<td>0</td>
<td>0.0304</td>
<td>0</td>
<td>0.0267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{mR}$</td>
<td>0</td>
<td>0.2921</td>
<td>0</td>
<td>0.2872</td>
<td>0</td>
<td>0.1622</td>
<td>0</td>
<td>0.2298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_R$</td>
<td>0</td>
<td>0.0314</td>
<td>0</td>
<td>0.0306</td>
<td>0</td>
<td>0.2652</td>
<td>0</td>
<td>0.3231</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{R2}$</td>
<td>0</td>
<td>0.2321</td>
<td>0</td>
<td>0.2763</td>
<td>0</td>
<td>0.4139</td>
<td>0</td>
<td>0.4250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_A$</td>
<td>1</td>
<td>0.0002</td>
<td>1</td>
<td>5.91·10^{-7}</td>
<td>1</td>
<td>2.31·10^{-26}</td>
<td>1</td>
<td>2.81E-24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_C$</td>
<td>0</td>
<td>0.0864</td>
<td>0</td>
<td>0.0872</td>
<td>0</td>
<td>0.0534</td>
<td>0</td>
<td>0.0314</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{CR}$</td>
<td>1</td>
<td>0.0004</td>
<td>1</td>
<td>6.83·10^{-6}</td>
<td>1</td>
<td>1.70·10^{-12}</td>
<td>1</td>
<td>1.81·10^{-12}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{AR}^*$</td>
<td>0</td>
<td>0.4792</td>
<td>0</td>
<td>0.5875</td>
<td>0</td>
<td>0.5001</td>
<td>0</td>
<td>0.6455</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{mA2}^*$</td>
<td>0</td>
<td>0.6458</td>
<td>0</td>
<td>0.8601</td>
<td>0</td>
<td>0.6640</td>
<td>0</td>
<td>0.8972</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D$</td>
<td>1</td>
<td>0.0001</td>
<td>0</td>
<td>0.0087</td>
<td>0</td>
<td>0.0300</td>
<td>0</td>
<td>0.0498</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_F$</td>
<td>0</td>
<td>0.8906</td>
<td>0</td>
<td>0.3945</td>
<td>0</td>
<td>0.1312</td>
<td>0</td>
<td>0.2759</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{on,Abasal}$</td>
<td>0</td>
<td>0.7757</td>
<td>0</td>
<td>0.6191</td>
<td>0</td>
<td>0.0399</td>
<td>0</td>
<td>0.0383</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Parameters used only for the RA complex
+ Parameters used only for the antisense RNA interactions

**Table E.3:** Results from K-S test for the simulations with $\chi = 1.1 - 10$ ($k_{Fr} > k_{Fa}$)
### Appendix E. Analysis of the GBL System Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
<th>Value 5</th>
<th>Value 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{5-}^*$</td>
<td>0.2894</td>
<td>0.224</td>
<td>0.4995</td>
<td>0.5639</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d6}$</td>
<td>0.4617</td>
<td>0.381</td>
<td>0.6114</td>
<td>0.4632</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{6-}$</td>
<td>0.9829</td>
<td>0.998</td>
<td>0.7968</td>
<td>0.7717</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d7}$</td>
<td>0.0464</td>
<td>0.028</td>
<td>2.26e-5</td>
<td>3.64e-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{7-}$</td>
<td>0.2420</td>
<td>0.302</td>
<td>0.7223</td>
<td>0.6698</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d8}$</td>
<td>0.2519</td>
<td>0.185</td>
<td>1.81e-7</td>
<td>1.54e-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{8-}$</td>
<td>0.6721</td>
<td>0.858</td>
<td>0.9978</td>
<td>0.9921</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d9}$</td>
<td>0.8657</td>
<td>0.822</td>
<td>0.1550</td>
<td>0.1397</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{9-}$</td>
<td>0.5798</td>
<td>0.438</td>
<td>0.0110</td>
<td>0.0381</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d10}$</td>
<td>0.3085</td>
<td>0.318</td>
<td>0.9668</td>
<td>0.9340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{10-}$</td>
<td>0.8001</td>
<td>0.852</td>
<td>0.2213</td>
<td>0.4428</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d11}$</td>
<td>0.9144</td>
<td>0.781</td>
<td>0.5875</td>
<td>0.6088</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{11-}$</td>
<td>0.5914</td>
<td>0.741</td>
<td>0.5769</td>
<td>0.5657</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d12}$</td>
<td>0.5636</td>
<td>0.719</td>
<td>0.0808</td>
<td>0.0694</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{12-}$</td>
<td>0.3637</td>
<td>0.470</td>
<td>0.7488</td>
<td>0.7291</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{C0}$</td>
<td>0.0030</td>
<td>0.001</td>
<td>6.22e-6</td>
<td>8.77e-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{on,R}$</td>
<td>0.3342</td>
<td>0.260</td>
<td>0.855</td>
<td>0.5279</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{on,A}$</td>
<td>0.1847</td>
<td>0.372</td>
<td>0.593</td>
<td>0.3567</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{ar+}$</td>
<td>0.7624</td>
<td>0.631</td>
<td>0.082</td>
<td>0.0793</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{ar+}$</td>
<td>0.4037</td>
<td>0.425</td>
<td>0.421</td>
<td>0.3837</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{R}$</td>
<td>0.4096</td>
<td>0.521</td>
<td>0.053</td>
<td>0.0993</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{A}$</td>
<td>0.7232</td>
<td>0.809</td>
<td>0.018</td>
<td>0.0344</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{mR}$</td>
<td>0.0416</td>
<td>0.088</td>
<td>0.015</td>
<td>0.0461</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{mA}$</td>
<td>0.4933</td>
<td>0.523</td>
<td>0.029</td>
<td>0.0204</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{R}$</td>
<td>0.1239</td>
<td>0.160</td>
<td>0.743</td>
<td>0.9894</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{R2}$</td>
<td>0.1858</td>
<td>0.127</td>
<td>0.995</td>
<td>0.9573</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{A}$</td>
<td>0.0001</td>
<td>2.23e-4</td>
<td>4.30e-24</td>
<td>1.45e-22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{C}$</td>
<td>0.0133</td>
<td>0.022</td>
<td>0.0483</td>
<td>0.0267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{CR}$</td>
<td>0.0100</td>
<td>0.003</td>
<td>4.10e-16</td>
<td>6.32e-16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{AR}$</td>
<td>0.7947</td>
<td>0.876</td>
<td>0.3953</td>
<td>0.1828</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d_{mA+}$</td>
<td>0.7857</td>
<td>0.854</td>
<td>0.4663</td>
<td>0.2939</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix E. Analysis of the GBL System Model

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d1}$</td>
<td>0</td>
<td>0.0267</td>
<td>0.071</td>
<td>0.1159</td>
<td>0</td>
<td>0.1459</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{1}$</td>
<td>0</td>
<td>0.0558</td>
<td>0.079</td>
<td>0.8092</td>
<td>0</td>
<td>0.9898</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d2}$</td>
<td>0</td>
<td>0.1676</td>
<td>0.731</td>
<td>0.2736</td>
<td>0</td>
<td>0.0193</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{2}$</td>
<td>0</td>
<td>0.1983</td>
<td>0.093</td>
<td>0.4867</td>
<td>0</td>
<td>0.2672</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d3}$</td>
<td>0</td>
<td>0.4453</td>
<td>0.550</td>
<td>0.6591</td>
<td>0</td>
<td>0.6347</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{3}$</td>
<td>0</td>
<td>0.4856</td>
<td>0.783</td>
<td>0.5663</td>
<td>0</td>
<td>0.4190</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d4}$</td>
<td>0</td>
<td>0.2770</td>
<td>0.328</td>
<td>0.0528</td>
<td>0</td>
<td>0.0096</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{4}$</td>
<td>0</td>
<td>0.1415</td>
<td>0.450</td>
<td>0.6588</td>
<td>0</td>
<td>0.2329</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d5}$</td>
<td>0</td>
<td>0.9565</td>
<td>0.991</td>
<td>0.2503</td>
<td>0</td>
<td>0.0947</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{5}$</td>
<td>0</td>
<td>0.3782</td>
<td>0.626</td>
<td>0.5509</td>
<td>0</td>
<td>0.3372</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d6}$</td>
<td>0</td>
<td>0.7155</td>
<td>0.252</td>
<td>0.7333</td>
<td>0</td>
<td>0.5880</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{6}$</td>
<td>0</td>
<td>0.4333</td>
<td>0.365</td>
<td>0.4252</td>
<td>0</td>
<td>0.4482</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d7}$</td>
<td>0</td>
<td>0.9978</td>
<td>0.897</td>
<td>0.8166</td>
<td>0</td>
<td>0.9957</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{7}$</td>
<td>0</td>
<td>0.0109</td>
<td>0.004</td>
<td>0.9533</td>
<td>0</td>
<td>0.9995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d8}$</td>
<td>0</td>
<td>0.6306</td>
<td>0.831</td>
<td>0.8850</td>
<td>0</td>
<td>0.4438</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{8}$</td>
<td>0</td>
<td>0.8485</td>
<td>0.651</td>
<td>0.3948</td>
<td>0</td>
<td>0.8651</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d9}$</td>
<td>0</td>
<td>0.7159</td>
<td>0.461</td>
<td>0.5977</td>
<td>0</td>
<td>0.4255</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{9}$</td>
<td>0</td>
<td>0.7205</td>
<td>0.859</td>
<td>0.7212</td>
<td>0</td>
<td>0.8697</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d10}$</td>
<td>0</td>
<td>0.5275</td>
<td>0.582</td>
<td>0.5344</td>
<td>0</td>
<td>0.6210</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{10}$</td>
<td>0</td>
<td>0.3776</td>
<td>0.328</td>
<td>0.6142</td>
<td>0</td>
<td>0.8685</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d11}$</td>
<td>0</td>
<td>0.9206</td>
<td>0.455</td>
<td>0.7020</td>
<td>0</td>
<td>0.6706</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k_{11}$</td>
<td>0</td>
<td>0.6743</td>
<td>0.669</td>
<td>0.2024</td>
<td>0</td>
<td>0.0965</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Parameters used only for the RA complex
+ Parameters used only for the antisense RNA interactions

TABLE E.4: Results from K-S test for the simulations with $\chi = 0.1 - 0.9 (k_{FR} < k_{FA})$
### Analysis of the GBL System Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d12}^*$</td>
<td>0 0.0709 0 0.298 0 0.2652 0 0.3152</td>
</tr>
<tr>
<td>$k_{12}^*$</td>
<td>0 0.8764 0 0.546 0 0.8596 0 0.8056</td>
</tr>
<tr>
<td>$K_C$</td>
<td>0 0.0170 0 0.029 0 0.5462 0 0.2013</td>
</tr>
<tr>
<td>$k_{onR}$</td>
<td>0 0.6154 0 0.440 0 0.1188 0 0.0506</td>
</tr>
<tr>
<td>$k_{onA}$</td>
<td>0 0.4313 0 0.663 0 0.6256 0 0.3201</td>
</tr>
<tr>
<td>$K_{ar}^+$</td>
<td>0 0.7768 0 0.256 0 0.2549 0 0.6157</td>
</tr>
<tr>
<td>$k_{ar}^+$</td>
<td>0 0.5301 0 0.312 0 0.8482 0 0.3535</td>
</tr>
<tr>
<td>$P_R$</td>
<td>0 0.6696 0 0.574 0 0.2977 0 0.4510</td>
</tr>
<tr>
<td>$P_A$</td>
<td>0 0.6166 0 0.325 0 0.1792 0 0.4320</td>
</tr>
<tr>
<td>$d_{mR}$</td>
<td>0 0.9224 0 0.171 0 0.3007 0 0.3534</td>
</tr>
<tr>
<td>$d_{mA}$</td>
<td>0 0.8160 0 0.345 0 0.5158 0 0.4637</td>
</tr>
<tr>
<td>$d_R$</td>
<td>0 0.2017 0 0.151 0 0.0139 0 0.0253</td>
</tr>
<tr>
<td>$d_{R2}$</td>
<td>0 0.2066 0 0.290 0 0.5964 0 0.7874</td>
</tr>
<tr>
<td>$d_A$</td>
<td>1 0.0002 1 0.000 1 0.0007 1 0.001</td>
</tr>
<tr>
<td>$d_C$</td>
<td>0 0.0060 0 0.013 0 0.1561 0 0.0402</td>
</tr>
<tr>
<td>$d_{CR}$</td>
<td>0 0.0295 0 0.011 0 0.0040 0 0.0019</td>
</tr>
<tr>
<td>$d_{AR}^*$</td>
<td>0 0.6336 0 0.751 0 0.3669 0 0.4764</td>
</tr>
<tr>
<td>$d_{mRA}^+$</td>
<td>0 0.0967 0 0.046 0 0.5526 0 0.7448</td>
</tr>
<tr>
<td>$D$</td>
<td>0 0.0163 0 0.028 0 0.4502 0 0.7946</td>
</tr>
<tr>
<td>$k_F$</td>
<td>0 0.6876 0 0.929 0 0.2237 0 0.0748</td>
</tr>
<tr>
<td>$k_{onAbasal}^*$</td>
<td>0 0.7780 0 0.261 0 0.7901 0 0.8290</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>1 1.27\times10^{-14} 1 2.14\times10^{-16} 1 1.05\times10^{-21} 1 3.09\times10^{-25}</td>
</tr>
</tbody>
</table>

*Parameters used only for the RA complex

*Parameters used only for the antisense RNA interactions
Ensemble modelling strategies for the exploration of the \( \gamma \)-butyrolactone network in \textit{Streptomyces coelicolor}

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

2018

Areti Ioanna TSIGKINOPOULOU
School of Chemistry/ Faculty of Science and Engineering
# Contents

Contents 2

List of Figures 4

List of Tables 7

Abstract 10

Declaration 12

Acknowledgements 13

1 Introduction and literature review 15
   1.1 Systems and synthetic biology ............................. 15
   1.2 Design and analysis of computational models for biological networks 18
   1.3 Deterministic frameworks of analysis ............................ 26
   1.4 Parameter uncertainty and ensemble modelling ...................... 28
   1.5 Multivariate distributions and thermodynamic consistency .......... 32
   1.6 The γ-butyrolactone (GBL) regulatory system in *Streptomyces coelicolor* 35
   1.7 Aims and Overview of the Thesis .................................. 38

2 Respectful modelling: Addressing uncertainty in dynamic system models for molecular biology 41
   2.1 Computational models in current research: Success and scepticism .... 42
   2.2 A respectful approach to biological modelling ...................... 44
   2.3 Respectful modelling showcases and applications ..................... 48
   2.4 Conclusion ...................................................... 55

3 Defining informative priors for ensemble modelling in systems biology 57
   3.1 Introduction ..................................................... 58
      3.1.1 Applications of the protocol ................................. 61
      3.1.2 Overview of the protocol ..................................... 61
   3.2 Materials ....................................................... 63
   3.3 Procedure ....................................................... 63
      3.3.1 Step 1: Find and document parameter values .................. 63
      3.3.2 Step 2: Judge plausibility of values and assign weights .......... 66
### Contents

3.3.3 Step 3: Calculate the Mode and Spread of the probability distribution ........................................... 68
3.3.4 Step 4: Calculate the location and scale parameters ($\mu$ and $\sigma$) of the log-normal distribution .......... 70
3.3.5 Step 5: Account for thermodynamic consistency ................................................................. 71
3.4 Troubleshooting ......................................................................................................................... 76
3.5 Anticipated results .................................................................................................................... 77
3.6 Conclusion ................................................................................................................................. 86

4 Unravelling the $\gamma$-butyrolactone network in *Streptomyces coelicolor* under uncertainty ............... 87
   4.1 Abstract .................................................................................................................................... 87
   4.2 Introduction .............................................................................................................................. 88
   4.3 Methods .................................................................................................................................... 91
      4.3.1 Recreation of the published GBL and QS models ............................................................. 91
      4.3.2 Improved model of the scbR/scbA gene regulatory network .............................................. 96
   4.4 Results ..................................................................................................................................... 115
      4.4.1 Results from the recreation of GBL and QS models ......................................................... 115
      4.4.2 Results from the improved model of the GBL network ...................................................... 127
   4.5 Discussion ............................................................................................................................... 137

5 Summary and future perspectives .................................................................................................. 143

A Functions for the design of informative priors protocol ................................................................ 149
   A.1 Functions *CalcModeSpread* and *weightedMedian* ........................................................... 149
   A.2 Function *CreateLogNormDist* ............................................................................................ 156
   A.3 Function *Multivariate3param* ............................................................................................. 158
   A.4 Function *Multivariate5param* .............................................................................................. 160
   A.5 Function *MultivariateBiBi* ................................................................................................... 165
   A.6 Function *MultivariateUniBi* .................................................................................................. 169

B Supplementary case study for the design of informative priors protocol ............................................ 173

C Parameter information for recreated $\gamma$-butyrolactone and quorum sensing systems .................. 195

D Parameter information and uncertainty for the model of the $\gamma$-butyrolactone System ................. 203
   D.1 Binding of $R_2$ to $O_R$ operator .......................................................................................... 203
   D.2 Binding of $R_2$ to $O_A$ operator .......................................................................................... 211
   D.3 Binding of $R_2$ to $A$ ........................................................................................................... 218
   D.4 Binding of $R_2$ to $C$ ........................................................................................................... 223
   D.5 Binding of $AR_2$ to $O_A'$ operator ..................................................................................... 230
   D.6 Synthesis of $C$ ...................................................................................................................... 238
List of Figures

1.1 The cycle of hypothesis-driven research in Systems Biology ........ 16
1.2 Transcriptional regulatory network motifs ................................. 19
1.3 A typical sequence of events in signalling pathways ................... 20
1.4 Intertwined behaviour of signalling, regulatory and metabolic layers. . 21
1.5 Frameworks for the representation and analysis of biological networks .. 23
1.6 A qualitative comparison on modelling methods based on a relative scale 24
1.7 Properties of a bivariate normal distribution ............................... 34
1.8 Schematic diagram of the ScbR/ScbA system ............................... 36
1.9 Overview of the thesis ................................................................. 39
2.1 Respectful modelling outline ...................................................... 44
2.2 Flow chart of steps in *T. brucei* modelling ................................. 46
2.3 Flow chart of steps in *T. brucei* modelling ................................. 49
2.4 Distribution of the $K_m$ values retrieved from the BRENDA database.. 50
2.5 Sampling strategy for maintaining thermodynamic consistency of parameter sets ................................................................. 53
2.6 Trapped in a local optimum .......................................................... 54
3.1 Workflow of the protocol for design of informative priors ............... 62
3.2 Example of a back-of-the-envelope approach for estimating plausible parameter values ................................................................. 65
3.3 Calculation of the weighted median by *CalcModeSpread* function .... 69
3.4 Overview of the process followed by the *Multivariate3param* function 72
3.5 Overview of the process followed by the *Multivariate5param* function 75
3.6 Estimated probability distributions for parameters $K_{D1}$ and $k_1$, plotted on a log-scale ................................................................. 83
3.7 The bivariate distribution for parameters $k_1$ and $k_{on1}$ and the two marginal distributions ................................................................. 85
4.1 Steps followed for the design of an improved predictive model for the GBL system ................................................................. 91
4.2 Schematic representation of the published GBL and QS models ........ 95
4.3 Schematic representation of the potential mechanisms of the ScbA/ScbR system ................................................................. 99
4.4 Modelling approach for a growing cell culture in a fermentor .......... 108
4.5 Fitting the logistic growth curve to the reported experimental data by Nieselt *et al.* ................................................................. 114
4.6 Simulations results from the replication of published model on GBL system in Copasi .................................................. 116
4.7 Simulations results under parameter uncertainty on the published GBL model ................................................................. 118
4.8 Simulations results from replication of deterministic model on QS system (lux02 operon) ......................................................... 119
4.9 Identification of the bistability zone in the QS system .................... 120
4.10 Concentration profile of GFP/GFP_{max} during the simulations under parameter uncertainty on the QS system .................. 121
4.11 Comparison between the expected and the actual parameter values of the bistable QS models ........................................... 125
4.12 Variance of the bistability zone under parameter uncertainty .......... 126
4.13 Transcriptomics data for scbR and scbA genes by Nieselt et al. ...... 127
4.14 Highest ranked log-likelihood profiles for the 4 scenarios of the GBL model ............................................................. 129
4.15 Summary of log-likelihood analysis of the improved GBL model for the 4 scenarios ........................................................ 130
4.16 Comparison of the highest ranked models of the 4 scenarios with the transcriptomics data .................................................. 131
4.17 Effect of promoter strength on the improved GBL model for the 4 scenarios ..................................................................... 133
4.18 Comparison between the log-likelihood profiles of scenario B with varying promoter strengths .............................................. 134
4.19 Comparison between the expected and the actual parameter values of the highly ranked GBL models ...................................... 136
5.1 GUI prototype for the design of informative probability distributions 145
5.2 Future perspectives and contribution of the thesis ......................... 147
B.1 Schematic representation of the T. brucei glycolysis model ............... 174
B.2 Effect of reducing TPI on the steady-state flux of glucose, pyruvate and glycerol ................................................................. 193
List of Tables

1.1 Tools for visualisation and analysis of biological systems .......................... 26
3.1 Hypothetical example of documentation for hexokinase parameters in chicken liver cells retrieved from the literature .......................... 66
3.2 Standardized description of parameter information in Table 3.1, for use in Step 2 .......................................................... 66
3.3 Documentation of the experimental conditions reported in the literature for the hypothetical example in Table 3.1 and Table 3.2 ............... 67
3.4 Calculation of standardized weights from the information in Table 3.3 ...... 67
3.5 Final input table required for Step 3 for the hypothetical example in Tables 3.1–3.4 ......................................................... 68
3.6 Troubleshooting table ............................................................... 76
3.7 Summary of parameter values retrieved from literature for case study ...... 78
3.8 Summary of reported parameter values with their associated uncertainty 79
3.9 Summary of parameter values and assignment of weights for case study 80
3.10 Input matrix of $K_{D1}$ values found in the literature .......................... 81
3.11 Input matrix of $k^1_1$ values found in the literature .......................... 81
3.12 Summary of log-normal distribution properties for parameters $K_{D1}$, $k_{on1}$ and $k_{o1}$ ................................................................. 83
4.1 Chemical reactions for the published GBL model by Mehra et al.\textsuperscript{[1]} ............ 92
4.2 Chemical reactions for the published QS model by Weber et al.\textsuperscript{[2]} ............ 93
4.3 List of species for the ScbR/ScbA model ........................................ 101
4.4 Chemical reactions for the ScbR/ScbA model ........................................ 101
4.5 Differential equations for the ScbR/ScbA model ........................................ 103
4.6 Information on the probability distributions for the model parameters ........ 109
4.7 Parameters for cell culture growth ...................................................... 114
4.8 K–S Test results for the 85 models of the QS system that exhibited bistability ................................................................. 123
4.9 Summary of K–S test results for all scenarios and for different promoter strength combinations ......................................................... 135
B.1 Case study model reactions ............................................................... 175
B.2 Summary of parameter values with their associated uncertainty retrieved from literature for the trypanosome case study .......................... 175
B.3 Weighting scheme used to assess the plausibility of parameter values found for reactions in glycolysis pathway .......................... 183
B.4  Summary of parameter values and assignment of weights for case study  184
B.5  Input matrices  ........................................  186
B.6  Summary of log-normal distribution properties for parameters in the trypanosome glycolysis pathway  .......................  186
“The best material model of a cat is another, or preferably the same, cat.”

Norbert Wiener

Philosophy of science, 1945, 12(4), p.320
Abstract

Systems biology employs predictive models that guide hypothesis-driven research. As the use of computational models is increasingly gaining ground within the biological community and is moving from proof of concept to real-world applications, such as the engineering of synthetic biology circuits, there is a higher need for clear and testable predictions. However, parameter uncertainty is an important hindrance in the model analysis process.

This thesis carefully considers the so called “parameter problem” and suggests a framework that explicitly considers parameter uncertainty by sampling parameters from probability distributions. Thus, the pitfalls that traditional fitting strategies face can be avoided. By employing a “respectful modelling” approach that rigorously quantifies the confidence associated with the modelling results and keeps all options in view rather than focusing on a single maximum likelihood solution, models can be made adaptable in the face of emerging experimental results.

Another important consideration is the quality of the parameter distributions. As the issue of generating informative priors has not been successfully addressed yet, the thesis suggests a novel protocol that aims to fill this gap. The protocol concerns the collection of parameter values from a diverse range of sources (literature, databases and experiments), assessing their plausibility, and creating log-normal probability distributions, while maintaining the thermodynamic consistency of the model.

Finally, the suggested methodological developments are put in action, in order to explore the γ-butyrolactone regulatory system in *Streptomyces coelicolor*, a Gram-positive, soil bacteria. This small but complex system involves two genes (*scbR* and *scbA*) and regulates the antibiotic production through a mechanism of action that is not yet fully elucidated. The scenarios that have been suggested involve the formation of a putative ScbR–ScbA protein complex, potential transcriptional interference and antisense RNA interactions. The thesis describes the replication and ensemble modelling analysis of the previously published GBL models, as well as a model on the similarly structured quorum sensing (QS) system, in order to explore the behaviours of the two systems and allow comparisons between them. Additionally, a new, versatile and adaptable computational model was designed which considered all three suggested mechanisms in different combinations and with varying promoter strengths. The analysis of the model suggests that the most significant influence in the system’s behaviour seems to come from antisense RNA interactions, combined with an aggressive scbR promoter. The model can be used to question and refine our understanding of the system’s activity.
and could even suggest a different biological role than the one originally assumed. The model also indicates key experiments which could further elucidate the role of the system and the interactions of its components and ultimately lead to the design of robust and sensitive systems which can be used in synthetic biology and biotechnology.
**Declaration**

No portion of the work referred to in this 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.

**Copyright**

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 she 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, trademarks 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=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Librarys regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University’s policy on Presentation of Theses.
Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Rainer Breitling for the continuous support of my PhD study and the related research, for his patience, motivation, and immense knowledge. His guidance helped me throughout the time of my research and writing of this thesis. I could not have imagined having a better advisor and mentor for my PhD project.

Besides my advisor, I would like to thank my co-supervisor Prof. Pedro Mendes for his expert advice on computational modelling and on the use of the COPASI software.

This work would not materialise without our experimental collaborators, Prof. Eriko Takano, Marc Biarnes Carrera and Adam Amara who offered continuous advice, support and experimental data, and for that I am deeply grateful.

I would like to express my very great appreciation to my fellow group members and friends Francesco Del Carratore, Dr. Aliah Hawari and Megan Uttley for their help, for the stimulating discussions and brainstorming sessions, and for contributing in my work, each in their own special way.

Special thanks to my fellow PhD student in the Manchester Institute of Biotechnology, Chrysa Zerva, for her problem solving interventions in MATLAB and for the unforgettable travels around the globe.

I would also like to extend my thanks to Dr. Marc Weber and Prof. Javier Buceta for sharing their modelling software (Colony) and providing useful advice on the quorum sensing system.

Last but not least, I would like to thank my parents, my sister, my cat and my friends in Greece and in Manchester for being there for me and supporting me throughout my PhD and my life in general.
Blank page
Chapter 1

Introduction and literature review

1.1 Systems and synthetic biology

Although there is no general consensus on a formal definition of Systems Biology\[^3\], it can be described as the computational and mathematical assembly and analysis of complex biological networks. It does so by focusing more on the nature of the interactions between macromolecules and the resulting functional states of the system, rather than on the components themselves. An integrative approach is employed instead of the traditional reductionism in order to discover emergent functional properties in living organisms.\[^4,5\]

Systems biology provides a framework for predictive, hypothesis-driven research\[^3\] whose basic steps are presented in Figure 1.1. Prediction emerges from the generation of models based on hypotheses that can be experimentally tested and validated\[^5\]. The models are then refined and improved iteratively\[^6\] in order not to only map the interactions between components, but also the response of the system to perturbations\[^3\].

As it becomes apparent that the most prominent diseases of our times such as cancer, diabetes, viral infections and heart disease are affected by multiple factors and not by single molecules, a holistic examination of the biological organisms is required.\[^4\] In this respect, systems biology whose nature provides the base for such an approach, can have significant applications in medicine, such as drug target identification and personalized medicine\[^8\], disease prevention and healthier life style development.
Chapter 1. Introduction

Figure 1.1: The cycle of hypothesis driven research in Systems Biology. Research is motivated by a relevant area of biological interest, where alternative hypotheses can be formed. Thus, a hypothesis is generated which forms the base for designing of a number of mathematical models representing the reaction network under investigation. The models are then analysed computationally by a variety of methods in order to interpret and predict biological functions that emerge from the reconstructed network and which need to be experimentally tested. Models which are inconsistent with established experimental facts are amended or discarded. Afterwards, appropriate “wet lab” experiments are designed and performed and their results are analysed. The models are refined accordingly and the inadequate ones are rejected. The surviving models that are consistent with the experimental data provide a new biological insight and the cycle begins again. The development of new technologies for measurements and analytical methods along with the advancement of computational science and of the “omics” technologies, reinforce systemic research and improve the accuracy and quality of the results.

Another important goal of systems biology is to move away from abstract mathematical modelling and closer to assisting with the design of novel and improved biological functions via “synthetic biology”. This emerging interdisciplinary area involves the application of engineering principles to biology in order to design and build organisms with specific genetic traits that might not exist in nature. In this respect, synthetic biology has made very rapid advances in recent years in engineering metabolic pathways or regulatory components to improve production of specific compounds. Successful applications include the production of an antimalarial drug, the creation
of a library of 74 novel chemical compounds\textsuperscript{14} and the expression of natural products in heterologous hosts.\textsuperscript{15}

A key natural property that is of great interest for synthetic biology is bistability. It is one of the main features of the quorum sensing systems\textsuperscript{16} that will be the main focus of this thesis. Bistable switches are usually encountered in the decision making processes\textsuperscript{17} during various phases of the cell cycle (e.g. cell differentiation or apoptosis). Their two key properties of memory (the system remembers its “on” state despite a potential drop in the stimulus level) and bimodality (at intermediate stimulus levels the cell population is distributed between two states), make them very attractive building blocks in synthetic gene circuits.\textsuperscript{18–20}

Undoubtedly, computational models of cellular pathways and circuits are increasingly becoming essential tools for generating clear, testable predictions about the behaviour of complex cellular machineries (Section 2.1). Their use is currently moving beyond the proof-of-concept stage towards real-world applications, such as the engineering and optimisation of biological microorganisms for the production of specific chemicals and biofuels.\textsuperscript{21–25} In this respect, computational models are an extremely valuable addition to the synthetic biology and biotechnology toolbox.

However, the field has still a lot of potential for growth. The upsurge of wet lab data availability and accuracy combined with increasing computer power and the development of new tools, algorithms and standardized interfaces are bound to be the driving forces in the development of systems biology. Also, the focus is gradually moving from simply adding systems components, to making the systems descriptions more complete, accurate and realistic.\textsuperscript{4} To achieve this, a combination of computational methods (e.g. constraint-based\textsuperscript{26,27} and kinetic methods\textsuperscript{28,29}) is employed in order to perform systemic analysis on multiple levels (e.g. from specific pathways to genome-scale modelling).
1.2 Design and analysis of computational models for biological networks

Cellular functions require the collaborative action of a number of components (e.g. proteins, enzymes, genes) within the cell which interact with each other, and thus form a network. For instance, a signalling cascade pathway involves a series of chemical reactions initiated by a compound (stimulus) binding on a receptor and is transduced within the cell by second messengers (other signalling molecules) that propagate the signal and eventually induce a cellular response. In each network there is a number of different elements with individual roles and properties along with a number of different relations which connect them to each other and to the external environment. The system’s properties thus rely on the properties of its components, as well as on the links between them.\(^5,8,30\)

Alongside its investigation by experimental methods, a biological process can also be described by using mathematical models. The appropriate model should be designed to reflect the essential properties of the system. However, depending on the nature of the problem and the purpose and intention of the investigator, different aspects of the same system may be highlighted.\(^30\)

There are different types of biological networks that can be designed, such as gene regulatory networks, signal transduction networks and metabolic networks.

**Gene Regulatory Networks**

Regulatory networks contain the information on gene expression control within cells. They are usually represented by nodes (proteins, transcription factors, mRNA and DNA regulatory elements) and directed incoming or outgoing edges. The edges imitate the physical binding between proteins or between transcription factors and DNA, thus inhibiting or promoting transcription of target genes.

In this way, the series of events in the different stages of the process that eventually lead to gene expression are studied. Often, particular structural motifs and patterns occur (Figure 1.2), which can be studied to gain insight into the function of the overall system.\(^5,31,32\) Although the mechanisms of such networks are still partially understood,
progress is being made and data collection and integration techniques are facilitating the analysis and mapping of such networks. The reconstruction can be performed either by using data from literature or from experimental methods such as yeast one-hybrid (Y1H), and ChIP-Seq or ChIP-Chip techniques.

**Signal Transduction Networks**

Signal transduction networks investigate the reception and propagation of a signal, either intracellular or extracellular, and the sequence of events that are consequently triggered by the disruption of the cell’s homoeostasis. These networks are usually represented by nodes being proteins, chemicals and other macromolecules and directed edges which depict the interactions between them which lead to the signal transmission (Figure 1.3). Common patterns also emerge from this type of networks.
Metabolic Networks

Metabolic networks aim to thoroughly describe the biochemical reactions which take place in a cell or organism. The representation of such a system is hierarchical, starting from nodes (biochemical metabolites) connected by directed or undirected edges. These edges represent the reversible or irreversible reactions that convert one metabolite to another along with the enzymes that catalyse them. These reactions compose metabolic pathways which in turn eventually build together the metabolic network. The reconstruction of metabolic networks requires the gathering of genome sequencing data from databases and studies of orthologous genes from various species.
Chapter 1. Introduction

Figure 1.4: Intertwined behaviour of signalling, regulatory and metabolic layers. Signalling networks receive environmental stimuli and accordingly trigger cascades of information which eventually reach regulatory networks’ transcription factors and activate or repress target genes. These changes in gene expression may affect enzyme concentrations which catalyse metabolic reactions. Metabolic networks provide the necessary resources for the maintenance of homeostasis and the phenotype expression of the cell.[36]

Although usually the above discussed networks are designed and analysed individually, an emerging trend in systems biology is the collective analysis of signalling, regulatory and metabolic networks and of the interplay between them, in order to study cellular behaviour.[37] This approach is based on the interconnection of these different levels of cellular organisation (Figure 1.4) in order to react to external stimuli and ultimately maintain the cell homeostasis.[36]

The focus of this work was the design and analysis of the regulatory/signalling network for the production of antibiotics controlled by \(\gamma\)-butyrolactones in *Streptomyces coelicolor* (Chapter 4).

Even though computational modelling cannot replace experimentation, it can definitely enhance it and offers a lot of distinct advantages. Among others, it is cheaper and more environmentally friendly, it allows testing of scenarios and time courses that cannot be measured in an experiment, as well as the imposition of perturbations which are
not possible in the real system (i.e. perturbations without directly changing other system’s components). Overall, models allow researchers to capture their beliefs on the function of a biological system and, amidst their incomplete knowledge, update their thinking as more evidence becomes available. In order to be able to do so, there are some important considerations which must be taken into account while designing the model of a biological system.

- The quantities involved must be carefully considered. These can be either parameters or kinetic rates (quantities with assigned values), variables (quantities with changeable values) or constants (quantities with fixed values). The same quantity can be either of the three depending on the model.

- It is possible that different system structures may produce similar outputs, therefore making the output itself insufficient to predict the internal organisation of the system. Therefore, the model’s structure uniquely determines the behaviour of the system, but the opposite does not apply.

- Thermodynamic consistency should always be respected along with all the elementary laws of physics, such as conservation of mass. The model’s reactions must comply with the basic properties of chemical transformations (stoichiometry, relative and absolute rates) and the network’s functional states are determined by physicochemical constraints.

There are a variety of frameworks for the analysis of models. The most prominent types of frameworks currently used for the representation and analysis of biological systems are summarised in Figure 1.5. These can be roughly classified in three main categories depending on the approach they are taking on the network reconstruction and on the system’s properties that they are focusing on: structural and stoichiometric approaches, kinetic approaches and hybrid approaches. The first category assumes either static conditions to analyse topological features of the network (i.e. Boolean) or pseudo-steady state conditions, and constrained relations among fluxes to predict cellular responses (i.e. Flux Balance Analysis). The second category focuses on the time
evolution of the system under detailed flux specifications either in discrete or continuous states, and the third one entails a combination of methods from different categories in order to achieve a more realistic depiction of the system.\textsuperscript{[36]}

![Figure 1.5: Frameworks for the representation and analysis of biological networks, working on different levels of integration. Subcategories are indicated by different colours. The classification is not absolute, as many of the methods can be incorporated in more than one categories (i.e. Petri Nets can be included in Logical/Discrete, Hybrid or Stochastic network analyses)](image)

It is thus becoming obvious that there is a large number of different frameworks which can be employed in order to design and analyse a biological network. Each has its advantages and disadvantages and can be used individually or in combination with other methods, depending on the aims of the researcher. The choice of a modelling method is determined by the aspects of the system which need to be examined, the availability of data and information on the system, the computational resources and on how realistic we wish the estimation of the model to be. For example, logical methods provide a more qualitative analysis and might be a good choice when the topology of a system needs to be examined. On the other hand, kinetic methods are better suited for a time course study of a system’s components.
Chapter 1. Introduction

Figure 1.6: A qualitative comparison on modelling methods based on a relative scale. (Figure adapted from [36]) On the one end of the scale there are the logical models which are more abstract and require the least amount of data, but at the same time can display only qualitative dynamic behaviour. FBA-derived methods provide good metabolic predictions but require more data and is more difficult to analyse. On the other end of the scale, single-molecule level models are the most detailed but are computationally expensive. Approximations to stochastic simulation algorithms (SSAs) decrease the computational burden but sacrifice some detail for better performance. Hybrid models provide a good compromise between information requirements, computational effort and accuracy of outcome [36,38–40].

The systems properties might also favour or disqualify a particular modelling technique. For instance, for a small system with a number of molecules in the range of 10–100 a stochastic or discrete approach could be very useful in order to capture the effect of noise, while in a larger system this approach would be both computationally burdensome and unnecessary as the system behaviour would approximate a deterministic time course simulation. Each method incorporates a different level of detail and requires a different computational effort, thus there is no “good” or “bad” method, only a more suitable one for each occasion. A schematic comparison between different modelling methods is presented in Figure 1.6.

Additionally, there are different tools to perform visualisation and analysis of biological networks, some of which are summarized in Table 1.1. Popular interfaces can be diagrammatic (i.e. JDesigner), spreadsheet (i.e. JigCell Model Builder), forms-based
(COPASI) or text-based (i.e. Cellerator, Dizzy), depending on the needs and preferences of the users. A combination of these tools is provided by some packages.\(^4\)

In order to facilitate the communication between different modelling tools and subsequently an integrated analysis pipeline, model descriptions are being standardised. This practice is becoming applied to every aspect of the model building process, from file formats to variable names and equation structures and to mandatory archiving of the resulting models. The Systems Biology Markup Language (SBML)\(^{41}\), an XML-based machine readable language, is supported by almost all software tools in the field and is prevailing as a model representation format that facilitates model sharing and replication. Additionally, models are starting to comply with some basic community guidelines that describe the minimum information that needs to accompany a published model (MIRIAM)\(^{42}\). Moreover, centralised model repositories such as BioModels\(^{43}\), JWS Online\(^{44}\) and CellML Model repository\(^{45}\) have been created to enable the distribution and curation of biological systems’ models.
1.3 Deterministic frameworks of analysis

As previously discussed (Figure 1.5), there are different modelling techniques for the analysis of biological systems. These can be roughly classified in three main categories depending on the approach they are taking on the network reconstruction and on the system’s properties that they are focusing on: structural and stoichiometric approaches, kinetic approaches and hybrid approaches. The first category assumes either static conditions to analyse topological features of the network (i.e. Boolean) or pseudo-steady state conditions and constrained relations among fluxes to predict cellular responses (i.e. Flux Balance Analysis). The second category focuses on the time evolution of the system under detailed flux specifications either in discrete or continuous states, and the third one entails a combination of methods from more than one categories in order
to allow the exploration of different behaviours of the system.\textsuperscript{[36]} During the course of this work, continuous deterministic (Ordinary Differential Equations) were employed combined with the explicit consideration of parameter uncertainty in order to design an improved computational model of a bacterial regulatory system (Chapter 4).

**Continuous methods**

Continuous modelling is a widely used framework for sub-networks of biological systems, as it provides quantitative predictions of networks’ behaviour through time and allows a series of analytical methods. However, these models require a significant amount of knowledge and data (e.g. reactions, kinetic rates, parameter values) on the system under investigation, which might be a problem when investigating large networks.\textsuperscript{[36]}

Ordinary Differential Equations (ODEs) are a well established quantitative approach which models the behaviour of various dynamic networks, such as signalling, regulatory or metabolic pathways, and their responses to perturbations. They are in the form:

\[
\frac{dx_i}{dt} = f_i(x_1, \ldots, x_n) \tag{1.1}
\]

where \(x_i\) is the quantity of the \(i\)-th model species and \(f_i\) is the rate of change of \(x_i\), for \(n\) quantities in the modelled system. ODEs have been used successfully for the analysis of complex systems in combination with other methods\textsuperscript{[57]} as well as for the analysis of combined types of networks.\textsuperscript{[58, 59]} ODEs have been employed for the deterministic analysis of the models built during this project in order to describe the behaviour of the system when the number of molecules of the model’s species is large enough so that fluctuations are negligible. In this case, the model can give a sufficiently accurate description of continuous concentrations at a population level where noise is averaged out. However, in reality, chemical reactions might be more accurately described as discrete stochastic processes, where reactants may encounter each other randomly, and may react or not in any given collision.

Partial Differential Equations (PDEs) can also be used for modelling biological systems when there are more than one independent variables in the system (e.g. time and space).\textsuperscript{[60, 61]} Their general form is:
Chapter 1. **Introduction**

\[ f \left( x_1, \ldots, x_n, u, \frac{\partial u}{\partial x_1}, \ldots, \frac{\partial^2 u}{\partial x_1 \partial x_1}, \ldots, \frac{\partial^2 u}{\partial x_n \partial x_n}, \ldots \right) = 0. \tag{1.2} \]

A typical example of a problem with multiple independent variables is diffusion of a chemical from the cell cytoplasm to the external environment. This substance will have different concentration in the two compartments, so in order to determine the system’s behaviour the partial differential equation

\[
\frac{\partial F}{\partial t} = D \left( \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \frac{\partial^2 F}{\partial z^2} \right) \tag{1.3}
\]

can be used to represent the concentration of solute at point \((x, y, z)\) at time \(t\). Thus, the partial derivative of concentration with respect to time varies with the sum of the second partial derivatives with respect to each spatial dimension.\textsuperscript{62} Although PDEs have not been used in this work, they would provide an interesting future contribution to the analysis of the bacterial system which was investigated, in order to better study the role and effect of diffusion on the colony behaviour.

### 1.4 Parameter uncertainty and ensemble modelling

Biological models require the knowledge of kinetic parameters in order to generate reasonable outputs. However, the parameter values that are needed in order to model a biological system, are usually unknown or uncertain, due to lack of experimental data, measurement errors or biological variability.\textsuperscript{63} Lack of knowledge on agents or even reactions involved in the system, multiple measured values reported in literature and parameters of specific species that are too expensive or time consuming (or even impossible) to measure, are commonly encountered issues when modelling and show our limited or incomplete knowledge of the system’s properties. All these factors contribute to the so called “parameter problem”,\textsuperscript{48,64,65} the determination of the parameter values employed in a computational model in a way that promotes the credibility of the model’s output.

The most widely adopted approach in model building is the determination of the required parameters values by fitting them to quantitative data. With the development of
non-linear optimisation algorithms\(^66-\)\(^68\) and global optimisation methods,\(^69-\)\(^71\) parameter fitting has been made easier and more approachable for the biological community. The latter ensure better efficiency and robustness and can be classified as deterministic\(^72-\)\(^75\) stochastic\(^76-\)\(^78\) and hybrid\(^79\) strategies. Deterministic methods take advantage of the problem’s structure and ensure global convergence for problems that meet specific smoothness and differentiability conditions.\(^80\) They are very promising and powerful and provide a level of confidence that the global optimum will be located. However, there are still limitations to their applications, mostly related to the high computational cost associated with the size and complexity of the model.\(^80\) Global stochastic methods do not require any assumptions and mostly rely on probabilistic approaches. In contrast to the deterministic ones, these methods arrive with relative efficiency in the vicinity of the solution, but convergence to the global optimum cannot be guaranteed. Nonetheless, the key benefit of these methods is that they provide a very good solution with reasonable computational effort. The most prominent groups of stochastic global optimisation methods are adaptive stochastic methods,\(^81\) clustering methods,\(^82\) evolutionary computation (EC),\(^83,\)\(^84\) simulated annealing (SA)\(^85\) and meta-heuristics.\(^86\) In order to overcome the limitations of the aforementioned methods, hybrid strategies have been employed in dynamic optimisation\(^87,\)\(^88\) and parameter estimation problems.\(^89\)

Although the model fitting approach is prevalent in systems biology, it should be used with caution, as they can lead to overfitting and thus overlook alternative hypotheses that could equally well explain the results. The highly rugged likelihood surface of the biological systems’ models often leads to that situation, as the likelihoods of models with considerably different sets of parameter values can often be quite similar. This issue is especially important as the current use of models’ predictions are sought to explain complex bio-systems and guide the engineering of micro-organisms. It is therefore important to explore the full parameter space rather than focus on the maximum likelihood solution and keep all possible alternatives into view. In this way, models will be more adaptable to changes that may alter the likelihood landscape as new experimental data become available (Chapter 2).

A way to achieve this, is to incorporate parameter uncertainty in the models from
their early building stage, as it allows the identification of alternative structures and parameter values and the testing of their plausibility. Additionally, it allows predictions to be performed within a level of confidence and can guide further experiments. In order to allow predictive model building in the face of parameter uncertainty, modellers are increasingly employing “ensemble modelling” strategies. In this case, a probability distribution is defined for each parameter, which can have either an arbitrary or standard shape (i.e. log-normal distributions).\textsuperscript{[63,90]} Thus, by sampling reasonable parameter values from the defined distributions, an “ensemble” of outputs is generated, which accurately reflects the confidence limits for each model prediction. The resulting models can then undergo the same kind of further analysis as typical (fixed-parameter) dynamic models, such as global sensitivity analysis\textsuperscript{[91]}, metabolic control analysis,\textsuperscript{[92,93]} enzyme cooperativity analysis\textsuperscript{[94]} and cofactor analysis.\textsuperscript{[95]} This approach was employed during the course of this project in order to study a regulatory system in the bacteria \textit{Streptomyces coelicolor}. Furthermore, in order to contribute to the ensemble modelling toolbox, a novel protocol is presented in Chapter 3, which addresses the subject of defining informative distributions for parameter sampling, based on the information retrieved from literature.

\begin{wrapfigure}{r}{0.5\textwidth}
\textbf{Why sample parameters from distributions?}
\begin{itemize}
  \item The method that has been traditionally followed by the biological community to define a model’s parameters, is fitting them to experimental data. It is thus reasonable to wonder why sampling parameter values from distributions and generating a collection of models is a better approach. Since the number of parameter values and of their possible combinations is infinite, even a sample of 1,000,000 parameter sets will not cover even 1% of the total amount. Why do we therefore consider sampling a more reliable method than fitting? The answer to this question is directly related to the concept and use of sampling in science. As it is usually impossible to gather data from the entire population that is being studied, data is collected and analysed from a subset, which serves as an indicator for the entire population. Although samples always constitute a small proportion of the total population, they can be used to make statistical inferences about the entire population, provided that they are representative. The real question therefore should be: how much we trust our sample in order to make such assumptions and what we can do to improve its quality?

  An important step in this direction is to employ an appropriate sampling method. By using probability distributions, a large percentage of bias (at least of the systematic kind) is eliminated. Although we specify a range of values from which the parameter samples will be drawn, this range is based on experimental data and scientific knowledge, and is further constrained by thermodynamic feasibility. It is therefore possible to make
\end{itemize}
\end{wrapfigure}
predictions on the behaviour of the system within specified confidence intervals and if necessary, improve the sampling method or range of plausible values. Last but not least, a biological system will never have fixed parameter values in real life, as there will always be fluctuations and stochastic events which will affect it. Therefore, in order to achieve a more realistic representation of the system, a range of possible parameter values need to be explored.

There are, however, a number of challenges that need to be addressed by the modeller when applying this strategy. These concern mostly the design of the probability distributions describing parameter uncertainty (i.e. choosing the most plausible parameter value and how the plausibility changes around this value), and the actual parameter sampling process (i.e. sampling of interdependent parameters). This will be discussed in detail in Chapters 2 and 3.

Why log-normal distributions?
The distribution shape that may seem like the most logical choice to describe plausible parameter values is often a normal distribution, as it has some distinctive advantages. It can be very useful due to the central limit theorem and it is manipulated algebraically much more easily than other distributions. Although the normal distribution is a good choice to describe physical quantities that are the sum of independent processes, the log-normal distribution is the most “natural” one, as it is ubiquitous in nature across different fields. From social sciences and economics, to linguistics and ecology, there are numerous studies where the generated data fit a log-normal distribution. This is even more true in medicine and biological sciences, as the mean values are low, the variances are large and the values cannot be negative (i.e. latency periods of infectious diseases, species abundance etc.).[96]

The explanation for this occurrence is quite simple. The phenomena determined by numerous independent and additive forces can generally be described by a normal distribution. However, when the determining forces are not additive but “multiplicative” and arise from non-linear fluctuations in successive steps, the resulting errors are usually log-normally distributed.[97,98] The law of proportionate effect (Gibrat’s law) i.e. the change in a variable at any step of the process in a random proportion of its previous value is gives rise to a description by a log-normal distribution.[98] In life sciences, chemistry and physics are disciplines where the prevailing operation is multiplication. Reaction velocities, equilibrium conditions even surface and volume measurements are regulated by factors that act in a multiplicative way. In this respect, the log-normal distribution seems to be an accurate description of life itself through its different aspects and therefore the most realistic shape to use when describing parameter uncertainty. This comes in contrast with the human mind that is more in favour of the normal distribution.
1.5 Multivariate distributions and thermodynamic consistency

In some cases, parameters cannot be sampled separately either because they are statistically dependent, subject to thermodynamic constraints or depend on another common parameter.\[63\] Therefore, thermodynamic consistency is also an important factor that needs to be considered to decide if the combinations of parameters are plausible. For instance, a very common occurrence in biological systems are reversible reactions. In this case the dependency can be defined via the equilibrium constant, which denotes the relationship between the kinetic parameters for the “forward” and “reverse” components of the reaction. Let’s assume a reaction that is known to have an equilibrium constant very close to 1, i.e. its standard Gibbs free energy $\Delta G^0 = 0$. There is not much information about the rate of the reaction, so each of the two parameters is sampled from a very broad distribution. If the additional thermodynamic information is not taken into account, there will often be cases where values will be sampled from the “fast” end of the spectrum for the forward reaction rate, and from the “slow” end for the backward rate (or vice versa). Thus, inconsistent pairs of the two parameters will be generated. In this case, thermodynamic consistency requires that we discard such samples and only keep those where the two reaction rates are relatively similar, depending on the information we have on the equilibrium constant.

In order to address this problem during this project, a joint probability distribution (multivariate distribution) for the interconnected parameters (i.e. $k_{\text{forward}}$ and $k_{\text{reverse}}$) was employed, in order to ensure that each of the generated values for both of them are constrained within a specified range. Additionally, this approach ensures that their dependency on each other and on the equilibrium constant $K_{eq}$ is taken into account and quantified appropriately.

Although recent works have made important steps in the maintenance of thermodynamic consistency while sampling\[99–102\], it has mostly been applied while sampling from non-informative priors. The rigorous definition of the informative prior distributions that would be used as an input into such frameworks, has not yet been addressed in
Chapter 1. Introduction

the context of systems biology. This was therefore the focus of this project and the protocol
designed for the definition of informative priors in Chapter 3 includes this method as an optional final step.

The conditions that need to be met in order for a multivariate normal distribution \( \mathcal{N}_k(\mu, \Sigma) \) of a random vector \( x = (X_1, \ldots, X_k)' \) to exist are the following:

- Every linear combination of its components \( Y = a_1 X_1 + \ldots + a_k X_k \) is normally distributed. Thus, for any constant vector \( a \in \mathbb{R}^k \), the random variable \( Y = a' x \) has a univariate normal distribution.

- A random \( l \)-vector \( z \) exists, whose components are independent standard normal random variables, as well as a \( k \)-vector \( \mu \) and a \( k \times l \) matrix \( A \), such that \( x = Az + \mu \). \( l \) is the rank of the covariance matrix \( \Sigma = AA' \).

- There is a \( k \)-vector \( \mu \) and a symmetric, nonnegative-definite \( k \times k \) matrix \( \Sigma \), such that the characteristic function of \( x \) is \( \phi(u) = \exp(iu' \mu - \frac{1}{2} u' \Sigma u) \).

In the bivariate case (2-dimensional nonsingular case), the probability density function of a vector \( [XY]' \) is:

\[
 f(x, y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1-\rho^2}} \exp\left( -\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X \sigma_Y} \right] \right) \tag{1.4}
\]

where \( \rho \) is the correlation between \( X \) and \( Y \) and \( \sigma_X > 0 \) and \( \sigma_Y > 0 \). In this case, \( \mu = \left( \mu_X, \mu_Y \right) \), \( \Sigma = \begin{pmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{pmatrix} \).

The parameters are assumed to be independent if \( \rho = 0 \), so there is no correlation between them. Otherwise, the resulting bivariate iso-density loci plotted in the \( x,y \)-plane are ellipses (Figure 1.7). As the correlation parameter \( \rho \) increases, these loci appear to be squeezed to the following line:

\[
 y(x) = \text{sgn}(\rho) \frac{\sigma_Y}{\sigma_X} (x - \mu_X) + \mu_Y \tag{1.5}
\]

The two initial probability density functions \( f(X) \) and \( f(Y) \) are called marginal distributions, because each distribution gives the probability of various parameter values.
Samples generated from a bivariate normal distribution, shown along with the 2-sigma ellipse, the two marginal distributions, and the two 1-D histograms.

(b) Bivariate normal joint probability density function along with the two marginal PDFs

**Figure 1.7:** Properties of a bivariate normal distribution plotted in MATLAB from two marginal distributions with \( \mu=0 \) and Covariance Matrix= \([1 \ 0.6; 0.6 \ 2]\)

To be sampled without taking into account the values of the other distribution. Assuming one parameter is dependent on the other, the conditional distribution of \( X \) given \( Y \) is:

\[
X \mid Y = y \sim \mathcal{N}\left( \mu_X + \frac{\sigma_X}{\sigma_Y} \rho(y - \mu_Y), (1 - \rho^2)\sigma_X^2 \right)
\]

where \( \rho \) is the Pearson product-moment correlation coefficient between \( X \) and \( Y \) and 
\[E(X \mid Y = y) = \mu_X + \frac{\sigma_X}{\sigma_Y} \rho(y - \mu_Y)\] is the conditional expectation or conditional mean.

In the case of \( k_{\text{forward}} \) and \( k_{\text{reverse}} \), if the parameters are described by normal distributions the method is efficient as long as the value of \( K_{eq} \) is a known constant. In this case, the two marginal distributions are \( k_{\text{forward}} \) and \( k_{\text{forward}} \cdot K_{eq} (=k_{\text{reverse}}) \), meaning that \( k_{\text{reverse}} \) is dependent on the values of \( k_{\text{forward}} \) and \( K_{eq} \). However if the value of
the $K_{eq}$ also becomes uncertain, the bivariate distribution cannot be generated. This is due to the fact that if both $k_{\text{forward}}$ and $K_{eq}$ are normal distributions, their product will not be normally distributed and the first condition for the existence of a multivariate distribution is no longer met.

This is not the case, however, for the log-normal distributions, as any product of two log-normal random variables is also log-normally distributed. Therefore, assuming two log-normal distributions $\ln X \sim \mathcal{N}(\mu_{\ln X}, \sigma^2_{\ln X})$ and $\ln Y \sim \mathcal{N}(\mu_{\ln Y}, \sigma^2_{\ln Y})$, their product $Z$ will be the log-normal distribution $\ln Z \sim \mathcal{N}(\mu_{\ln X} + \mu_{\ln Y}, \sigma^2_{\ln X} + \sigma^2_{\ln Y})$ and its parameters will be $\mu_{\ln Z} = \mu_{\ln X} + \mu_{\ln Y}$, $\sigma^2_{\ln Z} = \sigma^2_{\ln X} + \sigma^2_{\ln Y}$. Additionally, since the natural logarithms of the values of a log-normal distribution are normally distributed, the problem can be reduced to the case of two normal distributions and the formulas described above can be employed. This property further supports the choice of log-normal distributions for describing parameter uncertainty in a biological model.

1.6 The $\gamma$-butyrolactone (GBL) regulatory system in *Streptomyces coelicolor*

The network that has been investigated in most detail in this thesis was the $\gamma$-Butyrolactone (GBL) regulatory system, of *Streptomyces coelicolor*. These are Gram-positive, filamentous, soil-dwelling bacteria, which are known as prolific source of secondary metabolites, such as antibiotics (actinorhodin (Act), undecylprodigiosin (Red), methylenomyacin (Mmy), and calcium-dependent antibiotic (CDA)).\cite{106,107} As the end-products can be toxic even to the producing organisms,\cite{108} antibiotic production is carefully coordinated in the bacterial population. One way *Streptomyces* can regulate secondary metabolite production is through the use of small diffusible molecules, known as $\gamma$-butyrolactones, in a manner analogous to acyl homoserine lactone (AHL)-based quorum sensing (QS).\cite{109} At this point, more than fifteen different GBLs have been identified in various Streptomyces species.\cite{108,110}

Recent studies and reviews\cite{112,113} point to the hypothesis that the GBL regulatory system (Figure 1.8) involves a small, yet complex two gene network composed by a synthase ($scbA$ (SCO6266)) and a butyrolactone receptor ($scbR$ (SCO6265)), which
The scbA gene product is presumed to catalyse the condensation of dihydroxyacetone phosphate with a beta–ketoacid to produce three different butyrolactones (SCB1, SCB2 and SCB3).\textsuperscript{108,114} On the other hand, scbR is a TetR-like\textsuperscript{115} DNA-binding protein known to regulate its own transcription and that of scbA. Additionally, ScbR protein has a $\gamma$-butyrolactone binding domain at its C-terminal and a DNA binding domain at its N-terminal\textsuperscript{116}. This structure allows it to directly regulate the production of a cryptic metabolite and indirectly the production of blue pigmented actinorhodin (Act) and prodigiosins (e.g. undecylprodigiosin, Red). These genes are divergently encoded and their promoter regions overlap 53bp, as has been shown by gel retardation assays and DNase I footprinting studies.\textsuperscript{117} Complementary to these findings, the discovery of cis asRNA has also been reported in different studies.\textsuperscript{118–120} The regulatory role of the system’s overlapping topology and of the antisense RNA in biological decision making has been demonstrated in number of studies concerning different bacteria,
such as *E. faecalis*\textsuperscript{[121]} and *S. enterica*.\textsuperscript{[122]} It is therefore possible that the formation of an antisense RNA between the *scbA* and *scbR* transcripts could play an important role in the regulation of the GBL system.

Transcription analyses have shown that both genes are mainly active during transition from logarithmic growth to stationary phase. It is presumed that SCBs slowly accumulate into the media and upon reaching a concentration threshold promote a coordinated switch-like transition to antibiotic production by binding to ScbR. However, the mechanism of this network is not fully defined, although several alternative scenarios have been proposed.\textsuperscript{[1,108,123]} The complete elucidation of this system could potentially lead to the design of robust and sensitive systems with significant applications as orthologous regulatory circuits in synthetic biology and biotechnology.\textsuperscript{[111]}

In 2008, Mehra *et al.*\textsuperscript{[1]} proposed a deterministic model involving a putative heterodimer of the two proteins (ScbR-ScbA) which would bind to a different site upstream of *scbA* and further activate its transcription. More recently, Chatterjee *et al.*\textsuperscript{[123]} proposed an alternative model of the system, where ScbA only catalyses the SCB biosynthesis and the promoter overlap and antisense RNA formation between *scbA* and *scbR* are the sole driving forces of the precise switch-like transition. During both previous modelling efforts on this system, the work was mostly focused on exploring the parameters which induced bistability in the system, without so much considering how closely the model’s output matches the experimental data.

This system therefore provides a good opportunity for the design and analysis of a model explicitly acknowledging uncertainty which will examine all possible scenarios within a range of reasonable parameters. The ultimate aim of this project was that the improved model would successfully describe the mechanism of the GBL regulatory system and allow reliable predictions of its behaviour. The work detailed in Chapter 4 was performed in close collaboration with experimental microbiologists who provided their expert advice in the biological aspects, as well as experimental data for the model validation.
1.7 Aims and Overview of the Thesis

The thesis is presented in journal format and most of the chapters are in the form of a manuscript. Where this is the case, it is clearly identified on the chapter title page and the nature of the authors contributions are also detailed there. The overview of the thesis is summarised in Figure 1.9. The thesis has two main aims: 1) the development of appropriate computational methods that will facilitate the application of ensemble modelling in the analysis computational models, and 2) the use of these methods to explore the γ-butyrolactone regulatory system and unravel its elusive mechanism.

In Chapter 2 a more “respectful” modelling framework in systems biology is proposed, discussing the importance of a realistic parameter estimation and a rigorous uncertainty consideration while modelling metabolic networks and signalling pathways.

Chapter 3 introduces a fast, reliable and robust protocol that will enable the design of log-normal probability distributions based on assessing the plausibility of information retrieved from literature, databases and experiments. These distributions can then be used as informative priors in an ensemble modelling framework. Furthermore, the protocol ensures the conservation of the thermodynamic consistency of the model.

In Chapter 4, the methods discussed in the previous chapters are put to use, first to recreate and comprehensively refactor the previously published model on the GBL system by Mehra et al.\cite{1} in the context of an ensemble modelling framework. The same was done for a published model on the quorum sensing system by Weber et al.\cite{2} as the structural similarities between the two systems enabled the comparison of their respective behaviours. The second part of this chapter focuses on the design and analysis of an improved multifunctional GBL model which considers the most important of the previously proposed mechanisms in various combinations while allowing the incorporation of parameter uncertainty. The new GBL model elucidates some important features of the system and even suggests it might have a different biological role than the one originally considered. Furthermore, it points to a number of future experiments that will further clarify the interactions between the system’s components.

Finally, a general discussion along with future research perspectives are presented in Chapter 5.
Figure 1.9: Overview of the thesis. The project started from two different directions: the development of methods for the explicit consideration of uncertainty and the exploration of the GBL system mechanism. These two were combined in order to design an improved, versatile tool that will enable accurate predictions with specified confidence levels that will guide future experimentation and engineering of microorganisms.
Blank page
Chapter 2

Respectful modelling: Addressing uncertainty in dynamic system models for molecular biology

Areti Tsigkinopoulou, Syed Murtuza Baker, Rainer Breitling*

Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, Faculty of Science and Engineering, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom.

*Corresponding author: Rainer Breitling, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom, Tel.: +44161 306 5117

Keywords: Ensemble modelling, uncertainty, distributions, biological systems, Bayesian statistics

Preface

“Respectful modelling: Addressing uncertainty in dynamic system models for molecular biology” was published in Trends in Biotechnology in June 2017. The draft manuscript was submitted on 28th of September 2016, revised on 5th of December 2016, accepted on 15th of December 2016, and published online on 13th of January 2017. AT and RB wrote the manuscript, SMB provided useful feedback and corrections. The published
manuscript has been formatted in a style consistent throughout the thesis, opposed to 
*Trends in Biotechnology* formatting. The position of figures and boxes within the text is 
as consistent as possible with the published manuscript.

**Abstract**

Although there is still some scepticism within the biological community regarding the 
value and significance of quantitative computational modelling, important steps are con-
tinually being taken in order to enhance its accessibility and predictive power. We view 
these developments as essential components of an emerging “respectful modelling” 
framework, which has two critical aims: 1) Respecting the models themselves and 
facilitating the reproduction and update of modelling results by other scientists. 2) 
Respecting the predictions of the models, and rigorously quantifying the confidence as-
associated with the modelling results. This respectful attitude will guide the design of 
higher-quality models and facilitate the use of models in modern applications, such as 
engineering and manipulating microbial metabolism by synthetic biology.

**2.1 Computational models in current research: Success 
and scepticism**

Quantitative computational models of cellular pathways and circuits are essential tools 
for generating clear, testable predictions about the behaviour of complex cellular ma-
chineries\[124,125\]. Their use is currently moving beyond the proof-of-concept stage to-
wards real-world applications, such as engineering and optimising biological microor-
ganisms to produce specific chemicals and biofuels\[126-128\], as has been shown most 
successfully for various terpenoids\[129\] and for succinic acid\[130\]. Their applications in 
identifying potential drug targets in metabolic or signalling pathways are also rapidly 
advancing\[131-133\]. An increasing number of success stories demonstrate that computa-
tional models have a lot to offer to biologists, from surveying cellular development\[134\] 
and exploring signalling pathways\[135\] and genetic circuits\[136\], to investigating potential 
treatments for cancer\[137,138\]. However, modelling is not yet part of the mainstream
of biological practice, even in fields like synthetic biology that intend to embrace an engineering approach to manipulating biological complexity\textsuperscript{[139–141]}. 

Major progress is also being made in the case of genome-scale models\textsuperscript{[142–144]}, which match enzyme-coding genes with predicted reactions in metabolic pathways. Until recently, their nature limited the application of genome-scale models to stoichiometric constraint-based approaches, studying the fluxes through the system without being able to predict metabolite levels or the dynamics of responses to internal or external perturbations. However, efforts are being made to incorporate detailed information on enzyme regulation and kinetic mechanisms into such constraint-based models, thus transforming them into genome-scale kinetic models of metabolism\textsuperscript{[145–147]}. 

We aim to outline (Figure 2.1) “respectful modelling”, an emerging set of closely related concepts and techniques that together enhance and advance earlier modelling approaches to make computational models both more approachable and more relevant for the work of experimental biologists.

Many molecular biologists still doubt the relevance of computational models of cellular systems. One sometimes even hears the truism that “All models are wrong, but some are useful”\textsuperscript{[148]}, cited as if it meant “all models are false and most are useless”. In fact, models are meant to be explicit descriptions of our implicit knowledge about the function of a biological system. Therefore, they are ideally no more wrong than less-formal reasoning about the same system, yet they offer numerous advantages: most importantly, they put any assumptions about how a cellular mechanism works out into the open for everyone else to check and criticize, and they allow predictions about non-obvious (“emergent”) behaviours of a system that follow compellingly from these assumptions, even when the complexity and non-linearity of the systems make simple back-of-the-envelope arguments impossible.

Yet, beyond the anecdotal observation that models are sometimes flippantly dismissed as “always wrong”, there are more serious indications that computational models of cellular pathways and circuits are not yet fully respected as scientific tools in molecular biology. For example, complex systems models are very rarely updated and developed further; with a few rare exceptions, such as some constraint-based models\textsuperscript{[149,150]}, there are not many examples of “versioned” and iteratively improved models. Moreover,
predictions of computational models in systems biology are usually reported without error bars. In contrast, the incorporation of uncertainty in models has long been common practice in other fields\cite{151,152}, from natural hazard insurance to economic forecasts, where the predictions and their associated confidence make a real difference to people’s lives. For example, in the prediction of human-induced climate change, error bars and the exploration of multiple scenarios are the dominant feature in all predictions, to the extent that they are even presented and discussed in the popular press\cite{153}.

2.2 A respectful approach to biological modelling

So, what is needed to establish the same “respectful” attitude to modelling in biology, so that it can meaningfully contribute to areas such as synthetic biology or personalized medicine? Two aspects seem to be central:
1. Respect for the model itself as a resource that can grow and develop. If we respect our models, we should not treat them as one-off exercises. Instead, we need to make them understandable and reproducible by other scientists and update them regularly and continuously rather than discard them once new discoveries are made (or a new PhD student takes on the project). Therefore, reaction mechanisms and parameter values (such as assumed rate constants, enzyme concentrations, or substrate affinities) need to be documented, so that others can build on our work, and models that capture the full breadth of alternative hypotheses need to be maintained, instead of fixating on one “maximum likelihood” or preferred model. In this way, the model can improve iteratively as new data become available, instead of having to be rebuilt each time.

2. Respect for the output of the model and for the predictions it generates. We should strive for accurate predictions that can be used for rational decision-making. When respecting our models, in the sense that we really care about their predictions, we need to care about the associated level of uncertainty, and we need to identify the different levels of confidence associated with each alternative prediction. Otherwise, we could not place reasonable bets on different possible outcomes. Therefore, we need model predictions with confidence intervals, correctly reflecting our current (limited) knowledge about the topology and dynamics of the biological system. We can only ignore confidence intervals as long as we are not interested in the real-world implications of our quantitative model predictions, i.e., when we only treat them as little more than random numbers that illustrate the elegant model-building and analysis algorithms. Once we have a true stake in generating accurate quantitative predictions, because we want to apply them in the design of actual biological experiments or the engineering of real living systems, we will listen much more carefully to what the models have to tell us. Looking for the uncertainty in the model predictions is not a sign of disrespectful mistrust, but an immediate consequence of taking the model serious as a predictive, rather than illustrative, tool.
Chapter 2. Respectful modelling

In trypanosome parasites in the human bloodstream, energy metabolism is mostly restricted to glycolysis, which takes place in unique, specialized organelles, the glycosomes; its enzymes are considered promising targets for newly developed drugs against sleeping sickness.

The steps followed during the model design and analysis are depicted in Figure 2.2. Starting from a highly curated model of trypanosomal energy metabolism, the experimental uncertainty of every enzyme kinetic parameter was determined by an extensive exploration of the original literature. Parameter sources, evidence for alternative model topologies (extra reactions), as well as any calculations (e.g., for the parameter means and standard deviations) were documented in a dedicated Wiki-based database (A). Several versions of the model were created, in which alternative groups of metabolites with different molecular weight could freely diffuse across the membrane, representing the uncertainty about possible glycosome permeability resulting from recent evidence that the organelle membrane contains non-selective pores (B).

For each of the alternative models, ranging from very tight to very leaky glycosomes, plausible combinations of parameter values were sampled according to the documented uncertainty by using a random number generator in accordance with each parameters assumed probability distribution. This resulted in a large collection of model variants, each using a different set of parameter values (C). The ensembles of models were then subjected to the same types of analysis as traditional dynamic models, e.g. determining steady-state concentrations of metabolites and calculating the control coefficients of enzymatic reactions that could identify the most promising drug targets (D).

Figure 2.2: Flow chart of steps followed during the dynamic modelling of *T. brucei* energy metabolism with explicit consideration of parameter uncertainty. [89,154,155]
Chapter 2. Respectful modelling

The analysis of the resulting ensembles of model predictions provided several interesting insights that had remained “hidden” in classical maximum-likelihood analyses of individual models. For example, it revealed unexpected fragilities in the existing models. Two metabolites, 3-phosphoglycerate and pyruvate, seemed to accumulate to impossible concentrations in many of the models, indicating that critical regulatory loops are probably still unaccounted for in our current understanding of trypanosome metabolism. Moreover, control of glycolytic flux seemed to be more widely distributed between several key reaction steps, rather than being largely restricted to the rate-limiting glucose uptake transporter. The results also showed that the models that predicted steady-state metabolite concentrations and fluxes that most closely matched experimental observations, for the largest number of plausible parameter sets, were the ones with glycosomes permeable to small metabolites up to the size of fructose 6-phosphate and fructose 1,6-bisphosphate. This result challenges the current consensus view of trypanosome glycolysis that critically depends on highly controlled trans-membrane fluxes but is in good agreement with the presence of recently discovered unspecific glycosomal pores.

In fact, these two aspects are closely related. They both imply that our models should not be rigidly fixed but need to capture alternative scenarios, alternative parameter values\(^\text{156}\), alternative circuit topologies\(^\text{157}\), and generally alternative hypotheses about the biological system that is being studied. A recent study on the energy metabolism of the protozoan parasite Trypanosoma brucei,\(^\text{90,154}\) the causative agent of sleeping sickness, illustrates how an explicit treatment of uncertainty can yield new insights even for well-studied model organisms (Box 2.1). The “respectful modelling” data identify a number of previously unexplored key experiments. In the trypanosome example, for instance, the exact level of permeability of the glycosomes for specific glycolytic intermediates could be determined by a more targeted measurement, or the sensitivity of trypanosomes to the inhibition of glycosomal enzymes with unexpectedly high control coefficients (i.e. triose phosphate isomerase) could be directly measured.

Therefore, these two aspects represent different levels of alternatives: on one hand, the degree of confidence in our predictions can be determined by examining ensembles of equally plausible models. On the other hand, models can be updated and re-used as our biological understanding evolves, by changing our assumptions about which models (and parameter values) are plausible and which are not.

The advantages of respectful modelling are obvious:

1. It quantitatively indicates each prediction’s robustness, which is helpful for identifying the most suitable experiments to carry out next: do we already have enough
confidence in a prediction to build an expensive study around it, or are there specific uncertainties that we first need to reduce by targeted measurements?

2. It enables collaborative and iterative work on model building and model improvement by explicitly identifying and documenting alternative model structures and parameter values.

3. It allows managing alternative hypotheses about the functions of complex biological systems within a unified modelling framework, instead of having slightly different, incompatible, and often incomparable models associated with each hypothesis.

4. Finally, and perhaps most importantly, it makes modelling approaches available for systems in which quantitative information is incomplete and uncertain, thus unveiling otherwise inaccessible biological phenomena. This may also help bridging the conceptual gap between quantitative and qualitative modelling\[158\].

### 2.3 Respectful modelling showcases and applications

Several important steps have recently been taken to provide the ingredients for a respectful modelling approach.

1. Model descriptions are being standardised. This practice is becoming applied to every aspect of the model building process, from file formats to variable names and equation structures and to mandatory archiving of the resulting models. The Systems Biology Markup Language (SBML)\[159\], an XML-based machine readable language, is supported by almost all software tools in the field and is prevailing as a model representation format that facilitates model sharing and replication. Additionally, models are starting to comply with some basic community guidelines that describe the minimum information that needs to accompany a published model (MIRIAM)\[42\] along with conventions that facilitate data integration in model building, such as SBtab\[160\]. Moreover, centralised model repositories such as BioModels\[161\], JWS Online\[162\], CellML Model repository\[163\] and the
BiGG database\textsuperscript{[164]} have been created to enable the distribution and curation of biological system models.

2. Parameter information is being rigorously documented. The inclusion of supplementary material on the sources and values of the parameters (e.g., enzyme and substrate affinities, transcription rates, etc.) in published models is increasing\textsuperscript{[165–168]}. For example, recently published updated models of central carbon metabolism in trypanosome parasites included a Wiki page dedicated to each biochemical reaction of the system\textsuperscript{[90,154,169]}. In this way, detailed information on the sources for each parameter value was provided, along with descriptions of the underlying calculations and assumptions, as well as alternative model versions with different topologies. This approach not only increases the accountability of modellers for the critical decisions made during model building; it also greatly facilitates the reconstruction, validation and updating of models by successive generations of researchers.

\textbf{Box 2.2: Generating probability distributions to describe uncertainty}

Describing the uncertainty associated with our knowledge of parameter values in molecular systems models is challenging.

The natural choice for the shape describing the range of plausible parameter values is often a log-normal distribution: there will be a most likely value (the mode of the distribution), negative values are not allowed, and the distribution is symmetrical, in the sense that values that are \( x \) times larger than the most likely estimate, are just as plausible as values that are \( x \) times smaller (Figure 2.3). More specifically, the mode is the value \( x_o \) for which the condition \( f(x_o \ast \delta) = f(x_o) \) is fulfilled for all real numbers \( \delta \) (where \( f \) is the probability density function).

In contrast to common assumptions, the log-normal distribution is ubiquitous in nature across different fields and often a far better description of the data than a normal distribution. This is particularly seen in the medical and biological sciences, when the mean values are low and the
variances are large, yet the values cannot be negative, whether they be latency periods of infectious diseases in epidemiology, species abundances in ecological studies, or enzyme kinetic parameters as in the present discussion [56]. The reason for this predominance of log-normal distributions is simple: when the change in a variable (observable) at each step of the process is proportional to its current value (i.e., when the process follows Gibrat’s law of proportionate effect), the resulting observations will be log-normally distributed [57]. In the life sciences, biochemistry, biophysics and population ecology are disciplines where processes characterized by Gibrat’s law are most obvious: reaction velocities, surface and volume measurements, and population growth (also at the molecular level) are regulated by factors that act in a multiplicative way, rather than in the additive way that would be required for data to become normally distributed.

The challenge for the biological systems modeller is to decide on the appropriate values describing the distribution: what is the most likely value (described by the mode of the distribution) and how rapidly does the plausibility of the values decrease when moving away from this value (described by the standard deviation). These in turn determine the location and scale parameters $\mu$ and $\sigma$ of the log-normal distribution. In many cases, a good estimate of the most likely parameter value exists, e.g. from actual experimental measurements. In other cases, related parameters have been measured, e.g., the kinetics of similar enzymes or even of the same enzyme in different species or conditions. In the extreme case, no measurements are available at all; but even in that case, an informed guess is usually possible. For instance, the most plausible $K_m$ value of an uncharacterized novel enzyme might be the average of all $K_m$ values ever recorded in a comprehensive enzyme database like BRENDA, which comprehensively records available experimental data on enzyme kinetic parameters from all domains of life. The spread of the distribution, its standard deviation, would in such a case reasonably be determined by the range of all reported values (Figure 2.4).

It is also necessary to decide if there are any hard (biophysical) thresholds, beyond which a parameter value is not only implausible, but impossible; in other words, if the distribution should be truncated at the extremes. In every case, it is important to remember that determining plausible parameter distributions is not a mechanical exercise but has to be based on actual biological and biophysical reasoning. If there are arguments to support the idea that a given log-normal distribution does not match the expected range of plausible values, it needs to be adjusted accordingly.
3. Moreover, the detailed documentation of parameter sources enables the quantification of the uncertainty associated with each value, so that model predictions can take these uncertainties into consideration intuitively. A formal acknowledgement of uncertainty in computational models of biological systems not only describes the modellers’ beliefs and confidence in model structure and model parameters, but it also explicitly identifies alternative structures and parameter values and their associated plausibility\cite{90,154}. This approach mirrors the process of scientific progress by continuously exploring alternative hypotheses. Thus, rather than paralyzing the model’s analysis, the explicit acknowledgement of uncertainty actually enables the flexible evolution of biological models, which at each stage honestly represent our current knowledge about a biological system. This is in contrast to traditional modelling approaches, where alternative hypotheses and assessments of uncertainty are managed only in an ad hoc process taking place implicitly in the brains of the modellers and their collaborating expert biologists. Furthermore, acknowledging uncertainty and integrating it during the model building phase allows making predictions that are associated with specified confidence intervals, which can guide further experimentation\cite{170,171}.

4. To transition from the collected information about parameter uncertainty to the resulting quantitative assessment of our confidence in specific model predictions, data-driven parameter sampling strategies have been formulated\cite{63,90,154,155,169}. They employ informative distributions to describe what the modellers (and their biologist collaborators) consider as plausible values for each parameter. These distributions correspond to the priors in a Bayesian statistical framework\cite{172,173}, and can indeed be used for a Bayesian statistical analysis\cite{174,175} to update the parameter values when new experimental evidence becomes available, but they most importantly can be used in the next step to create an entire ensemble of plausible models by sampling values for each parameter from its corresponding distribution (Box 2.2). Experimental data, biological background knowledge, and biophysical plausibility (Box 2.3) can all contribute to defining the most appropriate distributions that capture our current state of knowledge accurately, not exaggerating the
uncertainty, but also not being overconfident about specific values or connections in the network.

**Box 2.3: Thermodynamic consistency**

Once the plausible values for each parameter have been described, there is another important factor that needs to be considered to decide if combinations of parameters are plausible: thermodynamic consistency. For example, imagine a reaction that is known to have an equilibrium constant very close to 1; i.e., its standard Gibbs free energy $\Delta G^o = 0$. We are trying to determine the kinetic parameters for the forward and backward component of this reaction, and we don’t know much about the rate of the reaction, so we sample each of the two parameters from a very broad distribution. If we do not take the additional thermodynamic information into account, we will often end up with sampling the forward reaction rate from the “fast” end of the spectrum, and the backward rate from the “slow” end, or vice versa. Thermodynamic consistency requires that we discard such samples and only keep samples where the two reaction rates are very similar (how similar will in turn depend on our uncertainty about the equilibrium constant) (Figure 2.5).
Figure 2.5: Sampling strategy for maintaining thermodynamic consistency of parameter sets. In order to ensure the thermodynamic feasibility of a parameter combination in the case of interconnected parameters (i.e. forward and backward reaction, $k_{on}$, $k_{off}$) a bivariate distribution is created. When the two marginal distributions are non-correlated, the generated points that represent parameter pairs form a circle. When the equilibrium dissociation constant, $K_D$, is exactly known, the parameters $k_{on}$ and $k_{off}$ are tightly correlated through the $K_D$ value, and the points form a straight line. Finally, if $K_D$ is approximately known (a distribution of values for $K_D$ exists) the resulting points of the bivariate system form an ellipse. The thickness and orientation of the ellipse depend on the magnitude of the correlation between the two marginal distributions and on the degree of uncertainty on the values of $k_{on}$, $k_{off}$ and $K_D$. This case represents the realistic scenario when modelling, as usually the parameter values are approximately known. The first case (no correlation) does not respect thermodynamic consistency and is therefore undesirable. The second case, although taking into account the dependency of the two parameters, is in most cases unrealistic since the value of a parameter such as the equilibrium constant is rarely exactly known.

5. Approaches that explicitly acknowledge model uncertainty have been developed and employed\cite{156}, such as Markov Chain Monte Carlo (MCMC) methods\cite{176–178}, ensemble modelling\cite{179,180}, and global sensitivity analysis\cite{91,172,181}. Such techniques are based on the concept of sampling parameters from their associated probability distributions and thus creating a collection of models can undergo
further analysis in a similar way as typical dynamic models. In contrast to analyses of the local effects of variable parameter values around their preferred value, the entire resulting landscape of solutions can be surveyed without focusing only on one optimal solution, which may in the future be rejected once new data become available. This approach is also highly preferable over methods that try to fit a single “maximum likelihood” set of parameters based on the best match to experimental data\[174\]. Such a “fitting” strategy is very popular in applications including transcriptional dynamics\[182,183\], epigenetics\[184\], neuronal dynamics\[185\], and population-level epidemiology\[186\].

![Image](image_url)

*Figure 2.6: By focusing only on the “best fit solution” a modeller might end up trapped in a local optimum and thus miss interesting alternatives which in the future might be supported by new experimental data. Instead, by surveying the entire landscape of solutions without focusing on one particular peak, respectful (Bayesian) modelling remains adaptable to future developments, especially when studying complex systems with rugged likelihood surfaces, where the availability of new experimental data can result in rapid changes in the relative likelihood of alternative model scenarios. (Image courtesy of Danai Triantafyllopoulou)*

However, fitted parameters are also very likely to be trapped in transient and spurious global optima, given that the likelihood surface for complex biological systems is inherently extremely rugged (highly non-convex), and are always in
danger of leading to overfitting and a failure to identify alternative hypothesis that could explain the results equally well\cite{124,156}. Moreover, parameter values, once fitted, are rarely updated in the face of new experimental results. Avoiding fitting parameter values according to the maximum likelihood is particularly important in the case of biological models: their complexity and non-linearity lead to a highly non-convex (i.e. rugged) likelihood surface, which means that models with considerably different sets of parameter values can have quite similar likelihoods (Figure 2.6). If one of them is preferred based on a momentarily higher likelihood, alternative options are easily overlooked later on. In order to avoid this pitfall, all options are kept in view in a respectful modelling approach, so that the model can be easily adapted. In this regard, and in important details of the parameter distributions\cite{187}, the respectful modelling approach closely approximates the Bayesian inference processes supposedly implemented in the human neocortex\cite{188}.

2.4 Conclusion

It is probably clear by now that all of these developments offer reasonable and useful modifications of the way we build and analyse computational models for molecular biology. But why will they make a difference? They get us closer to the ideal expressed in another famous saying about computational modelling: “Models are not meant to be descriptions, pathetic descriptions, of nature; they are designed to be accurate descriptions of our pathetic thinking about nature” (J. Black, cited in\cite{189}). To achieve this ideal, each model has to allow us to capture the “pathetic” aspects of our thinking, the uncertainties and incompleteness of the evidence, and to evolve as our thinking evolves on the basis of new experiments\cite{190,191}.

In the future of molecular systems biology and modelling, nothing is certain except uncertainty itself. The increasing use of automated model building strategies will only increase the challenge\cite{192}, as models grow in size and the specific refinement of individual parameters by targeted experiments becomes even less feasible than it is now. The adoption of a respectful modelling framework will promote and facilitate collaboration
within the biological community, stimulating the use of models as a tool for fundamental research, but also as a valuable guide for the predictive engineering of biological systems and their informed manipulation by increasingly personalized drugs.
Chapter 3

Defining informative priors for ensemble modelling in systems biology

Areti Tsigkinopoulou¹, Aliah Hawari¹, Megan Uttley², Rainer Breitling¹*

¹ School of Chemistry, University of Manchester, Manchester, United Kingdom
² School of Health Sciences, University of Manchester, Manchester, United Kingdom

*Corresponding author: Rainer Breitling, Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom, Tel.: +44161 306 5117

Keywords: Ensemble modelling, uncertainty, informative priors, biological systems, Bayesian statistics

Preface

“Defining informative priors for ensemble modelling in systems biology” has been submitted to Nature Protocols on 7th of September 2017 and it is currently under review. AT designed and developed the mathematical strategy and the MATLAB functions of the protocol and wrote the manuscript. AH designed the first step of the protocol concerning the criteria of the assignment of weights, performed tests with diverse case studies and provided feedback for the improvement of the protocol and the manuscript. MU tested the protocol and provided insightful advice on the improvement of the manuscript and the computational functions. RB supervised the project and wrote the manuscript.
Abstract

Ensemble modelling is a powerful addition to the systems biological toolbox, providing the ability to use quantitative kinetic models in situations of incomplete and uncertain knowledge of system parameters, while systematically assessing confidence in model predictions. The application of ensemble modelling requires the objective and reproducible translation of parameter data into probability distributions with informative shapes, as well as approaches to sample parameters from these distributions without violating the thermodynamic consistency of the overall model. Although a number of pioneering frameworks for ensemble modelling have been published recently, the issue of generating informative priors has not been addressed yet. Here we present a protocol which aims to fill this gap, by enabling the formal consideration of uncertainty in the form of probability distributions. The protocol concerns the collection of parameter values from a diverse range of sources (literature, databases and experiments), assessing their plausibility, and creating log-normal probability distributions, which can be used as priors in an ensemble modelling framework. Furthermore, the protocol enables sampling from the generated distributions while maintaining the thermodynamic consistency of the model. The aim of this novel protocol is to facilitate the design and use of informative distributions for ensemble modelling in fields such as synthetic biology and systems medicine, where a quantitative and rigorous assessment of the confidence of model predictions is necessary to guide the efficient manipulation of complex biological systems.

3.1 Introduction

Kinetic biomolecular models, such as mechanistic models of metabolism or cellular signalling pathways, require an accurate knowledge of all kinetic parameters in the system in order to generate predictive outputs. However, the available experimental information on these parameters is often incomplete or uncertain. To address this challenge and to allow predictive model building in the face of parameter uncertainty, recent publications have advocated the use of “ensemble modelling” strategies\(^{[193–197]}\), where reasonable parameter values are sampled from probability distributions of plausible values.
(derived from experimental information). This results in an “ensemble” of predictions that accurately reflect the confidence limits for each model prediction. This approach is already being successfully used in different fields, such as economics\cite{198}, engineering\cite{199} and climate predictions\cite{200}, but is still not prevalent in biology, although considerable progress is being made in this direction\cite{201–203}. In contrast to established “parameter fitting” approaches, ensemble modelling has only recently started gradually becoming supported by systems biology modelling tools, thus implementing this conceptually straightforward approach in a rigorous and reproducible manner has been a challenge for biologists.

One of the most prominent problems concerns the definition of the probability distributions describing parameter uncertainty. Recent contributions have achieved remarkable results in the design of robust and reliable frameworks that will promote the use of ensemble modelling in systems biology\cite{146,204–206}, such as toolkits and methods for parameter inference and model selection\cite{207–210}, generation of probability distributions,\cite{211} and ensuring thermodynamic consistency while sampling\cite{99–102}. However, the rigorous definition of the informative prior distributions that would be used as an input into such frameworks has not yet been addressed in the context of systems biology, although there is a common consensus on its importance. Informative priors are necessary for a full Bayesian analysis, as they allow the modellers to make inferences about the system under investigation and enable model checking. Although non-informative priors may appear as more objective, they can lead to (mathematically) improper posteriors and thus impede the model analysis\cite{212}, and they unavoidably involve an intentional disregard for available experimental information.

Among available informative priors, there has recently been a shift from uniform (flat) or normal priors to log-normal distributions, which more accurately describe the uncertainty associated with our knowledge about biological kinetic parameters\cite{213–216} (more extensively discussed in Chapter 2). Unfortunately, in contrast to standard normal distributions, log-normal distributions are often far less intuitive to work with. What is the most likely value? How does the plausibility change when using an alternative parameter value? Unlike for normal distributions, there is no direct link between the distribution’s characteristic parameters (i.e., $\mu$ and $\sigma$) and the mean, median and mode
of a log-normal distribution. This creates challenges when translating biologists’ beliefs regarding the plausibility of parameter values into the shape of the desired distribution, which is further compounded when limited (or no) experimental information on a parameter value is available.

Another issue that ensemble modellers have been facing and trying to address is maintaining the thermodynamic consistency of the system when sampling uncertain parameters. For example, when the forward and reverse rate of a reversible reactions are sampled independently, many combinations of values will potentially be inconsistent with our knowledge about the position of the thermodynamic equilibrium of the reaction\[63,216\], e.g. because they are sampled from opposite extremes of the respective distributions. In cases like this, the values of interdependent parameters need to be consistent with the common parameter (e.g., the equilibrium constant) and therefore should not be sampled separately. The modeller faces the challenge of having to ensure that implausible combinations of parameters are avoided or discarded, a non-trivial task which increases the complexity of the model building process\[63\].

This protocol aims to address these issues by facilitating and partially automating the process of defining informative priors for ensemble modelling. In particular, we are focusing on encoding the quality of the information retrieved from a variety of sources in accordance with the modeller’s beliefs about the reliability of each experimental value. The protocol outlines a number of standard steps to follow in order to collect information from the literature and databases, to systematically rank this information according to its relevance for the model under construction, and to use it to define log-normal distributions which can be used for sampling parameter values. As an optional step, the protocol addresses the issue of thermodynamic consistency and dependent parameters, by detailing the procedure to sample simultaneously from two or more interdependent distributions under constraints.

This protocol offers a procedure to capture parameter information in probability distributions in a rigorous and standardized way. It also allows the modeller to translate their beliefs and knowledge about the parameter values in a straightforward and intuitive way into probability distributions. Therefore, through this protocol, informative parameterisation for ensemble modelling is made accessible to experimental biologists.
interested in applying these state-of-the-art methodologies, who might be intimidated by the apparent complexity of the process. At the same time, the protocol is useful for computational biologists who wish to improve the information content of their models using informative priors, to optimise the outputs of existing ensemble modelling frameworks. Although the protocol is mostly focused on biological modelling, its concepts could also be applied to ensemble modelling in other disciplines.

### 3.1.1 Applications of the protocol

An earlier version of this protocol has been successfully implemented in a recent study on the energy metabolism of the protozoan parasite *Trypanosoma brucei*[^90^154]. As a result, a number of key experiments were identified, thus highlighting the benefits of an explicit consideration of uncertainty when modelling. The new protocol includes the following improvements:

- Improved systematic evidence weighting scheme;
- Facilitated definition of the Mode and Spread of parameter values based on user beliefs;
- Standardized visual validation, encouraging users to manually modify and correct the distributions once they have been generated;
- Integration of systematic approach to ensure thermodynamic consistency.

### 3.1.2 Overview of the protocol

The protocol aims to translate biological knowledge into informative probability distributions for every parameter, without the need for excessive mathematical and computational effort from the user. It assumes that a functioning kinetic model of the relevant biological system is already available and focuses on the procedures specific to the application of an ensemble modelling approach. The major protocol steps are the following (as detailed in Figure 3.1):

- Collection and documentation of parameter information from the literature, databases and experiments.
Chapter 3. Defining informative priors for ensemble modelling in systems biology

Figure 3.1: Workflow of the protocol from the parameter collection to the generation of probability distributions and thermodynamic consistency for interconnected parameters. (*Abbreviations: CDF: Cumulative Density Function; erf: error function; N/A: not available; SD: standard deviation*

- Plausibility assessment of each value collected from the literature using a standardized evidence weighting system. This includes assigning weights with regards to the quality, reliability and relevance of each experimental data point.

- Calculation of the *mode* (most plausible value) and *Spread* (*multiplicative standard deviation*) using the weighted parameter data sets. These values are found using a function which calculates the weighted median and weighted standard deviation of the log transformed and normalised parameter values.

- Calculation of the location (*\( \mu \)*) and scale (*\( \sigma \)*) parameters of the log-normal distribution based on the previously defined *mode* and *Spread*.

- If required, ensure the thermodynamic consistency of the model by creating multivariate distribution for interdependent parameters (e.g., the rates of forward and reverse reactions are connected via the equilibrium constant of the reaction, and must not be sampled independently).
In order to facilitate and automate the necessary calculations and model constructions, annotated computer scripts have been created for the most challenging steps and are provided as supplementary materials.

### 3.2 Materials

- Parameter values retrieved from databases and literature
- A standard personal computer
- A numerical computation and visualization software (for example R\textsuperscript{[217]} or Matlab\textsuperscript{[218]})

### 3.3 Procedure

#### 3.3.1 Step 1: Find and document parameter values

A. Collect and document all available parameter values from literature, databases (e.g., BRENDA\textsuperscript{[219]} and BioNumbers\textsuperscript{[220]}) and experiments (hypothetical example displayed in Table 3.1).

i) **No values:** If no values can be retrieved for a reaction occurring in a particular organism, broaden the search to include data from phylogenetically related organisms (the uncertainty should then be adjusted accordingly). When no values can be retrieved for a certain parameter, the broaden the search constraints; e.g., if the Michaelis–Menten constant, $K_M$, for a particular substrate is unknown, it might be reasonable to retrieve all $K_M$ values from a database like BRENDA\textsuperscript{[219]} in order to determine the range of generally plausible values for this parameter. In cases where no parameter values are available, an indirect estimate is often possible, based on back-of-the-envelope calculations\textsuperscript{[221]} (Figure 3.2) and fundamental biophysical constraints, which will have a larger uncertainty than a direct experimental
measurement but will still be informative-situations where absolutely no information about a plausible parameter range is available will be extremely rare.

ii) **One value:** If only a single experimental value will be available for a particular parameter, use this value as the Mode (most plausible value) of the probability distribution. Sometimes, an experimental error (e.g., a standard deviation) is reported for the value and can be used to estimate the Spread of plausible values; if this is not available an arbitrary multiplicative error can be assigned, based on a general knowledge of the reliability of a particular type of experimental technology. This information, of course, can be complemented using the previous strategy of retrieving related values from databases to optimise the estimate of the Spread of plausible values and avoid over-confident reliance on a single observation. Proceed to **Step 4**.

iii) **More than one value:** Whenever possible, it is important to collect as many experimental data with some implications for the range of plausible values, including those from a diverse range of methodologies, experimental conditions and biological sources. These will be pruned and weighted in the next step. Proceed to **Step 2**.

B. Document the reported experimental uncertainty (if available), in addition to the actual parameter value (Table 3.2).

i) **Uncertainty is reported in literature:** Uncertainty is often reported as an (additive) standard deviation (SD) or a standard error of the mean (SEM). As the functions employed in the subsequent steps require standard deviations as inputs, reported standard errors should be converted to standard deviations for consistency. This can be achieved by multiplying by the square root of the sample size ($n$), as per the equation: $SD = SEM \cdot \sqrt{n}$. In cases where the sample size is not reported, a conservative value of $n=3$ can be assumed.

ii) **Uncertainty is not reported in literature:**
Figure 3.2: Example of a back-of-the-envelope approach for estimating plausible parameter values. Calculation of the translation rate for an *E. coli* 310 aa size protein (i.e. thrB, homoserine kinase) when no information can be found on the actual translation, but there are available values for *E. coli* RNA transcription rates. By assuming that the translation rate is 1/3 the transcription rate, a rough estimation can be made in order to generate the parameter probability distribution. When compared to the actual values of *E. coli* the translation rates as reported in the literature (0.032 - 0.068 s⁻¹), the calculated range (0.029 - 0.097 s⁻¹) shows a surprisingly good match. The green highlighting shows the area of values covered by the reported literature values.\[220,221\]

a) If the modeller is absolutely certain about a parameter value, the standard deviation can be set to zero. Although it is extremely rare that there is no uncertainty about a parameter at all, this option can be useful when designing the backbone topology of a model, prior to performing ensemble modelling, or when parts of the model are to be kept constant in all simulations.

b) If the modeller has no information about the uncertainty of a reported parameter value, an arbitrary standard deviation can be assigned (e.g., the script provided for the next step assumes a default multiplicative standard deviation of 10% in such cases).
Chapter 3. Defining informative priors for ensemble modelling in systems biology

Table 3.1: Hypothetical example of documentation for hexokinase parameters in chicken liver cells retrieved from the literature

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Reported values</th>
<th>Uncertainty</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_M$ (μM)</td>
<td>0.5</td>
<td>N/A</td>
<td><em>J. Poultry</em> 2(3),2002</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>±0.2 (S.D.)</td>
<td><em>Quack Res.</em> 4(6),1998</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>N/A</td>
<td><em>Antarctica Biol.</em> 1(2),2004</td>
</tr>
</tbody>
</table>

*Note:* Although uncertainty is usually reported in an additive form ($x \pm SD/SEM$), theoretical biophysical considerations indicate that multiplicative errors often provide a better description about our actual uncertainty about experimental data in biology. For example, additive errors can imply that the confidence interval includes negative values. Thus, whenever possible, multiplicative standard deviations (Value $\times \div SD$) should be recorded. The script provided for the calculation of the probability distributions in Step 2 can handle both additive and multiplicative standard deviations. In order for the script to be able to distinguish between additive and multiplicative uncertainty, an extra column is added in the input file, where the uncertainty type is reported. This can be either 0 (additive standard deviation) or 1 (multiplicative standard deviation).

Table 3.2: Standardized description of parameter information in Table 3.1, for use in Step 2

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Reported values</th>
<th>Standard Deviation</th>
<th>Multiplicative/ Additive</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_M$ (μM)</td>
<td>0.5</td>
<td>1.1</td>
<td>Multiplicative (1)</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>0.2</td>
<td>Additive (0)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1.1</td>
<td>Multiplicative (1)</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>4.7434</td>
<td>Additive (0)</td>
</tr>
</tbody>
</table>

3.3.2 Step 2: Judge plausibility of values and assign weights

When multiple parameter values are reported in the literature, assign appropriate weights according to their source. For the construction of larger models it can be helpful to use a standardized weighting scheme, to increase the internal consistency and reproducibility of the process. The weights are arbitrary, but not in the sense of “random” and “inconsistent”, but rather in the sense that they consistently capture an expert “arbiter’s” opinion and experience regarding the relative reliability of individual types of data. For
example, standard weights can be assigned by answering a standard set of questions on
the origin and properties of each parameter value:

- Is the value derived from an \textit{in vitro} or \textit{in vivo} experiment, and is this the same
  environment as in the experiment that will be simulated?

- Is the value derived from the modelled organism, from a phylogenetically related
  species, or from a completely unrelated system?

- Does the value concern the same enzyme (or protein / gene / receptor etc. de-
  depending on the case), a related class/superclass of enzyme, or a generic enzyme
  completely unrelated to the one used in the model?

- Are the experimental conditions (pH, temperature) the same, closely similar, or
  completely different from the ones simulated in the model?

According to the answer to each question, a standard partial weight can be assigned,
and these weights can be multiplied to yield a total evidence weight for each experi-
mental value (see Tables 3.3 and 3.4).

\textbf{T}able 3.3: Documentation of the experimental conditions reported in the literature
for the hypothetical example in Table 3.1 and Table 3.2

| Parameter values $K_M$ (μM) | Experimental Conditions
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{In vivo}/ \textit{in vitro}</td>
<td>Organism</td>
</tr>
<tr>
<td>\textit{In vitro}</td>
<td>Chicken</td>
</tr>
<tr>
<td>\textit{In vivo}</td>
<td>Duck</td>
</tr>
<tr>
<td>\textit{In vitro}</td>
<td>Penguin</td>
</tr>
<tr>
<td>\textit{In vitro}</td>
<td>Crocodile</td>
</tr>
</tbody>
</table>

\textbf{T}able 3.4: Calculation of standardized weights from the information in Table 3.3

<table>
<thead>
<tr>
<th>Parameter values $K_M$ (μM)</th>
<th>\textit{In vivo}/ \textit{in vitro} same as model?</th>
<th>Organism (same/related/unrelated)</th>
<th>Enzyme</th>
<th>Conditions (pH &amp; temperature)</th>
<th>Total Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>\textit{In vitro}</td>
<td>4 (Same)</td>
<td>4 (Same)</td>
<td>4 (Same)</td>
<td>64</td>
</tr>
<tr>
<td>4.4</td>
<td>\textit{In vivo}</td>
<td>2 (Related)</td>
<td>2 (Related)</td>
<td>1 (Unrelated)</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>\textit{In vitro}</td>
<td>2 (Related)</td>
<td>4 (Same)</td>
<td>2 (Related)</td>
<td>16</td>
</tr>
<tr>
<td>20</td>
<td>\textit{In vitro}</td>
<td>1 (Unrelated)</td>
<td>2 (Related)</td>
<td>4 (Same)</td>
<td>8</td>
</tr>
</tbody>
</table>
3.3.3 **Step 3: Calculate the Mode and Spread of the probability distribution**

A. Create a final input table which will include the parameter values, errors and weights that have been documented in Steps **Step 1** and **Step 2** (as shown in Table 3.5).

*Note:* When no uncertainty estimate (e.g. standard deviation or standard error) is provided for an experimental parameter value, a multiplicative standard deviation of 10% (multiplicative factor of 1.1) is automatically assigned to the parameter by the `CalcModeSpread` function, unless the uncertainty is explicitly set to zero. The standard assignment of the 10% multiplicative standard deviation is based on an extensive survey of reported error values in biology, but it remains arbitrary and can and should be modified by the user as appropriate.

**Table 3.5:** Final input table required for **Step 3** for the hypothetical example in Tables 3.1–3.4

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Reported values</th>
<th>Standard Deviation</th>
<th>Total Weight</th>
<th>Multiplicative/ Additive</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_M$ (μM)</td>
<td>0.5</td>
<td>NaN</td>
<td>64</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>0.2</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>NaN</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>4.7434</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

B. Calculate the Mode and Spread of the probability distribution best describing our knowledge of the parameter values (Box 3.1) using the function `CalcModeSpread` (see Appendix A.1).

**Box 3.1: How the `CalcModeSpread` function works**

Using the summary table (Table 3.5) of the information gathered in Steps 1 and 2, the `CalcModeSpread` function calculates the Mode and Spread based on the weighted median and weighted standard deviation of the log-transformed experimental parameter values. The output of `CalcModeSpread` is used to create the final log-normal distribution in the next step, but its main use is to allow the model constructor an intuitive plausibility check of the translation process: the Mode represents the value most likely to be sampled, and the Spread by default represents a “multiplicative standard deviation”, the log-normal equivalent of the (additive) standard deviation of a normal distribution, i.e. it is calculated as a multiplicative error so that 68.27% of the sampled
values will lie within the interval [Mode/Spread, Mode*Spread].

The \texttt{CalcModeSpread} function starts by calculating the weighted median and the weighted standard deviation of the log-transformed literature values, taking into account their uncertainty (e.g., standard deviation or multiplicative error) and reliability weighting specified in \textbf{Step 2}. In order to take into account the location dependency in the Gaussian representation of log-normal data, during the log-transformation the values of $\mu$ and $\sigma$ are estimated through the following formulas:

\begin{align*}
\sigma^2 &= \log(1 + \frac{\sigma_Y^2}{\bar{Y}^2}) \quad (3.1) \\
\mu &= \log(\bar{Y}) - \frac{1}{2}\sigma^2 \quad (3.2)
\end{align*}

where $\bar{Y}$ and $\sigma_Y^2$ are the mean and variance of the parameter value in the original scale and $\mu$ and $\sigma$ are the estimated values calculated in the log scale.

The calculations are based on the weighted median of the literature values rather than the (weighted) mean, as it is less likely to be biased by outliers and thus provide a more accurate description of the central tendency in the dataset. The algorithm implemented in \texttt{CalcModeSpread} defines the weighted median as the unique value $\bar{Y}$ with the property that exactly half of the probability weight is assigned to values in the interval $[1, \bar{Y}]$. This approach, which expands earlier published algorithms for the calculation of weighted medians, circumvents the restriction that the weighted median can only be one of the values of the dataset (or the arithmetic mean of two adjacent values). The effects of different parameter weights and uncertainty on the weighted median are illustrated in Figure 3.3.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.3.png}
\caption{Calculation of the weighted median of two values (modes) $10^3$ and $10^6$ (log-transformed to 6.9078 and 13.82 respectively) with varying uncertainty and weights. The modes of the distributions are indicated by the blue lines and the weighted medians (WMedian) by the green lines. The weighted median moves away from data that are particularly uncertain and moves towards parameters with higher perceived relevance. This subtle adjustment would not be possible with previous implementations of the weighted median.}
\end{figure}

Once the weighted median and the weighted standard deviation have been calculated, they are exponentiated to yield the Mode and Spread (multiplicative standard deviation) of the final log-normal distribution.
3.3.4 Step 4: Calculate the location and scale parameters ($\mu$ and $\sigma$) of the log-normal distribution

The next part of the protocol is the calculation of the location ($\mu$) and scale ($\sigma$) parameters of the log-normal distribution describing the parameter uncertainty (Box 3.2). The function `CreateLogNormDist` (see Appendix A.2) performs this surprisingly non-trivial transformation of the intuitive values of the Mode and Spread into the mathematically relevant properties of the log-normal distribution.

**Box 3.2: How the CreateLogNormDist function works**

The `CreateLogNormDist` function is based on the multiplicative symmetric property of the log-normal distribution. Log-normal distributions are symmetrical in the sense that a value that is $x$ times larger than the most likely estimate (Mode) are just as plausible a value that is $x$ times smaller. More specifically, the Mode of the distribution is the value $x_0$ for which the condition $pdf(x_0 \delta) = pdf(x_0 / \delta)$ is fulfilled for all positive real numbers $\delta$, (where pdf is the probability density function). By exploiting this property, the `CreateLogNormDist` function calculates an interval of values from the Mode and Spread provided. By default, 68.27% (or, more precisely $erf(\frac{1}{\sqrt{2}})$, where $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$) of the sampled values should be in the confidence interval $[x_{\text{min}} = \frac{\text{Mode}}{\text{Spread}}, \ x_{\text{max}} = \text{Mode} \times \text{Spread}]$.

The default definition of the Spread can also be modified by the user, e.g. by requesting that 99% of values should be contained within the confidence interval this flexibility can help in eliciting parameter estimates from experimentalist, especially in situations of high uncertainty. Based on the $x_{\text{min}}$ and $x_{\text{max}}$ values, a two-by-two system of the equations containing the cumulative distribution function (CDF) of the log-normal distribution and the Mode can be solved, in order to derive the location parameter $\mu$ and the scale parameter $\sigma$ of the corresponding log-normal distribution:

$$\begin{cases} CDF_{\lognormal}(x_{\text{max}}) - CDF_{\lognormal}(x_{\text{min}}) = erf(\frac{1}{\sqrt{2}}) \\ Mode = e^{\mu - \sigma^2} \end{cases}$$

(3.3)

where $CDF = \frac{1}{2} + \frac{1}{2} erf[\frac{ln(x) - \mu}{\sqrt{2} \sigma}]$ and $x_{\text{min}}$ and $x_{\text{max}}$ are the lower and upper bounds of the confidence interval as defined before. By substituting these into equation 3.3 the final form of the system is obtained:

$$\begin{cases} \frac{1}{2} erf[\frac{ln(\text{Mode} \times \text{Spread}) - \mu}{\sqrt{2} \sigma}] - \frac{1}{2} erf[\frac{ln(\text{Mode} \times \text{Spread}) - \mu}{\sqrt{2} \sigma}] = erf(\frac{1}{\sqrt{2}}) \\ Mode = e^{\mu - \sigma^2} \end{cases}$$

(3.4)

By algebraically solving this system of equations, the $\mu$ and $\sigma$ parameters are obtained, and it is then easy to calculate any property in the distribution (e.g., the geometric mean and variance).
As in many cases log-normal distributions have long tails which may lead to sampling from a larger range of values than intended, the distributions can be truncated at a desired value. This is again best done in a multiplicative way. For instance, the modeler can decide how many times higher than the maximum value or lower than the minimum value among the experimental data points the sampling range should spread.

3.3.5 **Step 5: Account for thermodynamic consistency**

This step concerns parameters that cannot be independently sampled, because they are subject to thermodynamic constraints. In order to address the issue of thermodynamic consistency, joint probability distributions (multivariate distributions) can be employed for the interconnected parameters.

A. **Reversible Mass Action Reaction:** Input the $\mu$ and the $\sigma$ of all three initial (log-normal) distributions along with the number of parameter samples that should be generated in the function `Multivariate3param` (see Appendix A.3). The function then generates the requested number of parameter triplets, by using the strategy detailed in Box 3.3.

**Box 3.3: Thermodynamic consistency in a reversible mass action reaction**

In a simple reversible mass action reaction, three interconnected parameters are involved (the association rate ($k_{on}$), the dissociation rate ($k_{off}$), and the equilibrium constant ($K_D$)) (Figure 3.7). The three values are connected by the equation $k_{off} = k_{on} \cdot K_D$, and the plausible combinations of rate constants will depend on our uncertainty about the equilibrium constant (i.e., if we have very precise information about the position of the equilibrium, only a very limited range of combinations will be consistent with this thermodynamic information, while in the absence of thermodynamic information, any combination of rate constants is allowed). Of course, in principle, each of the three parameters has equal status, and the multivariate distribution for sampling can be constructed in three different ways, each expressing the dependency between two sampled parameters in terms of the third.

For instance, the two marginal distributions can be $k_{on}$ and $k_{off} (=k_{on} \cdot K_D)$, where $k_{off}$ is dependent on the values of $k_{on}$ and $K_D$. An overview of the steps followed is outlined in Figure 3.4. In order to decide which parameter will be the dependent one, the geometric coefficient of variation (GCV)\[223] is calculated for each parameter ($GCV = e^\sigma - 1$), and the parameter with the largest geometric coefficient of variation is set as the dependent variable (this is the parameter with the largest uncertainty, and sampling from the multivariate distribution will result in an increase...
of variance for the dependent parameter). The calculation of the dependent parameter is enabled by another property of the log-normal distributions: any product of two independent log-normal random variables is also log-normally distributed.\[215\]

Therefore, for the two log-normal distributions $lnk_{on} \sim N(\mu_{lnk_{on}}, \sigma_{lnk_{on}}^2)$ and $lnK_D \sim N(\mu_{lnK_D}, \sigma_{lnK_D}^2)$, their product $k_{off}$ will also be a log-normal distribution $lnk_{off} \sim N(\mu_{lnk_{on}} + \mu_{lnK_D}, \sigma_{lnk_{on}}^2 + \sigma_{lnK_D}^2)$ and its location and scale parameters will be $\mu_{lnk_{off}} = \mu_{lnk_{on}} + \mu_{lnK_D}$, $\sigma_{lnk_{off}}^2 = \sigma_{lnk_{on}}^2 + \sigma_{lnK_D}^2$.

A similar argument applies for the quotient of two log-normal distributions, although in this case the parameter $\mu$ will be derived by the formula $\mu_{quotient} = \mu_{divident} - \mu_{divisor}$. The formula for the calculation of the parameter $\sigma$ does not change. Note that while log-normal distributions are sometimes counterintuitive to use, this useful and unique property alone would be sufficient justification for using them as often as possible in the definition of biologically informative priors, in preference over uniform or standard normal distributions.

Figure 3.4: Overview of the process followed by the Multivariate3param function

It is then very easy to transform the two marginal distributions $k_{on}$ and $k_{off}$ to normal ones, via the natural logarithm. The problem can therefore be reduced to the case of a multivariate normal distribution with the following probability density function: \[104, 224, 225\]

$$f(x, y) = \frac{1}{2\pi \sigma_X \sigma_Y \sqrt{1-\rho^2}} \exp \left( -\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - 2\rho(x-\mu_X)(y-\mu_Y) \right] \right)$$

where $\rho$ is the correlation between $X(=k_{on})$ and $Y(=k_{off})$ and $\sigma_X > 0$ and $\sigma_Y > 0$.

In this case, $\mu = \left( \frac{\mu_X}{\mu_Y} \right)$ and $\Sigma = \begin{pmatrix} \sigma_X^2 & \rho \sigma_X \sigma_Y \\ \rho \sigma_X \sigma_Y & \sigma_Y^2 \end{pmatrix}$ (covariance matrix).

In this way, thermodynamically consistent triplets of $k_{on}$, $k_{off}$ and $K_D$ parameters can be sampled, which can then be used for the model simulations.
Enzymatic Reactions: For the five-parameter Michaelis–Menten kinetics, input the $\mu$ and the $\sigma$ of all five initial distributions as derived in Steps 1–4 of the protocol, along with the number of parameter samples that should be generated the Multivariate5param function (see Appendix A.4). The function then generates the requested number of parameter sets by using the principles described above, by using the strategy described in Box 3.4. This strategy can be expanded to even more complex cases which would require more than five parameters (e.g., Bi–Bi, Uni–Bi, Ter–Ter mechanisms etc.), by using the internal mass action kinetic reactions to define the correlations of the external Michaelis–Menten parameters.

**Box 3.4: Thermodynamic consistency in a 5-parameter Michaelis–Menten enzymatic reaction**

The same strategy applied for the three parameter mass action kinetics (Box 3.3), can be expanded to slightly more complex reaction schemes, such as reversible enzymatic reactions. The simplest case in this category would be a reaction where a single substrate is transformed to a single product (Uni–Uni reaction) as per:

$$
E + A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} X \overset{k_3}{\underset{k_4}{\rightleftharpoons}} P + E
$$

This reaction is usually described by Michaelis–Menten kinetics, where the parameters $K_M^+, K_M^-, V_{max}^+$ and $V_{max}^-$ are linked with the $K_{eq}$ through the Haldane equation:

$$
K_{eq} = \frac{V_{max}^- K_M^-}{V_{max}^+ K_M^+}.
$$

In this case, the dependencies and correlations of the parameters are defined in two stages (Figure 3.5). In Stage 1, the parameter with the largest geometric coefficient of variation is set as the dependent one (as in the 3-parameter case) and is recalculated from the independent parameters through the Haldane equation. In Stage 2, the internal mass action kinetics parameters $k_1, k_2, k_3, k_4$, derived by using the King–Altman method, are used to define the correlations between the observed (external) parameters $K_M, V_{max}$ and $K_{eq}$. The internal parameters can be derived from the external ones manually or, for more complex reactions, algorithms and online tools that automate the process can be employed.

For example, the transformation equations for the Uni–Uni reaction are the following:
The marginal distributions in this case are for the parameters $K_{+M}$, $K_{-M}$, $V_{+max}$, $V_{-max}$ and $K_{eq}$, and the mass action kinetics parameters $k_1, k_2, k_3, k_4, K_{D1}$ and $K_{D2}$ are used to calculate the covariance matrix and thus define the correlations within the system. The mathematical properties are similar to the ones employed in the three parameter case, being based on the properties of products and quotients of lognormal distributions and additionally using the Fenton–Wilkinson approximation for the addition of two independent log-normal distributions:\cite{231,232}

$$
\sigma^2 = \ln \left[ \sum e^{2\mu_j + \sigma_j^2} (e^{\sigma_j^2/2} - 1) \right] + 1
$$

$$
\mu_Z = \ln \left[ \sum e^{\mu_j + \sigma_j^2/2} \right] - \frac{\sigma_Z^2}{2}
$$

where $\mu_j$ and $\sigma_j$ ($j = 1...n$) are the parameters of the initial log-normal distributions and $\mu_Z$ and $\sigma_Z^2$ are the parameters of the log-normal distribution representing their sum.
C. **Futile Cycles:** Finally, it is possible that substrate (futile) cycles\(^{233-235}\) may be a part of a biological model. For each closed loop in the metabolic pathways, the total Gibbs free energy change (\(\Delta G^o\)) has to be zero as an additional thermodynamic constraint. This means that the equilibrium constants of the reactions involved in the cycle will need to additionally comply with the constraint  
\[ K_{eq1} \cdot K_{eq2} \cdot \ldots \cdot K_{eqN} = 1. \]  
\(^{236}\) In order to address this aspect of the system, e.g., for a three reaction loop, correlate the parameters for two of the reactions independently as described above, and calculate the \(K_{eq}\) of the third reaction from the
Gibbs energy equation \( K_{eq3} = \frac{1}{K_{eq1}K_{eq2}} \). Finally, pass the values calculated for \( K_{eq3} \) on the internal system of equations (Michaelis–Menten, mass action etc.) of the third reaction and thus create the additional correlations.

### 3.4 Troubleshooting

**Table 3.6: Troubleshooting table**

<table>
<thead>
<tr>
<th>Step</th>
<th>Problem</th>
<th>Possible reason</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>General advice</td>
<td>MATLAB generates error message “Undefined function or variable ‘Function name’”.</td>
<td>One or more of the functions are not located in the MATLAB directory</td>
<td>Ensure that all the relevant functions are kept in the same directory and that it is the current folder in MATLAB.</td>
</tr>
<tr>
<td>3</td>
<td>MATLAB cannot calculate the weighted standard deviation</td>
<td>It is possible that one of the uncertainty values in the input table was 0 and the uncertainty type was defined as “multiplicative”</td>
<td>Ensure that the corresponding “uncertainty type” for all the uncertainty values that are zero is defined as “additive”.</td>
</tr>
<tr>
<td>3</td>
<td>MATLAB generates error message “The weights cannot have values smaller than 0.0001.”</td>
<td>The user input includes very small weights that can cause numerical errors during the calculation of the weighted median.</td>
<td>Ensure that all the weights employed are larger than 0.0001. If a total weight is divided into partial weights between replicates of the same experiment, ensure that partial weights also comply with this boundary.</td>
</tr>
</tbody>
</table>
3.5 Anticipated results

The practical implementation and anticipated results of the protocol are illustrated by a case study on a real-world modelling example (a tiny part of a much larger molecular model), which describes a reaction within the gene regulatory system of the Gram-positive bacterium *Streptomyces coelicolor*. This reaction involves a member of the TetR family of repressors, the ScbR homo-dimer ($R_2$), which binds to its own gene operator ($O_R$) and represses its mRNA transcription.

$$O_R + R_2 \rightleftharpoons O_R \cdot R_2$$

To represent this reaction three parameters are interconnected in the following rate law:

$$v = \frac{k_{\text{on}}}{K_{D1}} \cdot [O_R] \cdot [R_2] - k_1^- \cdot [O_R \cdot R_2]$$

The parameters of this reaction are the dissociation constant for binding of ScbR to $O_R$ ($K_{D1}$) and the dissociation rate of ScbR to $O_R$ ($k_1^-$). These two parameters are connected through the association rate constant ($k_{\text{on}}$) by the equation $k_{\text{on}} = \frac{k_1^-}{K_{D1}}$. The model will be used to simulate an *in vivo* experiment under physiological conditions where the pH is approximately $\sim 7$ and the temperature is $30^\circ C$.

**Step 1: Find and document parameter values**

A. A range of values were retrieved from literature from phylogenetically related organisms and the same protein-receptor family (Table 3.7). The values concerned the TetRtetO interaction and other tetR-like proteins binding to their corresponding operators. The literature data are summarised in Table 3.8.
Table 3.7: Summary of parameter values retrieved from literature for case study

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Remarks and References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.005</td>
<td></td>
<td>An early publication by Kleinschmidt et al.\cite{237} based on stopped-flow measurements at various salt concentrations reports a $K_A$ of $2 \cdot 10^{11} M^{-1}$, therefore a $K_D (\frac{1}{K_A})$ of 0.005 nM. The Tet repressor and TetO operator were derived by using an overproducing E. coli strain.</td>
</tr>
<tr>
<td></td>
<td>0.179</td>
<td></td>
<td>Data derived from equilibrium SPR analysis of TetR–tetO interaction report a $K_A$ of $5.6 \pm 2 nM^{-1}$,\cite{238} therefore a $K_D (\frac{1}{K_A})$ of $0.179 \pm 0.064$ nM. E. coli was used as a host for the expression of proteins and the experiments were conducted with synthetic tetO-containing fragments.</td>
</tr>
<tr>
<td>$K_{D1}$</td>
<td>4.4</td>
<td>nM</td>
<td>Additionally, a study on the TetR-like protein Rv3066 binding to the mmr operon in <em>M. tuberculosis</em> suggests a $K_D$ of $4.4 \pm 0.3$ nM.\cite{239}</td>
</tr>
<tr>
<td></td>
<td>0.1, 0.3 and 5.8</td>
<td></td>
<td><em>In vitro</em> studies on TetR-like protein ActR in <em>S. coelicolor</em> suggest a $K_D$ in the range of 0.1–5.8 nM.\cite{240}</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td></td>
<td>According to the study by Kleinschmidt et al.\cite{237}, the maximal association rate constant was $k_{on} = 3 \cdot 10^8 M^{-1} s^{-1}$. By taking into account the equilibrium association constant reported above ($K_A = 2 \cdot 10^{11} M^{-1}$), the dissociation rate constant can be calculated as per $k_1^- = \frac{k_{on}}{K_A} = \frac{3 \cdot 10^8}{2 \cdot 10^{11}} = 0.0015 s^{-1} = 0.09 min^{-1}$.</td>
</tr>
<tr>
<td>$k_1^-$</td>
<td>0.846 and 0.44</td>
<td>$min^{-1}$</td>
<td>A study on a TetR-like protein (RolR) which binds to its operator roLO and blocks the transcription of rolHMD (Gram positive and GC content ~ 50-60% bacteria) reports the dissociation rates $1.41 \cdot 10^{-2} s^{-1}$ (0.846 min$^{-1}$), and $7.34 \cdot 10^{-3} s^{-1}$ (0.44 min$^{-1}$)\cite{241}, measured by surface plasmon resonance (SPR).</td>
</tr>
<tr>
<td>$k_{on1}$</td>
<td>18</td>
<td>$nM^{-1} \cdot min^{-1}$</td>
<td>As reported in the study by Kleinschmidt et al.\cite{237} mentioned above, the maximal association rate constant was $k_{on} = 3 \cdot 10^8 M^{-1} s^{-1} = 18 nM^{-1} min^{-1}$.</td>
</tr>
</tbody>
</table>
B. The standard errors reported in the literature are converted to standard deviations by multiplying by the square root of the sample size, as per the equation: $SD = SEM \cdot \sqrt{n}$. In the cases where the sample size was not reported, a value of $n=3$ was assumed. The uncertainty type is set as “additive” for all parameter values.

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Reported values</th>
<th>Standard Deviation</th>
<th>Multiplicative/ Additive</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{D1}$ (nM)</td>
<td>0.005</td>
<td>N/A</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>0.179</td>
<td>0.064</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>0.3</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>N/A</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>N/A</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>5.8</td>
<td>N/A</td>
<td>Additive</td>
</tr>
<tr>
<td>$k_1^-$ (min$^{-1}$)</td>
<td>0.09</td>
<td>N/A</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>0.846</td>
<td>N/A</td>
<td>Additive</td>
</tr>
<tr>
<td>$k_{on1} (nM^{-1}min^{-1})$</td>
<td>18</td>
<td>N/A</td>
<td>Additive</td>
</tr>
</tbody>
</table>

**Step 2: Judge plausibility of values and assign weights**

Upon the detailed documentation of the parameter sources, values and uncertainties, the next step was to assign the appropriate weight to each parameter value. It is immediately obvious that not all of the experimental values are equally informative: the values reported in literature correspond to different protein–operator binding reactions, although all are part of the same family of repressors (TetR). Moreover, the organisms used in the experiments include *E. coli*, *M. tuberculosis* and *Corynebacterium glutamicum*, while the model target organism *S. coelicolor* was used only in the experiments by Ahn et al. Finally, the values were acquired by *in vitro* testing, although in some of the publications a strong correlation between *in vitro* and *in vivo* measurements is noted. Based on these considerations, plausibility weights were assigned to each parameter value (on an arbitrary, standardized scale), as shown in Table 3.9.
Chapter 3. Defining informative priors for ensemble modelling in systems biology

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Parameter value</th>
<th>SD</th>
<th>SD</th>
<th>Partial plausibility weights</th>
<th>pH &amp; temperature</th>
<th>Total weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{D1}$ (nM)</td>
<td>0.005</td>
<td>N/A</td>
<td></td>
<td>$In vitro$ (1) Unrelated (1) Related (2)</td>
<td>pH 8.0, 20°C (2)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>0.179</td>
<td>0.064</td>
<td></td>
<td>$In vitro$ (1) Unrelated (1) Related (2)</td>
<td>pH 8.0, 22°C (2)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>0.3</td>
<td></td>
<td>$In vitro$ (1) Related (2) Related (2)</td>
<td>pH 8.0, 22°C (2)</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>N/A</td>
<td></td>
<td>$In vitro$ (1) Same (4) Related (2)</td>
<td>pH 7.8, 30°C (4)</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>N/A</td>
<td></td>
<td>$In vitro$ (1) Same (4) Related (2)</td>
<td>pH 7.8, 30°C (4)</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>5.8</td>
<td>N/A</td>
<td></td>
<td>$In vitro$ (1) Same (4) Related (2)</td>
<td>pH 7.8, 30°C (4)</td>
<td>32</td>
</tr>
<tr>
<td>$k_1^-$ (min$^{-1}$)</td>
<td>0.09</td>
<td>N/A</td>
<td></td>
<td>$In vitro$ (1) Unrelated (1) Related (2)</td>
<td>pH 8.0, 20°C (2)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>0.846</td>
<td>N/A</td>
<td></td>
<td>$In vitro$ (1) Related (2) Related (2)</td>
<td>pH 7.5, 30°C (4)</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>0.44</td>
<td>N/A</td>
<td></td>
<td>$In vitro$ (1) Related (2) Related (2)</td>
<td>pH 7.5, 30°C (4)</td>
<td>16</td>
</tr>
<tr>
<td>$k_{on1}$ (nM$^{-1}$ min$^{-1}$)</td>
<td>18</td>
<td>N/A</td>
<td></td>
<td>$In vitro$ (1) Unrelated (1) Related (2)</td>
<td>pH 8.0, 20°C (2)</td>
<td>4</td>
</tr>
</tbody>
</table>

The information for each parameter is summarised in a table (Tables 3.10 and 3.11) which will be used as input for the functions that will generate the probability distributions.
TABLE 3.10: Input matrix of $K_{D1}$ values found in the literature

<table>
<thead>
<tr>
<th>Parameter value (nM)</th>
<th>Uncertainty*</th>
<th>Weight</th>
<th>Uncertainty type**</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>NaN</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>0.179</td>
<td>0.11085</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>4.4</td>
<td>0.5196</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>NaN</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>0.3</td>
<td>NaN</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>5.8</td>
<td>NaN</td>
<td>32</td>
<td>0</td>
</tr>
</tbody>
</table>

TABLE 3.11: Input matrix of $k_1^-$ values found in the literature

<table>
<thead>
<tr>
<th>Parameter value (nM)</th>
<th>Uncertainty*</th>
<th>Weight</th>
<th>Uncertainty type**</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.09</td>
<td>NaN</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>0.846</td>
<td>NaN</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>0.44</td>
<td>NaN</td>
<td>16</td>
<td>0</td>
</tr>
</tbody>
</table>

*Uncertainty values in this table correspond to the standard deviation calculated from SEM reported in literature.

** For the computational script provided, additive standard deviations are coded as “0”, multiplicative standard deviations as “1”.

NaN, “not a number”, indicates cases where no error was reported in the literature; the script will automatically replace these by a multiplicative error of 10% by default.

Step 3 and 4: Calculate the Mode and Spread and location and scale parameters ($\mu$ and $\sigma$) of the log-normal distribution

Using the data in Tables 3.10 and 3.11 as input for the CalcModeSpread and CreateL-ogNormDist functions (Script 3.1) gives the following results (summarised in Table 3.12):

SCRIPT 3.1: Script used to generate the distributions for $K_{D1}$ and $k_1^-$

```matlab
%Distribution for KD1
%Input summary table for CalcModeSpread
V=[0.005 NaN 4 0;0.179 0.11085 4 0; 4.4 0.5196 8 0; 0.1 NaN ...
  32 0; 0.3 NaN 32 0; 5.8 NaN 32 0];
```
%Calculate the mode and Spread
>> [Mode, Spread] = CalcModeSpread(V);

%Calculate the mu, sigma and the minimum and maximum values ... within the confidence interval
>> [mu, sigma, Xmin, Xmax] = CreateLogNormDist(Mode, Spread);

%Generate summary output table
>> KD1= table(Mode, Spread, mu, sigma, Xmin, Xmax);

KD1 =
  Mode  Spread  mu    sigma  Xmin   Xmax
  _____  _____  _____  _____  _____  _____
  0.29964 6.8357  0.15474 1.1662  0.043834 2.0482

%Distribution for k1
%Input summary table for CalcModeSpread
>> V=[0.09 NaN 4 0;0.846 NaN 16 0;0.44 NaN 16 0];

%Calculate the mode and Spread
>> [Mode, Spread] = CalcModeSpread(V);

%Calculate the mu, sigma and the minimum and maximum values ... within the confidence interval
>> [mu, sigma, Xmin, Xmax] = CreateLogNormDist(Mode, Spread);

%Generate summary output table
>> k1= table(Mode, Spread, mu, sigma, Xmin, Xmax);

k1 =
  Mode  Spread  mu    sigma  Xmin   Xmax
  _____  _____  _____  _____  _____  _____
  0.48898 1.9789 -0.37641 0.58226  0.2471  0.96762
TABLE 3.12: Summary of log-normal distribution properties for parameters $K_{D1}$, $k_{on1}$ and $k_{1}^{-}$

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$X_{min}$</th>
<th>$X_{max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{D1}$</td>
<td>0.299</td>
<td>6.84</td>
<td>0.1547</td>
<td>1.1662</td>
<td>0.044</td>
<td>2.05</td>
</tr>
<tr>
<td>$k_{1}^{-}$</td>
<td>0.489</td>
<td>1.98</td>
<td>-0.3764</td>
<td>0.5823</td>
<td>0.247</td>
<td>0.968</td>
</tr>
<tr>
<td>$k_{on1}$</td>
<td>-</td>
<td>-</td>
<td>-0.5312</td>
<td>1.3035</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 3.6 shows how the estimated overall distributions correspond to the literature data, their weights and experimental errors.

Step 5: Account for thermodynamic consistency

Since the two parameters are interconnected, it is necessary to take into account the thermodynamic consistency of the system. Values for the third parameter of the system ($k_{on1}$) also need to be retrieved from literature in order to impose the necessary constraints on the other two parameters. For $k_{on1}$, only one reported value was found, so rather than using the standard approach of collecting generic values from the literature, we made use of the fact that the three parameters are interconnected as per the formula: $k_{on1} = \frac{k_{1}^{-}}{K_{D1}}$. The location and scale parameters for this distribution can be
calculated from the $\mu$ and $\sigma$ of $K_{D_1}$ and $k_1^-$ based on the properties of the product-
s/quotients of log-normal distributions by using the formulas: $\mu_{k_{on1}} = \mu_{k_1^-} - \mu_{K_{D_1}}$ and
$\sigma^2_{k_{on1}} = \sigma^2_{k_1^-} + \sigma^2_{K_{D_1}}$.

The next step is to account for the thermodynamic consistency of the reaction. This
is achieved by creating a bivariate system as described in Step 5. Since information
for $k_{on1}$ in the literature was scarce, it was selected as the dependent parameter as per:

$k_{on1} = \frac{k_1^-}{K_{D_1}}$. In this way, a system of distributions can be created where each distribution
is constrained by the other two (Figure 3.7). The information for the three parameters
was then provided as input for the Multivariate3param function (Script 3.2):

```matlab
SCRIPT 3.2: Parameter information for the Multivariate3param function

1  %Input the parameters \mu and \sigma for the three distributions ... and the number of values that we want to sample
2  >> muKD=0.1547; sigmaKD=1.1662; mukon=-0.5312; sigmakon=1.3035; ...
3  mukoff=-0.3764; sigmakoff=0.5823; ParamNo=10;
4  %Generate parameter values from the three distributions
5  >> [KD1, kon1, koff1] = Multivariate3param(muKD, sigmaKD, mukon, ...
6    sigmakon, mukoff, sigmakoff, ParamNo)
7  %Output
8  KD1= kon1= koff1=
9  3.778  0.134  0.508
10  0.415  2.287  0.949
11  0.996  1.161  1.157
12  0.194  3.259  0.631
13  0.829  0.983  0.815
14  2.343  0.172  0.403
15  0.579  0.851  0.493
16  18.440  0.071  1.317
17  0.336  2.580  0.866
18  1.398  0.287  0.401
19  % Plot 3D scatter of generated values
20  >> figure(1)
```
Chapter 3. Defining informative priors for ensemble modeling in systems biology

>> scatter3(log(KD1),log(kon1),log(koff1),'.')

Figure 3.7: The bivariate distribution for parameters $k_{on1}$ and $k_1$ the two marginal distributions. Unless appropriate constraints are imposed, there is a risk of sampling thermodynamically infeasible parameter combinations. The scattered points demonstrate the sampling from the two distributions without considering their dependence and the ellipsoid contour lines represents the probability distribution of acceptable parameter combinations when taking into account thermodynamic consistency, through the relationship $K_{D1} = \frac{k_1}{KD_1}$.

By using this bivariate distribution, values for parameters $k_1$ and $k_{on1}$ were generated simultaneously in order to account for their dependency. The value for $K_{D1}$ was calculated from the parameters sampled from the other two distributions in an additional step, using the relationship $K_{D1} = \frac{k_1}{KD_1}$. Parameters that are sampled using the function `Multivariate3param` are not only consistent with our knowledge about each individual parameter, but also fulfill the relevant thermodynamic constraints imposed by
the fact that they describe a single reversible reaction. As the model’s rate law requires the parameters $K_D$ and $k^{-1}$, and not $k_{on}$, the user can ignore the values for the latter parameter. An additional extensive case study where the protocol has been implemented in the design of priors for a complete model of *Trypanosoma brucei* glycolysis, one of the classical proof-of-concept models of systems biology research, can be found in Appendix B.

### 3.6 Conclusion

The protocol we present here provides a systematic, rigorous, fast and consistent method to translate existing biological knowledge and modellers’ beliefs about a biological system into log-normal probability distributions to be used as informative priors in biological ensemble modelling approaches, and to sample from these distributions in a thermodynamically consistent manner. This will hopefully lead to a much improved quality of models, and potentially form a bridge in collaborations between computational and experimental biologists.
Chapter 4

Unravelling the $\gamma$-butyrolactone network in *Streptomyces coelicolor* under uncertainty

4.1 Abstract

*Streptomyces* species are Gram-positive soil-dwelling bacteria, which are known as a prolific source of secondary metabolites, such as antibiotics. Antibiotic production is coordinated in the bacterial population through the use of diffusible signalling molecules of the $\gamma$-butyrolactone (GBL) family. The GBL regulatory system involves a small, yet complex two-gene network (*scbR* and *scbA*), the mechanism of which has not yet been completely defined, and which governs a potentially bi-stable switch between the “on” and “off” states of antibiotic production. The work presented in this chapter focused on the creation and analysis of a versatile and adaptable computational model of this system, in close collaboration with experimental microbiologists. The potential influences by transcriptional interference, by antisense RNA interactions between the mRNAs of the two genes and the formation of a putative complex between scbR and scbA proteins were investigated. The effects of different promoter strengths were also a point of interest. The analysis of the model leads to the rejection of the relevance of the putative complex, but also suggests that transcriptional interference alone is not sufficient to explain the system’s behaviour. Antisense RNA interactions seem to be the
driving force that leads to the observed behaviour, combined with an aggressive \textit{scbR} promoter. The model can now be used to question and refine our understanding of the system’s activity and could suggest a different biological role than the one originally anticipated.

\textbf{4.2 Introduction}

The core aim of synthetic biology is the design and engineering of complex biological systems with functionalities that do not exist in nature. In order to accomplish this, novel regulatory circuits need to be developed, which will enable the precise control of gene expression over a wide range of conditions.\textsuperscript{[111]} The bacterium \textit{Vibrio fischeri} has been in the spotlight in recent years, due to its communication mechanism known as “quorum sensing” (QS).\textsuperscript{[242]} This organism produces acyl homoserine lactone (AHL) as a signal molecule and exports it to the environment. As the colony grows and the cell density increases, the concentration of AHL in the environment also rises, until it reaches a threshold which activates the expression of specific genes responsible for the emission of light.

The system responsible for the production of AHL is composed of two genes and their encoding regulatory proteins LuxI and LuxR. LuxI is the actual synthase of the autoinducer and LuxR, is a repressor belonging to the TetR family.\textsuperscript{[115]} AHL forms a complex with the LuxR protein and together they bind to a short sequence called \textit{lux} box which both enhances the transcription of LuxI (thus leading to even further accumulation of AHL) and activates the bioluminescence gene cluster. This positive feedback loop leads to a behaviour similar to a bistable switch, with the system being either in the “on” or in the “off” state, and with a zone of unstable intermediate states between them.

Although the quorum sensing circuit has been widely employed in synthetic biology with numerous successful applications,\textsuperscript{[243–245]} it has some important limitations, such as potential crosstalk between different systems due to the promiscuity of the signalling molecules or the promoters,\textsuperscript{[246]} and the problematic implementation in eukaryotic organisms.\textsuperscript{[247]} There is therefore a need to expand the available potential circuits which
will enhance the synthetic biology toolbox and pave the way to even more exciting applications.

A good candidate for this purpose could be the $\gamma$-butyrolactone (GBL) signalling circuits of *Streptomyces coelicolor*. Streptomycetes are Gram-positive, filamentous, soil bacteria, that produce antibiotics to eliminate their competitors in the face of hostile environmental conditions. As these compounds can be toxic even to the producing strains, the biosynthesis needs to be carefully regulated. This is achieved via the SCB1 $\gamma$-butyrolactones, a group of signalling molecules associated with the regulation of antibiotic production and some aspects of the secondary metabolism. These molecules are postulated to have a similar effect as the AHLs in quorum sensing, inducing a switch-like transition to the antibiotic producing state, which is also dependent on the growth phase. The structure of the circuit also has similarities to the quorum sensing system, as it involves two genes and their respective proteins (ScbR and ScbA). ScbR belongs to the TetR family of repressors and inhibits its own transcription, as well as the transcription of the divergently encoded ScbA, which is the synthase of the signalling molecule SCB1. Furthermore, it represses cpkO, a regulatory gene for the CPK antibiotic biosynthesis gene cluster. SCB1 binds to ScbR, effectively deactivating it and thus leading to the further production of butyrolactones. Upon reaching a specific SCB1 threshold (reported to be 250 nM), the CPK cluster is also activated, leading to the production of antibiotics.

Apart from this general scheme, little further mechanistic detail is known, although different hypotheses have been put forward. The two genes are transcribed in opposite directions and their promoters overlap by 53 bp. In previous studies it has been reported that divergent overlapping promoters are responsible for regulating expression of genes. This topology has been suggested to be also important in the GBL circuit, for the precise switch of the system at relatively low concentrations. Another scenario that has been suggested is the formation of a putative complex between ScbA and ScbR proteins which acts in a similar manner as the LuxR-AHL complex in quorum sensing and further enhances the transcription of the *scbA* gene. Finally, studies have shown that RNA transcripts from genes with overlapping promoters may bind to each other and interfere with each other’s activity (antisense RNA interactions). This
has also been suggested to occur in the GBL system and thus induce a form of internal regulation.\textsuperscript{[123,256]}

Previous modelling work on this system is limited to two published models which investigated some of these scenarios. Mehra \textit{et al.}\textsuperscript{[1]} proposed a model based on the scenario of the ScbA–ScbR complex formation and Chatterjee \textit{et al.}\textsuperscript{[123]} focused on the effects of the overlapping promoters and antisense interaction. Both models were deterministic and were based on the traditional parameter fitting strategies, discussed in Chapter 2. This system therefore provides a good opportunity for application of the ensemble modelling strategies complemented with the protocol described in Chapter 3, which will examine all possible scenarios within a range of reasonable parameters and thus enable a consensus regarding the possible processes that take place within this circuit. In order to accomplish this, the previous models were recreated and analysed in order to gain a better understanding of how a simpler form of the system behaves and identify potential fragilities and omissions which will in turn reveal the possibilities of improvement. At the same time, a recent successful model on the quorum sensing system by Weber \textit{et al.}\textsuperscript{[2]} was also recreated, as the similarities that the two systems share will enable a comparison of their respective behaviours (Figure 4.1).
Chapter 4. Modelling the $\gamma$-butyrolactone network in Streptomyces coelicolor

4.3 Methods

4.3.1 Recreation of the published GBL and QS models

The first step of the project was to study the previously published models of the GBL system and attempt to replicate their results. Their careful investigation would lead to
the identification of their weak points and omissions and therefore to potential ways in which they could be improved. The main focus was the published model by Mehra et al.,[1] but the alternative model by Chatterjee et al.[123] was also considered. The reactions of the published model by Mehra et al. are summarised in Table 4.1. A schematic representation of the model is shown in Figure 4.2a.

**Table 4.1:** Chemical reactions for the published GBL model by Mehra et al.[1]

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_R + R \xrightarrow{k_{OR}} O_R - R$</td>
<td>(1)</td>
</tr>
<tr>
<td>$O_{A'} + AR \xrightarrow{k_{OAAR}} O_{A'} - AR$</td>
<td>(2)</td>
</tr>
<tr>
<td>$O_A + R \xrightarrow{k_{OA}} O_A - R$</td>
<td>(3)</td>
</tr>
<tr>
<td>$A + R \xrightarrow{k_{AR}} AR$</td>
<td>(4)</td>
</tr>
<tr>
<td>$C_i + R \xrightarrow{k_{CR}} CR$</td>
<td>(5)</td>
</tr>
<tr>
<td>$S + A \xrightarrow{k_{prot}} C_i + A$</td>
<td>(6)</td>
</tr>
<tr>
<td>$\frac{[O_R]}{[O_{R,T}]} = \frac{K_{OR}}{K_{OR} + [R]}$</td>
<td>(7)</td>
</tr>
<tr>
<td>$\frac{[O_A]}{[O_{A,T}]} \cdot \frac{[O_{A'}]}{[O_{A',T}]} = \frac{KOA + [AR]}{(KOA + [R])(KOA' + [AR])}$</td>
<td>(8)</td>
</tr>
<tr>
<td>$\frac{[O_A]}{[O_{A,T}]} \cdot \frac{[O_{A'}]}{[O_{A',T}]} = \frac{KOA + KOA'}{(KOA + [R])(KOA' + [AR])}$</td>
<td>(9)</td>
</tr>
</tbody>
</table>

The reported results of the previous model were based on testing a defined range of parameters and evaluating a range where bistable behaviour was obtained. An optimum combination of parameters for bistability was also defined. Therefore, the initial analysis performed involved a deterministic simulation on the previous model by employing the parameter optimum values for bistability. In order to explore alternative software capabilities and ensure that the results would not be affected by differences in software performance, the simulations were performed in both MATLAB® by employing the stiff ordinary differential equation solver ode15s, as well as COPASI[29] software by employing the deterministic LSODA algorithm. All equations, parameter values, initial conditions (where reported) and assumptions were in accordance with the publication by Mehra et al. When an initial condition was not specified, an assumption or approximation was made. Moreover, as the publication did not specify a parameter for the basal
transcription rate of *scbA* mRNA, the value reported in the alternative model by Chatterjee *et al.*\[1\] was used. This choice was made because the two models reported similar or the same values for the majority of their parameters. The results were compared with each other and with the original paper’s figures.

Additionally, upon successful completion of the initial simulations and in order to study the model’s behaviour under perturbation in the parameter values, an uncertainty modelling analysis was performed. The first group of simulations included distributions, chosen so that the majority of the reported range of tested values would be sampled. Subsequently, a second group of simulations was performed by using more narrow distributions which only sampled values from within the reported range of bistability. For each group of simulations, 1000 parameter sets were generated from the distributions and then used as input to create a collection of model outputs. The total 2000 simulations were conducted by using the MATLAB ode15s solver to decrease computational time.

At the same time, a recently published model on the quorum sensing (QS) system by Weber *et al.*[2] was also recreated in order to study the interesting framework of analysis that was employed. This system of bacterial communication shares important traits with the GBL system, and the simulation methods which were employed to analyse it could be beneficial for the analysis of the improved GBL model. The analysis of the QS system according to the publication involved the simulation of a deterministic model of coupled ODEs and of a stochastic model which employed a modified Gillespie algorithm to account for cellular growth and division.

**Table 4.2:** Chemical reactions for the published QS model by Weber *et al.*[2]

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( DNA \xrightarrow{\alpha_{ki}} DNA + mRNA_{luxR} )</td>
<td>(1)</td>
</tr>
<tr>
<td>( DNA \xrightarrow{\alpha_{ki}} DNA + mRNA_{luxI::gfp} )</td>
<td>(2)</td>
</tr>
<tr>
<td>( DNA \cdot (luxR \cdot A)_2 \xrightarrow{k_R} DNA \cdot (luxR \cdot A)<em>2 + mRNA</em>{luxR} )</td>
<td>(3)</td>
</tr>
<tr>
<td>( DNA \cdot (luxR \cdot A)_2 \xrightarrow{k_i} DNA \cdot (luxR \cdot A)<em>2 + mRNA</em>{luxI::gfp} )</td>
<td>(4)</td>
</tr>
<tr>
<td>( mRNA_{luxR} \xrightarrow{pR} mRNA_{luxR} + luxR )</td>
<td>(5)</td>
</tr>
<tr>
<td>( mRNA_{luxI::gfp} \xrightarrow{pI} mRNA_{luxI::gfp} + luxI :: gfp )</td>
<td>(6)</td>
</tr>
<tr>
<td>( luxI :: gfp \xrightarrow{kA} A + luxI :: gfp )</td>
<td>(7)</td>
</tr>
</tbody>
</table>
Chapter 4. Modelling the \(\gamma\)-butyrolactone network in Streptomyces coelicolor

\[
luxR + A \quad \xleftarrow{k_1^{-1} / K_{d1}} \quad luxR \cdot A \quad \text{(8)}
\]

\[
2(luxR \cdot A) \quad \xleftarrow{k_2^{-1} / K_{d2}} \quad (luxR \cdot A)_2 \quad \text{(9)}
\]

\[
(luxR \cdot A)_2 + DNA \quad \xleftarrow{k_{lux}^{-1} / K_{dlux}} \quad DNA \cdot (luxR \cdot A)_2 \quad \text{(10)}
\]

\[
A \xrightarrow{D} A_{ext} \quad \text{(11)}
\]

\[
A \xrightarrow{dA} \emptyset \quad \text{(12)}
\]

\[
A_{ext} \xrightarrow{dA} \emptyset \quad \text{(13)}
\]

\[
mRNA_{luxR} \xrightarrow{dm_{R}} \emptyset \quad \text{(14)}
\]

\[
mRNA_{luxI::gfp} \xrightarrow{dm_{I}} \emptyset \quad \text{(15)}
\]

\[
luxR \xrightarrow{dR} \emptyset \quad \text{(16)}
\]

\[
luxI :: gfp \xrightarrow{dI} \emptyset \quad \text{(17)}
\]

\[
(luxR \cdot A)_2 \xrightarrow{dc_2} \emptyset \quad \text{(18)}
\]

\[
luxR \cdot A \xrightarrow{dc_1} \emptyset \quad \text{(19)}
\]

\[
DNA \quad \xrightarrow{ln(2)/\tau} \quad DNA + DNA \quad \text{(20)}
\]

\[
DNA \cdot (luxR \cdot A)_2 \quad \xrightarrow{ln(2)/\tau} \quad DNA \cdot (luxR \cdot A)_2 + DNA \quad \text{(21)}
\]
Upon successful completion of the initial simulations and in order to study the model’s behaviour under perturbation in the parameter values, an uncertainty modelling analysis was performed similar to the one performed on the GBL model. The probability distributions were defined by using a Spread (definition in Section 3.1.2) of 2.5 for all the parameters, except for the diffusion parameter (D) as this was assumed to have a direct effect in the concentration of the external autoinducer, therefore a wider range of values was chosen to be studied.

The complete information on the parameter distributions which were used for the simulations of both models are included in Appendix C. The simulation results for both
models can be found in Section 4.4.

4.3.2 Improved model of the scbR/scbA gene regulatory network

As previously discussed, the regulatory interactions in this system have not yet been fully elucidated. Therefore our aim was to explore the most important of the previously proposed mechanisms\cite{1,123} under the scope of realistic parameter values retrieved from literature. In order to achieve this, we designed a multifunctional unified model that includes all the potential mechanisms and enables their individual or combined use by switching certain reactions “on” and “off”. Many aspects of the model are adapted from the quorum sensing model by Weber et al.\cite{2}

A schematic representation of the regulatory interactions considered in our model is shown in Figure 4.3. The ScbR homo-dimer binds to the operators of both scbR and scbA genes and represses their activity. As reported by Bhukya et al.,\cite{257} two ScbR homo-dimers can bind to the operator. When one homo-dimer is bound, the mRNA transcription is already being repressed. As the concentration of ScbR rises, a second homo-dimer may bind to the already suppressed promoter and further enhance the suppression of the transcription. ScbA protein (A) through an enzymatic reaction with glycerol derivatives and $\beta$-keto acid derivative precursors (S) produces the $\gamma$-butyrolactones (C). Our model considers the production of C to be proportional to the concentration of A. $\gamma$-butyrolactone (C) then creates a complex with the ScbR protein ($C_2 \cdot R_2$) and thus effectively deactivate it, enabling further production of ScbA. The signalling molecules (C) diffuse passively between the cells and the environment ($C_e$) and thus accumulate in the culture medium. The model assumes that internal and external SCBs degrade at the same rate. Additionally we assume that all molecules are homogeneously distributed both in the cytoplasm and in the medium. DNA duplication, degradation of chemical species and their dilution due to cellular growth are also considered.

The following alternative scenarios are investigated:

A. The effect of transcriptional interference (collisions between the elongating RNAPs which lead to transcriptional termination) due to the overlap of the two genes’ promoter regions by 53 bp and by the convergent transcription of the two genes.
This results in a decrease in expression of full-length mRNAs from both promoters and production of truncated mRNAs.

B. The antisense effect conferred by convergent transcription of the \textit{scbR} and \textit{scbA} genes. In this case, transcripts with a segment of complementary sequence may lead to interactions between sense-antisense full length transcripts of the two genes, thus leading to the formation of a fast degrading complex of the two mRNAs and inhibition of translation.

C. The formation of a complex between ScbA and ScbR proteins (ScbA–ScbR), which relieves ScbR repression, while at the same time activating the transcription of \textit{scbA} and, in effect, the production of SCBs. This complex would serve as positive feedback loop with a role similar to the LuxR–AHL complex in the QS system.

D. All three above scenarios combined.
Figure 4.3

Chapter 4. Modelling the γ-butyrolactone network in Streptomyces coelicolor
Figure 4.3: Schematic representation of the potential mechanisms of the ScbA/ScbR system. The scbR and scbA genes are divergently encoded and their promoter regions (O_R and O_A respectively) overlap by 53bp. Due to this promoter structure, RNA polymerase collisions prevent the transcription from both promoters at the same time. The mRNAs transcribed from scbR (r) and scbA (a) may also form a complex (r-a) which rapidly degrades, thus resulting in translational inhibition. The ScbR protein forms a homo-dimer (R_2) which represses its own transcription as well as the transcription of scbA. The ScbA protein (A) is responsible for the production of the γ-butyrolactones (C) which bind to R_2 and prevent it from binding to the O_R and O_A promoters. Additionally, A may form a complex with R_2 which simultaneously prevents R_2 inhibition and activates the transcription of a by binding to a hypothetical O_A′ promoter. Finally, C diffuses freely from the cell to the external environment where it accumulates and can potentially diffuse into other neighbouring cells. Scenario A considers only the transcriptional interference, scenario B includes the transcriptional interference and the antisense RNA interactions, scenario C considers the transcriptional interference and the formation of the hypothetical AR_2 complex, and scenario D includes all of the proposed mechanisms.
The full model comprises two compartments (cell and environment), 37 chemical reactions and 51 parameters. The initial concentrations for all species are zero; the only exception are the promoters $O_R$ and $O_A$: for these one copy of each gene and therefore of its corresponding promoter is assumed for each cell. The list of the model species and the complete set of reactions for all scenarios are listed in Tables 4.3 and 4.4. The corresponding differential equations for each species are listed in Table 4.5.
Table 4.3: List of species for the ScbR/ScbA model

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Compartment</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>Operator site of ScbR upstream (scbR) promoter</td>
<td>Cell</td>
</tr>
<tr>
<td>OA</td>
<td>Operator site of ScbR within (scbA) promoter as well as putative operator site (O'_A) of AR(_2) complex</td>
<td>Cell</td>
</tr>
<tr>
<td>OR(_2)</td>
<td>Complex of R protein and operator site upstream (scbR) promoter</td>
<td>Cell</td>
</tr>
<tr>
<td>OR(_2)_2R(_2)</td>
<td>Complex of two R proteins and operator site upstream (scbR) promoter</td>
<td>Cell</td>
</tr>
<tr>
<td>OA(_2)</td>
<td>Complex of R protein and operator site within (scbA) promoter</td>
<td>Cell</td>
</tr>
<tr>
<td>OA(_2)_2R(_2)</td>
<td>Complex of two R proteins and operator site within (scbA) promoter</td>
<td>Cell</td>
</tr>
<tr>
<td>OA(_2)_AR(_2)</td>
<td>Complex of putative operator site presumed to activate (scbA) gene and AR(_2)</td>
<td>Cell</td>
</tr>
<tr>
<td>OA(_2)_AR(_2)_AR(_2)</td>
<td>Complex of operator site within (scbA) promoter with AR(_2) and R(_2)</td>
<td>Cell</td>
</tr>
<tr>
<td>OA(_2)_R(_2)_AR(_2)_AR(_2)</td>
<td>Complex of operator site within (scbA) promoter with AR(_2) and two R(_2) molecules</td>
<td>Cell</td>
</tr>
<tr>
<td>r</td>
<td>mRNA transcript of (scbR) gene</td>
<td>Cell</td>
</tr>
<tr>
<td>a</td>
<td>mRNA transcript of (scbA) gene</td>
<td>Cell</td>
</tr>
<tr>
<td>r_a</td>
<td>Complex of full length (scbR) and (scbA) mRNAs due to antisense effect</td>
<td>Cell</td>
</tr>
<tr>
<td>R</td>
<td>ScbR protein</td>
<td>Cell</td>
</tr>
<tr>
<td>R(_2)</td>
<td>ScbR homo-dimer</td>
<td>Cell</td>
</tr>
<tr>
<td>A</td>
<td>ScbA protein</td>
<td>Cell</td>
</tr>
<tr>
<td>C</td>
<td>SCB ((\gamma)-butyrolactone)</td>
<td>Cell</td>
</tr>
<tr>
<td>AR(_2)</td>
<td>Putative ScbA-ScbR complex</td>
<td>Cell</td>
</tr>
<tr>
<td>S</td>
<td>Glycerol derivative and (\beta)-keto acid derivative precursors</td>
<td>Cell</td>
</tr>
<tr>
<td>C(_2)_R(_2)</td>
<td>SCBs-ScbR complex</td>
<td>Cell</td>
</tr>
<tr>
<td>C(_e)</td>
<td>Extracellular SCB ((\gamma)-butyrolactone)</td>
<td>Environment</td>
</tr>
</tbody>
</table>

Table 4.4: Chemical reactions for the ScbR/ScbA model
Chapter 4. Modelling the γ-butyrolactone network in *Streptomyces coelicolor*

<table>
<thead>
<tr>
<th>Equation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( OR + R_2 \xrightleftharpoons[k_1']{k_1''/K_{d1}} OR \cdot R_2 )</td>
<td>(1)</td>
</tr>
<tr>
<td>( OR \cdot R_2 + R_2 \xrightleftharpoons[k_7']{k_7''/K_{d7}} OR \cdot 2R_2 )</td>
<td>(2)</td>
</tr>
<tr>
<td>( OA + R_2 \xrightleftharpoons[k_2']{k_2''/K_{d2}} OA \cdot R_2 )</td>
<td>(3)</td>
</tr>
<tr>
<td>( OA \cdot R_2 + R_2 \xrightleftharpoons[k_8']{k_8''/K_{d8}} OA \cdot 2R_2 )</td>
<td>(4)</td>
</tr>
<tr>
<td>( A + R_2 \xrightleftharpoons[k_3']{k_3''/K_{d3}} AR_2 )</td>
<td>(5)*1</td>
</tr>
<tr>
<td>( 2C + R_2 \xrightleftharpoons[k_4']{k_4''/K_{d4}} C_2 \cdot R_2 )</td>
<td>(6)</td>
</tr>
<tr>
<td>( OA + AR_2 \xrightleftharpoons[k_5']{k_5''/K_{d5}} OA \cdot AR_2 )</td>
<td>(7)*1</td>
</tr>
<tr>
<td>( OA \cdot R_2 + AR_2 \xrightleftharpoons[k_9']{k_9''/K_{d9}} OA \cdot R_2 \cdot AR_2 )</td>
<td>(8)*1</td>
</tr>
<tr>
<td>( OA \cdot 2R_2 + AR_2 \xrightleftharpoons[k_{10}']{k_{10}''/K_{d10}} OA \cdot 2R_2 \cdot AR_2 )</td>
<td>(9)*1</td>
</tr>
<tr>
<td>( OA \cdot AR_2 + R_2 \xrightleftharpoons[k_11']{k_11''/K_{d11}} OA \cdot R_2 \cdot AR_2 )</td>
<td>(10)*1</td>
</tr>
<tr>
<td>( OA \cdot R_2 \cdot AR_2 + R_2 \xrightleftharpoons[k_{12}']{k_{12}''/K_{d12}} OA \cdot 2R_2 \cdot AR_2 )</td>
<td>(11)*1</td>
</tr>
<tr>
<td>( A(S) \xrightarrow[K_C]{T_R} A + C )</td>
<td>(12)</td>
</tr>
<tr>
<td>( OR \xrightarrow[T_R]{T_R} OR + r )</td>
<td>(13)</td>
</tr>
<tr>
<td>( OA \xrightarrow[T_\alpha]{T_\alpha + T_\text{trans}} OA + a )</td>
<td>(14)</td>
</tr>
<tr>
<td>( OA \cdot AR_2 \xrightarrow[T_\alpha]{T_\alpha} OA \cdot AR_2 + a )</td>
<td>(15)*1</td>
</tr>
<tr>
<td>( a + r \xrightleftharpoons[k_{a}]{} r \cdot a )</td>
<td>(16)*2</td>
</tr>
<tr>
<td>( r \xrightarrow[p_R]{T_R} r + R )</td>
<td>(17)</td>
</tr>
<tr>
<td>( a \xrightarrow[p_A]{T_R} a + A )</td>
<td>(18)</td>
</tr>
<tr>
<td>( 2R \xrightleftharpoons[k_6']{k_6''/K_{d6}} R_2 )</td>
<td>(19)</td>
</tr>
<tr>
<td>( C \xrightarrow[D_D]{D} C_e )</td>
<td>(20)</td>
</tr>
<tr>
<td>( r \xrightarrow[d_{mR}]{D} \emptyset )</td>
<td>(21)</td>
</tr>
<tr>
<td>( R \xrightarrow[d_R]{D} \emptyset )</td>
<td>(22)</td>
</tr>
<tr>
<td>( R_2 \xrightarrow[d_{R2}]{D} \emptyset )</td>
<td>(23)</td>
</tr>
<tr>
<td>( a \xrightarrow[d_{mA}]{D} \emptyset )</td>
<td>(24)</td>
</tr>
</tbody>
</table>
\[
A \xrightarrow{d_A} \emptyset \quad \text{(25)}
\]
\[
C \xrightarrow{dc} \emptyset \quad \text{(26)}
\]
\[
C_2 \cdot R_2 \xrightarrow{dcR} \emptyset \quad \text{(27)}
\]
\[
AR_2 \xrightarrow{dAR_2} \emptyset \quad \text{(28)*1}
\]
\[
r \cdot a \xrightarrow{dmAR_2} \emptyset \quad \text{(29)*2}
\]
\[
O_R \cdot 2R_2 \xrightarrow{\mu} O_R \cdot 2R_2 + O_R \quad \text{(31)}
\]
\[
O_R \cdot R_2 \xrightarrow{\mu} O_R \cdot R_2 + O_R \quad \text{(32)}
\]
\[
O_A \cdot R_2 \xrightarrow{\mu} O_A \cdot R_2 + O_A \quad \text{(33)}
\]
\[
O_A \cdot AR_2 \xrightarrow{\mu} O_A \cdot AR_2 + O_A \quad \text{(34)}
\]
\[
O_A \cdot 2R_2 \xrightarrow{\mu} O_A \cdot 2R_2 + O_A \quad \text{(35)}
\]
\[
O_A \cdot R_2 \cdot AR_2 \xrightarrow{\mu} O_A \cdot R_2 \cdot AR_2 + O_A \quad \text{(36)}
\]
\[
O_A \cdot 2R_2 \cdot AR_2 \xrightarrow{\mu} O_A \cdot 2R_2 \cdot AR_2 + O_A \quad \text{(37)}
\]

*1 Reactions (5), (7)-(11), (15) and (28) occur only in scenarios C and D which include the activating complex AR2

*2 Reactions (16) and (29) occur only in scenarios B and D which include the antisense RNA interactions

**Table 4.5: Differential equations for the ScbR/ScbA model**

\[
\frac{d[O_R]}{dt} = -\frac{k_{-}}{K_{d1}} \cdot [O_R] \cdot [R_2] + k_{1}^{-} \cdot [O_R \cdot R_2] + \mu \cdot [OR \cdot 2R_2] + \mu \cdot [OR \cdot R_2]
\]
\[
\frac{d[O_A]}{dt} = \mu \cdot [OA \cdot 2R_2] + \mu \cdot [OA \cdot R_2] + \mu \cdot [OA \cdot AR_2] + \mu \cdot [OA \cdot R_2 \cdot AR_2]
\]
\[
+ \mu \cdot [OA \cdot 2R_2 \cdot AR_2] + \frac{k_{3}^{-}}{K_{d2}} \cdot [O_A] \cdot [R_2] + k_{2}^{-} \cdot [O_A \cdot R_2] - \frac{k_{2}}{K_{d5}} \cdot [O_A] \cdot [AR_2]
\]
\[
+ k_{5}^{-} \cdot [O_A \cdot AR_2]
\]
\[
\frac{d[O_R \cdot 2R_2]}{dt} = \frac{k_{3}^{+}}{K_{d2}} \cdot [O_R \cdot R_2] \cdot [R_2] - k_{4}^{-} \cdot [O_R \cdot 2R_2] - \mu \cdot [OR \cdot 2R_2]
\]
\[
\frac{d[O_R \cdot R_2]}{dt} = \frac{k_{4}^{+}}{K_{d2}} \cdot [O_R] \cdot [R_2] - k_{1}^{-} \cdot [O_R \cdot R_2] - \frac{k_{4}}{K_{d7}} \cdot [O_R \cdot R_2] \cdot [R_2] + k_{7}^{-} \cdot [O_R \cdot R_2] - \mu \cdot [OR \cdot R_2]
\]
\[
\frac{d[O_A \cdot R_2]}{dt} = \frac{k_{5}^{+}}{K_{d5}} \cdot [O_A] \cdot [R_2] - k_{2}^{-} \cdot [O_A \cdot R_2] + \frac{k_{3}}{K_{d6}} \cdot [O_A \cdot R_2] \cdot [R_2]
\]
\[
+ \frac{k_{8}^{+}}{K_{d6}} \cdot [O_A \cdot 2R_2] - \frac{k_{9}}{K_{d9}} \cdot [O_A \cdot R_2] \cdot [AR_2] + k_{10}^{-} \cdot [O_A \cdot R_2 \cdot AR_2] - \mu \cdot [OA \cdot R_2]
\]
\[
\frac{d[O_A \cdot AR_2]}{dt} = \frac{k_{6}^{+}}{K_{d6}} \cdot [O_A] \cdot [AR_2] - k_{5}^{-} \cdot [O_A \cdot AR_2] - \frac{k_{11}}{K_{d11}} \cdot [O_A \cdot AR_2] \cdot [R_2] + k_{11}^{-} \cdot [O_A \cdot AR_2 \cdot R_2] - \mu \cdot [OA \cdot AR_2]
\]
\[
\frac{dr}{dt} = T_R \cdot [O_R] - \frac{k_{m}}{K_{ar}} \cdot [r] \cdot [a] + k_{ar}^{-} \cdot [r] \cdot [a] - d_{mR} \cdot [r] - \mu \cdot [r]
Chapter 4. Modelling the γ-butyrolactone network in Streptomyces coelicolor

\[
\frac{d[a]}{dt} = \left( T_{basal} \cdot [O_A] \right) \cdot [T_A] \cdot [O_A] + T_A \cdot [O_A \cdot AR_2] - \frac{k_{ar}}{K_{ar}} \cdot [r] \cdot [a] + k_{ar}^{-} \cdot [r \cdot a] - d_{mA} \cdot [a] - \mu \cdot [a]
\]

\[
\frac{d[r \cdot a]}{dt} = \frac{k_{ar}^{-}}{K_{ar}} \cdot [r] \cdot [a] - k_{ar}^{-} \cdot [r \cdot a] - d_{mRA} \cdot [r \cdot a] - \mu \cdot [r \cdot a]
\]

\[
\frac{d[R]}{dt} = P_R \cdot [r] - \frac{k_{ar}^{-}}{K_{ar}} \cdot [R]^2 + k_{ar}^{-} \cdot [R]_2 - d_R \cdot [R] - \mu \cdot [R]
\]

\[
\frac{d[R_2]}{dt} = \frac{k_{ar}^{-}}{K_{ar}} \cdot [R]^2 - k_{ar}^{-} \cdot [R]_2 + k_{ar}^{-} \cdot [R]_2 - \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_k] \cdot [R]_2 + k_{ar}^{-} \cdot [O_R \cdot R_2]
+ k_{ar}^{-} \cdot [C_2 \cdot R_2] - \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_R \cdot [R_2] + k_{ar}^{-} \cdot [O_R \cdot 2R_2] - \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_A \cdot R_2] \cdot [R_2] + k_{ar}^{-} \cdot [O_A \cdot R_2 \cdot AR_2]
+ \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_A \cdot 2R_2] \cdot AR_2 - k_{ar}^{-} \cdot [AR_2] + k_{ar}^{-} \cdot [O_A \cdot 2R_2] \cdot AR_2 - d_{AR} \cdot [AR_2] - \mu \cdot [AR_2]
\]

\[
\frac{d[A]}{dt} = P_A \cdot [a] - \frac{k_{ar}^{-}}{K_{ar}} \cdot [A] \cdot [R_2] + k_{ar}^{-} \cdot [AR_2] - d_A \cdot [A] - \mu \cdot [A]
\]

\[
\frac{d[C]}{dt} = K_C \cdot [C] \cdot [C] \cdot [AR_2] - \frac{k_{ar}^{-}}{K_{ar}} \cdot [C]^2 \cdot [R_2] + k_{ar}^{-} \cdot [C_2 \cdot R_2] - d_C \cdot [C] - \mu \cdot [C] + D \cdot ([C_e] - [C])
\]

\[
\frac{d[AR_2]}{dt} = -\frac{k_{ar}^{-}}{K_{ar}} \cdot [A] \cdot [R_2] + k_{ar}^{-} \cdot [AR_2] - \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_A] \cdot [AR_2] + k_{ar}^{-} \cdot [O_A \cdot AR_2]
+ k_{ar}^{-} \cdot [O_A \cdot 2R_2] \cdot AR_2 - d_{AR} \cdot [AR_2] - \mu \cdot [AR_2]
\]

\[
\frac{d[C_2 \cdot R_2]}{dt} = \frac{k_{ar}^{-}}{K_{ar}} \cdot [C]^2 \cdot [R_2] - k_{ar}^{-} \cdot [C_2 \cdot R_2] - d_{CR} \cdot [C_2 \cdot R_2] - \mu \cdot [C_2 \cdot R_2]
\]

\[
\frac{d[C_e]}{dt} = \rho \cdot D \cdot ([C] - [C_e]) - d_C \cdot [C_e] - \mu \cdot [C_e]
\]

\[
\frac{d[O_A \cdot 2R_2]}{dt} = \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_A \cdot R_2] \cdot [R_2] - k_{ar}^{-} \cdot [O_A \cdot 2R_2] - \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_A \cdot 2R_2] \cdot [AR_2]
+ k_{ar}^{-} \cdot [O_A \cdot 2R_2] \cdot AR_2 - \mu \cdot [O_A \cdot 2R_2]
\]

\[
\frac{d[O_A \cdot R_2 - AR_2]}{dt} = \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_A \cdot R_2] \cdot [AR_2] - k_{ar}^{-} \cdot [O_A \cdot R_2 \cdot AR_2] + \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_A \cdot AR_2] \cdot [R_2]
- k_{ar}^{-} \cdot [O_A \cdot R_2 \cdot AR_2] - \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_A \cdot R_2 \cdot AR_2] \cdot [R_2]
+ k_{ar}^{-} \cdot [O_A \cdot 2R_2 \cdot AR_2] - \mu \cdot [O_A \cdot R_2 \cdot AR_2]
\]

\[
\frac{d[O_A \cdot R_2 \cdot AR_2]}{dt} = \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_A \cdot 2R_2] \cdot [AR_2] - k_{ar}^{-} \cdot [O_A \cdot 2R_2 \cdot AR_2]
+ \frac{k_{ar}^{-}}{K_{ar}} \cdot [O_A \cdot R_2 \cdot AR_2] \cdot [R_2] - k_{ar}^{-} \cdot [O_A \cdot 2R_2 \cdot AR_2]
- \mu \cdot [O_A \cdot 2R_2 \cdot AR_2]
\]

**State of promoters and transcription**

The mechanisms involved in the transcription are defined in reactions (1)–(4), (7)–(11) and (13)–(15). In order to describe the overlapping promoter effects, the transcription reactions of each gene need to take into account the strength and the state of the gene’s promoter (free or occupied), as well as the potential interference from the transcription
of the opposite gene. In order to accomplish that, the mathematical model for overlapping promoters proposed by Bendtsen et al.\cite{bendtsen2005} was employed. In our reactions, the transcription process depends on the transcription rates ($T_R$ and $T_A$). These are derived via the equations 4.1-4.8 by employing the promoter firing rates ($k_{FR}$, $k_{FA}$), the strength of the promoters ($\Omega_R$, $\Omega_A$), the occupancy of the promoters ($\theta_R$, $\theta_A$) and the rate of RNP binding on the promoters ($k_{on_R}$, $k_{on_A}$).

The model assumes that the strength of each promoter in the isolated state (uncoupled promoters) is equal to the number of RNA transcripts produced per unit of time. Therefore, the parameters $\Omega_R$ and $\Omega_A$ are set to be equal to the transcription rate constants of the isolated promoters. In order to describe the difference in the promoter strengths, the general firing rate $k_F$ is initially set to be common for both promoters, in accordance with the literature values. Afterwards, a heterogeneity factor ($\chi$) is employed to define the difference between the two individual firing rates, as per:

\[
k_{FR} = k_F \cdot \chi \quad \text{and} \quad k_{FA} = \frac{k_F}{\chi}.
\]

The promoters can be assumed to have equal strength in the isolated state ($\chi = 1$), so their firing rates are equal ($k_{FR} = k_{FA} = k$). Alternatively, $\chi$ can have values in the range between 0.1 and 1 in order for promoter $O_A$ to be stronger, or the reciprocal values (1 to 10) in order to reinforce promoter $O_R$.

In order to include the strength of the promoter in the derivation of the transcription rate, the following formulas are used:

\[
\frac{1}{k_{on_R}} = \frac{1}{\Omega_R} - \frac{1}{k_{FR}} \quad (4.1)
\]

\[
\frac{1}{k_{on_A}} = \frac{1}{\Omega_A} - \frac{1}{k_{FA}} \quad (4.2)
\]

The maximal occupancy for the isolated scbR promoter can be calculated as per:

\[
\theta_R^o = \frac{k_{on_R}}{k_{on_R} + k_{FR}} \quad (4.3)
\]

\[
\theta_A^o = \frac{k_{on_A}}{k_{on_A} + k_{FA}} \quad (4.4)
\]
The promoter aspect ratios $\alpha_R$ and $\alpha_A$ can then be derived from the equations:

$$\alpha_R = \frac{k_{onR}}{k_{FR}} = \frac{\theta_R^o}{1 - \theta_R^o}$$  \hspace{1cm} (4.5)

$$\alpha_A = \frac{k_{onA}}{k_{FA}} = \frac{\theta_A^o}{1 - \theta_A^o}$$  \hspace{1cm} (4.6)

In our case, as the two promoters are overlapping, there will be three possibilities:

a) $scbR$ promoter is occupied by the RNAP (probability = $\theta_R$), b) $scbA$ promoter is occupied by the RNAP (probability = $\theta_A$), c) both promoters are unoccupied (probability = $1 - \theta_R - \theta_A$). Therefore the occupancy for the coupled $scbR$ and $scbA$ promoters ($\theta_R$, $\theta_A$) is:

$$\theta_R = \frac{\alpha_R + 1}{\alpha_R + \alpha_A + 1}$$  \hspace{1cm} (4.7)

$$\theta_A = \frac{\alpha_A + 1}{\alpha_R + \alpha_A + 1}$$  \hspace{1cm} (4.8)

and the final transcription rate constant that is used for the model ODEs can be calculated as per: $T_R = k_F R \cdot \theta_R$ and $T_A = k_F A \cdot \theta_A$.

As shown in reactions (13) and (14), in the general case, the two genes are transcribed at the maximal rate when they are not repressed by the ScbR homo-dimer (species OA and OR). The transcription stops when even one homo-dimer binds to the gene operator ($O_R \cdot R_2$ and $O_A \cdot R_2$) and the suppression is further ensured by the binding of a second homo-dimer ($O_R \cdot 2R_2$ and $O_A \cdot 2R_2$) at a different time point (reactions (1)-(4)).

In the mechanisms that include the putative ScbA-ScbR complex ($AR_2$), reaction (14) only accounts for a transcription at a basal rate. In order to achieve the maximal rate, the $AR_2$ complex needs to bind to $O'_A$ ($O_A \cdot AR_2$) (reactions (7) and (15)). In our model, $O_A$ and $O'_A$ are considered together under the general name OA which can be found at different states. The transcription is again suppressed when the ScbR homo-dimers bind to OA ($O_A \cdot R_2 \cdot AR_2$ and $O_A \cdot 2R_2 \cdot AR_2$), even though the activating complex may also be bound (reactions (8)-(11)).
Cell growth and division

Our deterministic model does not consider cells as individuals, but assumes they are combined in one single common cytoplasm (Figure 4.4) whose total volume $V_{c,\text{tot}}$ can be calculated by the addition of the individual cell volumes ($V_c$). Therefore, the concentrations of the cellular species are averaged over all the cells of the culture ($V_{c,\text{tot}}$).

As the experimental time duration is over 60 hours, the effects of cell growth must be taken into consideration in addition to the various regulatory mechanisms. In our model, the number of cells is described by a logistic growth curve $N(t) = \frac{K}{1 + A e^{-\mu_{\text{max}} t}}$, where $A = \frac{K-N_o}{N_o}$, $K$ is the carrying capacity, $\mu_{\text{max}}$ is the maximal growth rate and $N_o$ is the initial number of cells (inoculum). The specific growth rate can be calculated as per: $\mu(N) = \mu_{\text{max}} \cdot \left(1 - \frac{N(t)}{K}\right)$ and describes the change in the growth rate according to the cell population status. By using this equation, the specific growth rate depends only on the actual number of cells and does not take into account their internal physiological state.

This means that the initial lag phase, where the growth rate is extremely low, is omitted and instead growth is assumed to be maximal at time zero and to slowly decrease as the population reaches the value of the carrying capacity. A more accurate way to describe the change in the specific growth rate would be the non-autonomous equation of the Baranyi & Roberts model\cite{259,260} which takes into account the lag phase by using an adjustment function. However, we also have to take into consideration the fact that the cells are still adapting to their new environment during the lag phase, so they do not produce any of the intracellular species of interest in a significant amount, and that the initial conditions defined in our model where all species except the DNA are set to zero, do not correspond to a natural state. For these reasons and in order to avoid over-fitting our model, we employ the non-modified equation but take into consideration the fact that the system is adjusting to the physiological conditions for the first 10–15 hours.

Therefore, the dynamics of the total volume of the cells is $V_{c,\text{tot}}(t) = V_0 \cdot N(t) = V_0 \cdot \frac{K}{1 + A e^{-\mu_{\text{max}} t}}$, where $V_0$ is the volume of a single cell. As a consequence, the cellular growth causes the dilution of the intracellular species concentrations in at a rate of $-\mu(N) \cdot C_x(t)$, where $C_x(t)$ is the concentration of a species X in time t. On the other
hand, the genetic material is duplicated during cell division. This is taken into account in our model through reactions (31) - (37), where DNA concentration is increased in a rate of $+\mu(N) \cdot (\text{Gene}_{\text{Operators bound}} + \text{Gene}_{\text{Operators free}})$ so that for DNA the cell growth dilution is compensated and the total DNA concentration is kept constant.

The volume of the culture medium is also affected by cellular growth. As the number of cells increases, the volume of the external environment decreases. So, by defining $V_{\text{tot}}$ as the total volume of the initial culture (medium and cells) which remains constant, the volume of the medium will be $V_{\text{ext}} = V_{\text{tot}} - V_{c;\text{tot}}$. By this equation it is obvious that the volume of the external environment will decrease as the number of cells is increases. The fraction ratio of the total cell volume and the external environment volume is $\rho = \frac{V_{c;\text{tot}}}{V_{\text{ext}}}$. The decrease in the external volume leads to an increase of the concentration of $C_e$ in at a rate of $+\mu(N) \cdot \rho \cdot [C_e](t)$. This also affects the diffusion of the external butyrolactones back to the cytoplasm. For this reason, the diffusion rate constant (D) of the reverse part of reaction (20) is additionally multiplied by the variable $\rho$.

---

**Figure 4.4:** Modelling approach for a growing cell culture in a fermentor. The cell culture is described by a unified volume with average and continuous concentrations of all species, including the DNA (small circles). $\gamma$-butyrolactones (small yellow pentagons) can be found both in the cytoplasm and in the culture medium. $V_{\text{tot}}$ represents the total volume of cells and medium injected in the fermentor at the beginning of the experiment and remains constant. $V_{c;\text{tot}}$ is the volume of the “cell” which increases in line with the cell population. $V_{\text{ext}}$ is the volume of the medium which decreases as the number of cells increase. All species in all compartments are assumed to be well-stirred.
Chapter 4. Modelling the γ-butyrolactone network in Streptomyces coelicolor

Parameters

For each parameter of the model, a probability distribution was designed according to the available information from literature and experiments. In order to achieve this, a dedicated MediaWiki-based website was created (http://www.systemsbiology.ls.manchester.ac.uk/wiki/index.php/MainPage) with the purpose of documenting parameter values along with explicit information on their sources and subsequent justification on beliefs about the most plausible values. By using this information, the log-normal probability distributions describing each parameter were inferred, according to the protocol described in Chapter 3.

The parameter information of the probability distributions is summarised in Table 4.6. The full information on the parameter values retrieved from literature and the design of the corresponding probability distributions is included in Appendix D.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Units</th>
<th>Mode</th>
<th>Spread</th>
<th>Location parameter (μ)</th>
<th>Scale parameter (σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d1}$</td>
<td>Dissociation constant for binding of $R_2$ to $O_R$</td>
<td>nM</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
</tr>
<tr>
<td>$K_{d7}$</td>
<td>Dissociation constant for binding of $R_2$ to $O_R \cdot R_2$</td>
<td>nM</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
</tr>
<tr>
<td>$k_1^-$</td>
<td>Dissociation rate for binding of $R_2$ to $O_R$</td>
<td>min$^{-1}$</td>
<td>0.489</td>
<td>1.98</td>
<td>0.37642</td>
<td>0.58225</td>
</tr>
<tr>
<td>$k_7^-$</td>
<td>Dissociation rate for binding of $R_2$ to $O_R \cdot R_2$</td>
<td>min$^{-1}$</td>
<td>0.489</td>
<td>1.98</td>
<td>0.37642</td>
<td>0.58225</td>
</tr>
<tr>
<td>$K_{d2}$</td>
<td>Dissociation constant for binding of $R_2$ to $O_A$</td>
<td>nM</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
</tr>
<tr>
<td>$K_{d8}$</td>
<td>Dissociation constant for binding of $R_2$ to $O_A \cdot R_2$</td>
<td>nM</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
</tr>
<tr>
<td>$k_2^-$</td>
<td>Dissociation rate for binding of $R_2$ to $O_A$</td>
<td>min$^{-1}$</td>
<td>0.489</td>
<td>1.98</td>
<td>0.37642</td>
<td>0.58225</td>
</tr>
</tbody>
</table>
### Chapter 4. Modelling the γ-butyrolactone network in Streptomyces coelicolor

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Units</th>
<th>Mode</th>
<th>Spread</th>
<th>Location parameter (μ)</th>
<th>Scale parameter (σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k^{-}_8$ [D.2]</td>
<td>Dissociation rate for binding of $R_2$ to $O_A \cdot R_2$</td>
<td>min$^{-1}$</td>
<td>0.489</td>
<td>1.98</td>
<td>0.37642</td>
<td>0.58225</td>
</tr>
<tr>
<td>$K_{d3}$ [D.3]</td>
<td>Dissociation constant for binding of $R_2$ to $A$</td>
<td>nM</td>
<td>9.9104</td>
<td>10.02</td>
<td>13.192</td>
<td>1.2977</td>
</tr>
<tr>
<td>$k^{-}_3$ [D.3]</td>
<td>Dissociation rate for binding of $R_2$ to $A$</td>
<td>min$^{-1}$</td>
<td>1879.7</td>
<td>23.13</td>
<td>9.9423</td>
<td>1.5503</td>
</tr>
<tr>
<td>$K_{d4}$ [D.4]</td>
<td>Dissociation constant for binding of $C$ to $R_2$</td>
<td>nM</td>
<td>1.58</td>
<td>730.63</td>
<td>5.9389</td>
<td>2.3412</td>
</tr>
<tr>
<td>$k^{-}_4$ [D.4]</td>
<td>Dissociation rate for binding of $C$ to $R_2$</td>
<td>min$^{-1}$</td>
<td>0.816</td>
<td>244</td>
<td>4.2872</td>
<td>2.119</td>
</tr>
<tr>
<td>$K_{d5}$ [D.5]</td>
<td>Dissociation constant for binding of $AR_2$ to $O_A$</td>
<td>nM</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
</tr>
<tr>
<td>$K_{d9}$ [D.5]</td>
<td>Dissociation constant for binding of $AR_2$ to $O_A \cdot R_2$</td>
<td>nM</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
</tr>
<tr>
<td>$K_{d10}$ [D.5]</td>
<td>Dissociation constant for binding of $AR_2$ to $O_A \cdot 2R_2$</td>
<td>nM</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
</tr>
<tr>
<td>$K_{d11}$ [D.5]</td>
<td>Dissociation constant for binding of $R_2$ to $O_A \cdot AR_2$</td>
<td>nM</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
</tr>
<tr>
<td>$K_{d12}$ [D.5]</td>
<td>Dissociation constant for binding of $R_2$ to $O_A \cdot R_2 \cdot AR_2$</td>
<td>nM</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
</tr>
<tr>
<td>$k^{-}_5$ [D.5]</td>
<td>Dissociation rate for binding of $AR_2$ to $O_A$</td>
<td>min$^{-1}$</td>
<td>0.489</td>
<td>1.98</td>
<td>0.37642</td>
<td>0.58225</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Units</td>
<td>Mode</td>
<td>Spread</td>
<td>Location parameter (μ)</td>
<td>Scale parameter (σ)</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>-------</td>
<td>------</td>
<td>--------</td>
<td>------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>$k_9$ $^{[D.5]}$</td>
<td>Dissociation rate for binding of $AR_2$ to $O_A \cdot R_2$</td>
<td>min$^{-1}$</td>
<td>0.489</td>
<td>1.98</td>
<td>0.37642</td>
<td>0.58225</td>
</tr>
<tr>
<td>$k_{10}$ $^{[D.5]}$</td>
<td>Dissociation rate for binding of $AR_2$ to $O_A \cdot 2R_2$</td>
<td>min$^{-1}$</td>
<td>0.489</td>
<td>1.98</td>
<td>0.37642</td>
<td>0.58225</td>
</tr>
<tr>
<td>$k_{11}$ $^{[D.5]}$</td>
<td>Dissociation rate for binding of $R_2$ to $O_A \cdot AR_2$</td>
<td>min$^{-1}$</td>
<td>0.489</td>
<td>1.98</td>
<td>0.37642</td>
<td>0.58225</td>
</tr>
<tr>
<td>$k_{12}$ $^{[D.5]}$</td>
<td>Dissociation rate for binding of $R_2$ to $O_A \cdot R_2 \cdot AR_2$</td>
<td>min$^{-1}$</td>
<td>0.489</td>
<td>1.98</td>
<td>0.37642</td>
<td>0.58225</td>
</tr>
<tr>
<td>$K_C$ $^{[D.6]}$</td>
<td>Synthesis rate of C</td>
<td>min$^{-1}$</td>
<td>0.094</td>
<td>60.49</td>
<td>0.87914</td>
<td>1.8017</td>
</tr>
<tr>
<td>$\Omega_R$ $^{[D.7]}$</td>
<td>Strength of the $OR$ promoter/ Transcription rate constant of the isolated $OR$ promoter</td>
<td>min$^{-1}$</td>
<td>0.8346</td>
<td>3.55</td>
<td>0.63056</td>
<td>0.90076</td>
</tr>
<tr>
<td>$k_F$ $^{[D.7,D.8]}$</td>
<td>Firing rate (elongation initiation rate) constant</td>
<td>min$^{-1}$</td>
<td>20.7</td>
<td>1.34</td>
<td>3.1107</td>
<td>0.28276</td>
</tr>
<tr>
<td>$k_{on R}$ $^{[D.7]}$</td>
<td>Rate of RNA polymerase binding to the $OR$ promoter</td>
<td>min$^{-1}$</td>
<td>0.867</td>
<td>4.09</td>
<td>0.78694</td>
<td>0.96428</td>
</tr>
<tr>
<td>$\Omega_{A_{basal}}$ $^{[D.8]}$</td>
<td>Basal transcription rate constant of the isolated $OA$ promoter</td>
<td>min$^{-1}$</td>
<td>0.0037</td>
<td>10.93</td>
<td>3.8324</td>
<td>1.3259</td>
</tr>
<tr>
<td>$\Omega_A$ $^{[D.8]}$</td>
<td>Strength of the $OA$ promoter/ Transcription rate constant of the isolated $OA$ promoter</td>
<td>min$^{-1}$</td>
<td>0.572</td>
<td>3.5</td>
<td>0.2514</td>
<td>0.89956</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Units</td>
<td>Mode</td>
<td>Spread</td>
<td>Location parameter</td>
<td>Scale parameter</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------------------</td>
<td>--------</td>
<td>------</td>
<td>---------</td>
<td>--------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>(k_{on}A)</td>
<td>Rate of RNA polymerase binding to the (O_A) promoter</td>
<td>min(^{-1})</td>
<td>0.587</td>
<td>3.875</td>
<td>0.35255</td>
<td>0.94054</td>
</tr>
<tr>
<td>(k_{on,\text{basal}})</td>
<td>Basal rate of RNA polymerase binding to the (O_A) promoter</td>
<td>min(^{-1})</td>
<td>0.00373</td>
<td>10.97</td>
<td>3.8297</td>
<td>1.3269</td>
</tr>
<tr>
<td>(K_{ar})</td>
<td>Dissociation rate for binding of r to a</td>
<td>nM</td>
<td>7.8</td>
<td>7.27</td>
<td>3.4665</td>
<td>1.1882</td>
</tr>
<tr>
<td>(k_{ar}^{-})</td>
<td>Dissociation rate for binding of r to a</td>
<td>min(^{-1})</td>
<td>0.223</td>
<td>4.8</td>
<td>0.43359</td>
<td>1.0326</td>
</tr>
<tr>
<td>(P_R)</td>
<td>Translation rate of R</td>
<td>min(^{-1})</td>
<td>0.744</td>
<td>3.2</td>
<td>0.42952</td>
<td>0.85176</td>
</tr>
<tr>
<td>(P_A)</td>
<td>Translation rate of A</td>
<td>min(^{-1})</td>
<td>0.5</td>
<td>3.2</td>
<td>0.053356</td>
<td>0.85327</td>
</tr>
<tr>
<td>(K_{db})</td>
<td>Dissociation constant for (R_2) (homo-dimer) formation</td>
<td>nM</td>
<td>3.89</td>
<td>1.9</td>
<td>1.6716</td>
<td>0.55779</td>
</tr>
<tr>
<td>(k_6^{-})</td>
<td>Dissociation rate for (R_2) (homo-dimer) formation</td>
<td>min(^{-1})</td>
<td>1.1997</td>
<td>21.7</td>
<td>2.5303</td>
<td>1.5324</td>
</tr>
<tr>
<td>(d_{mR})</td>
<td>Degradation rate of r</td>
<td>min(^{-1})</td>
<td>0.14</td>
<td>2.06</td>
<td>1.5916</td>
<td>0.60824</td>
</tr>
<tr>
<td>(d_R)</td>
<td>Degradation rate of R</td>
<td>min(^{-1})</td>
<td>0.00144</td>
<td>1.78</td>
<td>6.2837</td>
<td>0.50981</td>
</tr>
<tr>
<td>(d_{R_2})</td>
<td>Degradation rate of (R_2)</td>
<td>min(^{-1})</td>
<td>0.00144</td>
<td>1.78</td>
<td>6.2837</td>
<td>0.50981</td>
</tr>
<tr>
<td>(d_{mA})</td>
<td>Degradation rate of a</td>
<td>min(^{-1})</td>
<td>0.14</td>
<td>2.06</td>
<td>1.5916</td>
<td>0.60824</td>
</tr>
<tr>
<td>(d_A)</td>
<td>Degradation rate of A</td>
<td>min(^{-1})</td>
<td>0.00144</td>
<td>1.78</td>
<td>6.2837</td>
<td>0.50981</td>
</tr>
<tr>
<td>(d_C)</td>
<td>Degradation rate of C</td>
<td>min(^{-1})</td>
<td>0.0032</td>
<td>5.1</td>
<td>4.6234</td>
<td>1.0539</td>
</tr>
<tr>
<td>(d_{CR})</td>
<td>Degradation rate of (C_R \cdot R_2)</td>
<td>min(^{-1})</td>
<td>0.00144</td>
<td>1.78</td>
<td>6.2837</td>
<td>0.50981</td>
</tr>
<tr>
<td>(d_{AR})</td>
<td>Degradation rate of (AR_2)</td>
<td>min(^{-1})</td>
<td>0.00144</td>
<td>1.78</td>
<td>6.2837</td>
<td>0.50981</td>
</tr>
<tr>
<td>(d_{mRA})</td>
<td>Degradation rate of antisense RNA complex (r \cdot a)</td>
<td>min(^{-1})</td>
<td>4.9</td>
<td>14.1</td>
<td>2.5669</td>
<td>0.98835</td>
</tr>
</tbody>
</table>
The parameters for the cellular growth were derived by the experimental data for *Streptomyces* growth reported by Nieselt *et al.*\(^{261}\) By fitting the logistic equation for bacterial growth

\[ N(t) = \frac{K}{1 + \frac{K}{N_0} e^{-\mu_{\text{max}} t}} \]

to the number of cells (approximately calculated from the reported biomass values), the carrying capacity (K), the initial cell number (\(N_0\)) and the maximum growth rate were estimated. The prediction of the carrying capacity by the logistic equation was \(8.16 \cdot 10^{12}\) cells, which is very close to the \(8.24 \cdot 10^{12}\) cells calculated from the final biomass. Additionally, the initial number of cells predicted by the logistic equation was \(3.54 \cdot 10^9\) cells which is also very close to the inoculum value of \(10^9\) cells reported in the publication. The fitted data and the estimated logistic growth curve are shown in Figure 4.5. The parameter values for cellular growth are summarised in Table 4.7.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Units</th>
<th>Mode</th>
<th>Spread</th>
<th>Location parameter ((\mu))</th>
<th>Scale parameter ((\sigma))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D^{[D,22]})</td>
<td>Diffusion rate of C through the cell membrane</td>
<td>(\text{min}^{-1})</td>
<td>2.8</td>
<td>4.27</td>
<td>1.9947</td>
<td>0.98233</td>
</tr>
</tbody>
</table>
Chapter 4. Modelling the γ-butyrolactone network in Streptomyces coelicolor

Figure 4.5: Fitting the logistic growth curve (N(t)) to the reported experimental data by Nieselt et al.\textsuperscript{[261]} The carrying capacity (K) is the maximum plateau reached after 50h of growth. $\mu_{\text{max}}$ is the maximal growth rate achieved during the exponential phase of the growth. The exact parameter values used for fitting the experimental data to the equation are summarised in Table 4.7.

![Logistic growth curve fit](image)

**Figure 4.5:** Fitting the logistic growth curve (N(t)) to the reported experimental data by Nieselt et al.\textsuperscript{[261]} The carrying capacity (K) is the maximum plateau reached after 50h of growth. $\mu_{\text{max}}$ is the maximal growth rate achieved during the exponential phase of the growth. The exact parameter values used for fitting the experimental data to the equation are summarised in Table 4.7.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>Carrying capacity</td>
<td>$8.16 \cdot 10^{12}$</td>
<td>cells</td>
</tr>
<tr>
<td>$N_o$</td>
<td>Initial number of cells</td>
<td>$3.54 \cdot 10^9$</td>
<td>cells</td>
</tr>
<tr>
<td>$\mu_{\text{max}}$</td>
<td>Maximum growth rate</td>
<td>$0.004648$</td>
<td>$min^{-1}$</td>
</tr>
<tr>
<td>$V_0$</td>
<td>Volume of single cell</td>
<td>$1.2 \mu m^3$</td>
<td></td>
</tr>
<tr>
<td>$V_{\text{tot}}$</td>
<td>Total cell culture volume</td>
<td>$1.8 L$</td>
<td></td>
</tr>
</tbody>
</table>

In order to complete the model simulations, 5000 parameter sets were sampled from the defined probability distributions. By using the same parameter sets, 5000 simulations were performed for each of the four scenarios. The simulations assumed equal strength for both promoters ($k_{F_R} = k_{F_A}$, $\chi = 1$).
In order to also test the impact of the promoter strength, two additional groups of simulations were repeated with the same values for all parameters except from the heterogeneity factor $\chi$. In the first group, $\chi$ was sampled from a log-normal distribution within the range of 0.1 and 0.9 ($k_{Fr} < k_{Fa}$) and in the second group from a log-normal distribution between 1.1 and 10 ($k_{Fr} > k_{Fa}$). The detailed description on the design of the additional probability distributions is included in Appendix D.

4.4 Results

4.4.1 Results from the recreation of GBL and QS models

The simulations results from the recreation of the GBL model are presented in Figure 4.6. The plots show analogous behaviour for all species with the published model, although there are some small differences in the maximum concentration values attained by the species. The most distinctive difference is seen the concentration of ScbA and scbA mRNA. In the replicated model ScbA reached 291.77 nM, while in the published model’s plots it seems to reach a maximum of 210 nM. Similarly, in the replicated model scbA mRNA reached 8.67 nM while in the published model’s plots it seems to reach a maximum of 6.5 nM. The investigation on the bistable region resulted in a bistability zone range between 80 and 96 nM of SCB1 concentration, while the publication reported a range between 21 and 76 nM. These differences could be due to the initial values of the species which were assumed, as they were not explicitly reported in the publication, to the values of the parameters (as a parameter range for bistability is reported, but the simulations were conducted only with the reported single optimal value for bistability for each parameter), or to another assumption about the system’s state which was not clearly mentioned in the publication. It was also noted that all of these values were quite far from the threshold for the system switch on reported in the literature (250 nM).
Chapter 4. Modelling the γ-butyrolactone network in Streptomyces coelicolor

Figure 4.6: Simulations results from the replication of published model on GBL system in Copasi. The arrows near the curves point to their corresponding y-axis

The uncertainty modelling showed that the model works properly only within the range of values defined for bistability. The initial set of distributions included in Appendix D, covered the largest part of the range of tested values as reported in the publication. When used as input for the model, the simulations were either terminated prematurely (the model crashed) or exhibited abnormal results, such as species concentrations increasing to infinity or obtaining negative values. The simulations were
repeated with tighter confidence interval boundaries, so that the parameter values cover only the range of bistability (the Mode was set as the optimal value for bistability and the Spread was set to 1.2–1.4 for most of the species).

The second set of simulations showed better results (Figure 4.7) as none of the simulations were terminated prematurely and none exhibited abnormal behaviour. The threshold for optimal model behaviour was initially set at 0–800 nM for ScbR, 0–60000 nM for SCB1 and 0–2000 nM for ScbA. With these criteria, 6 models showed abnormal behaviour and 994 exhibited behaviour within the selected boundaries. However, when setting a stricter boundary of normal behaviour of 0–800 nM for ScbR, 10–60000 nM for SCB1 and 3–2000 nM for ScbA in order to eliminate the lower bound noise, all 1000 models were rejected. This shows that even if an individual species followed a “normal” trajectory during the simulation, some other species in the model never accumulated to concentrations compatible with the experimental information on the system. Finally, the SCB1 results showed a maximum value of 40000 nM, while in the bistable model the value increased above 120000 nM, which could mean that the behaviour reported in the publication could have been an outlier, rather than representative of the average model output.
Chapter 4. Modelling the γ-butyrolactone network in Streptomyces coelicolor

(a) Concentration profile of ScbA (A). The majority of the simulations showed qualitatively similar behaviour with the publication figures.

(b) Concentration profiles of ScbR. Diverse behaviours were observed for different parameter sets, some of them showing the system in the “off” state as reported in the publication (concentration of ScbR constantly above 270nM), some of them in the “on” state (concentration of ScbR dropping below 100nM) and some in the bistable state (concentration of ScbR decreasing but remaining between 110 and 270nM).

(c) Concentration profiles of intracellular SCB1 (Ci). The maximum value reached by the successful models was below 40000 nM, much lower than the value reported in the publication (120000 nM).

Figure 4.7: Simulations results from the parameter uncertainty test on the published GBL model by Mehra et al.\textsuperscript{[1]} in MATLAB, by using parameter distributions containing values from the range of bistability. The blue spaghetti plots show the results of the 1000 simulations and the red curves are the single deterministic run with the set of parameters defined for bistability.

With regards to the analysis of the QS system, the simulations’ results for the deterministic model are presented in Figure 4.8. The plots show analogous behaviour of GFP with the published model’s\textsuperscript{[2]} and the species concentration values obtained are also matching the reported values. The maximum GFP concentration in the recreated model was 1375.12 nM in simulations performed with Copasi and 1361.2 nM with
MATLAB, which are approximating the 1372.54 nM maximum value reported for the published model.

The region of bistability was also investigated (Figure 4.9) by simulating the steady state response of the system with the addition of small amounts of the external autoinducer $C_A^*$ from 0 nM to 100 nM. The process was then repeated by starting from fully induced cells at 100 nM of $C_A^*$ and investigate the steady state response of the system until all $C_A^*$ was removed. The analysis revealed a bistable region between 0 nM and 15 nM, in agreement with the published values.

*Figure 4.8:* Simulations results from replication of deterministic model on QS system (lux02 operon). The dark blue arrowed curve shows the response under 10 h induction time for gradually increasing $C_A^*$ concentration. The lower dark blue curve without arrow shows the transient response after 2 hours of induction from initially non-induced cells. The arrowed slashed lines show the autoinducer decreasing-concentration trajectories for cells weakly induced (2 hours) at $C_A^*=100$ nM, 75 nM and 50 nM. The value of $C_A^*$ is hourly decreased by 25%.
Figure 4.9: Identification of the bistability zone in the QS system. The steady state response of the system for increasing (blue curve) and decreasing (red curve) the external autoinducer concentration, revealed a bistable region of $0 \text{nM} < C_A < 15 \text{nM}$ (grey area), identical with the one reported in the publication.

The uncertainty analysis of the QS model produced results consistent with the published model behaviour (Figures 4.10a and 4.10c). In order to allow consistent comparison between the ensemble of results and the published model, GFP_{max} was set to be equal to the one maximum reported in the publication, and the ratio GFP/GFP_{max} was used for the plots. None of the simulations was terminated prematurely; however 190 models reached their maximum at very low values (ratio GFP/GFP_{max} was below 0.1).

As a next step, the bistability zone behaviour was also investigated under parameter uncertainty. All 1000 simulations were subjected to the steady state response under constant supply or depletion of the external autoinducer, by following the same procedure as with the single parameter set. The simulations showed that the bistability zone is also affected by the change in the parameter values. The majority of the models (915) did not exhibit hysteresis and only 85 models had even a small bistable region. From those, the majority had a bistability zone width of 1–2nM with the most prominent values for switching “on” and “off” being 1 nM and 2 nM of $C_A$. However, the bistability zone width extended up to 98, with some simulations making the “jump” even at 99 nM (Figures 4.12a, 4.12c and 4.12e).
Profiles from the 810 simulations with the initial set of parameters that produced high GFP concentrations (blue spaghetti plots) compared with the published deterministic model with the fixed set of parameters (red curve). Depending on the parameters used, GFP reached different concentration levels and the system became activated at different timepoints.

Profiles from the 651 simulations with the updated priors that produced high GFP concentrations (blue spaghetti plots) compared with the published deterministic model with the fixed set of parameters (red curve). The results seem converge towards the curve of the published model and all models have been activated in less than 45 hours, in contrast with the previous simulations where some of the models took over 55 hours to switch on.

Concentration profile of the median of the simulations with the initial set of parameters (dark blue line) between the Q1 and Q3 quartiles compared with the published deterministic model with the fixed set of parameters (red curve).

Concentration profile of the median of the simulations with the updated parameters (dark blue line) between the Q1 and Q3 quartiles compared with the published deterministic model with the fixed set of parameters (red curve). The median of the simulations is converging towards the published model’s curve.

Figure 4.10: Concentration profile of GFP/GFP\textsubscript{max} during the simulations under parameter uncertainty on the QS system. The system exhibited consistent behaviour with the publication and seemed to become activated under the different conditions tested. GFP\textsubscript{max} was set to be equal to the one maximum reported in the publication, in order to allow consistent comparison with the ensemble.

In order to test if any of the parameters are significantly different from their defined priors in the output of the simulations, the Kolmogorov–Smirnov test (K–S test)\textsuperscript{262} was performed between the parameters of the 85 models that showed bistability and the
initially sampled 1000 parameters. The ratio of the diffusion rate (D) over the doubling time (τ) was also tested, in order to investigate potential compensation between those two parameters. In order to account for the multiple testing problem, the significance level was set to 0.0024 (5% divided by the number of tests (21)), corresponding to a strict Bonferroni correction\textsuperscript{[263]}. The results (Table 4.8) showed that for the parameters $K_{d_{lux}}$ (dissociation constant of the LuxR-A complex to the lux promoter), D (diffusion rate) and τ (doubling time) we can reject the null hypothesis. The same is the case for the ratio of D over τ.
Table 4.8: K–S Test results for the 85 models of the QS system that exhibited bistability

<table>
<thead>
<tr>
<th>Parameter</th>
<th>K–S Test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d1}$</td>
<td>0</td>
<td>0.1317</td>
</tr>
<tr>
<td>$k_1$</td>
<td>0</td>
<td>0.4629</td>
</tr>
<tr>
<td>$K_{d2}$</td>
<td>0</td>
<td>0.0174</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0</td>
<td>0.7478</td>
</tr>
<tr>
<td>$k_A$</td>
<td>0</td>
<td>0.2714</td>
</tr>
<tr>
<td>$K_{dlux}$</td>
<td>1</td>
<td>0.00229</td>
</tr>
<tr>
<td>$k_{lux}$</td>
<td>0</td>
<td>0.4931</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>0.221</td>
</tr>
<tr>
<td>$a_R$</td>
<td>0</td>
<td>0.046</td>
</tr>
<tr>
<td>$a_I$</td>
<td>0</td>
<td>0.927</td>
</tr>
<tr>
<td>$d_A$</td>
<td>0</td>
<td>0.239</td>
</tr>
<tr>
<td>$d_{C2}$</td>
<td>0</td>
<td>0.85</td>
</tr>
<tr>
<td>$d_C$</td>
<td>0</td>
<td>0.7429</td>
</tr>
<tr>
<td>$d_R$</td>
<td>0</td>
<td>0.033</td>
</tr>
<tr>
<td>$d_I$</td>
<td>0</td>
<td>0.059</td>
</tr>
<tr>
<td>$d_{mR}$</td>
<td>0</td>
<td>0.321</td>
</tr>
<tr>
<td>$d_{mI}$</td>
<td>0</td>
<td>0.0746</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>2.48 · 10$^{-8}$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>1</td>
<td>2.34 · 10$^{-9}$</td>
</tr>
<tr>
<td>$d_{GF P}$</td>
<td>0</td>
<td>0.4627</td>
</tr>
<tr>
<td>Ratio D/T</td>
<td>1</td>
<td>3.07 · 10$^{-12}$</td>
</tr>
</tbody>
</table>

Figure 4.11 illustrates the difference in the expected distribution of the parameter values and the actual parameter values of the 85 models with bistable behaviour. The model seemed to require larger values for $K_{dlux}$ in order to exhibit bistable behaviour; however in the cases of D, $\tau$ and their ratio values from both extreme ends of the distribution seem to equally induce bistability, while the intermediate values are incompatible with bistable behaviour of the system.
By taking into account the K–S test results, the prior distributions for the parameters that had the most significant effects in the system (D and T) were updated in order to focus on the regions where the bistable results were observed. All the simulations were then repeated with the exact same values for the remaining parameters, in order to test how the system’s behaviour would be affected. This time, 349 models reached their maximum at very low values (ratio $\text{GFP}/\text{GFP}_{\text{max}}$ was below 0.1). The results for the remaining 651 models (Figures 4.10b and 4.10d) showed that the maximum concentration of GFP tends to converge towards the reported one in the publication. The system also seems to become activated faster than in the first simulation, with most of the models “jumping” in less than 45 hours, in contrast with the >55 hours observed in the previous simulations. The number of models that showed bistability was also increased, with 270 simulations showing a clear bistable region. The width of bistability region also became much more diverse compared with the initial simulations (Figures 4.12b, 4.12d and 4.12f).
Chapter 4. Modelling the $\gamma$-butyrolactone network in Streptomyces coelicolor

(a) Comparison of the initial distribution for the parameter $K_{dlux}$ with the actual values that showed bistability. It seems that the bistable behaviour is associated with the large values (>150 nM) for this parameter.

(b) Comparison of the initial distribution for the diffusion rate (D) with the actual values that showed bistability. Surprisingly, the system shows bistable behaviour for both very slow and very fast diffusion rates, but not for the intermediate values.

(c) Comparison of the initial distribution for the doubling time ($\tau$) with the actual values that showed bistability. Again, the system shows bistable behaviour for both small and very large doubling times, but not for the values in between.

(d) Comparison of the expected values for the ratio of D over $\tau$ and the actual ratios calculated from the bistable parameters.

Figure 4.11: Comparison between the expected parameter values according to the defined priors and the actual parameters that belonged to the models that exhibited bistability. The purple bar plots represent the log-ratio of the actual parameter values over the expected ones and the light blue lines in the background shows the original distribution from which parameter values were sampled for the ensemble model.
Chapter 4. Modelling the $\gamma$-butyrolactone network in Streptomyces coelicolor

(a) Simulations with initial set of parameters

(b) Simulations with updated priors for the doubling time ($\tau$) and the diffusion rate (D)

(c) Simulations with initial set of parameters

(d) Simulations with updated priors for the doubling time ($\tau$) and the diffusion rate (D)

(e) Simulations with initial set of parameters

(f) Simulations with updated priors for the doubling time ($\tau$) and the diffusion rate (D)

Figure 4.12: Variance of the bistability zone under parameter uncertainty. The steady state response of the system under different parameter sets, revealed a difference in the bistable region. Figures 4.12a and 4.12b show the number of simulations that were “switched on” at different concentrations of the external autoinducer $C_{A^*}$. Figures 4.12c and 4.12d show the number of simulations that were “switched off” at different concentrations of the external autoinducer $C_{A^*}$. Figures 4.12e and 4.12f show the width of the bistability zone of the simulations with different parameters. Upon updating our beliefs about the system’s priors, the number of simulations with bistable behaviour increased. The same happened with the diversity of the $C_{A^*}$ levels required for switching the system “on” and “off”, as well as the bistability zone width.
4.4.2 Results from the improved model of the GBL network

The analysis of the improved model was based on comparing the simulation results with transcriptomics data reported in the publication by Nieselt et al.\textsuperscript{[261]} Since the transcriptomics data and the simulations results are reported in different (and, in the case of the transcriptomics results, arbitrary) units and in order to compare their qualitative behaviour, the values for both were normalised as per: \( X_{i,\text{norm}} = \frac{X_i - \bar{X}}{SD} \)

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{transcriptomics_data}
\caption{Transcriptomics data for \textit{scbR} and \textit{scbA} genes as reported in the publication by Nieselt \textit{et al.}\textsuperscript{[261]} as normalized values of the corresponding microarray probe set for each time point. These data provided the criteria that were used as a measure of quality of the models.}
\end{figure}

The similarity between behaviour of the \textit{scbR} and \textit{scbA} simulations and the transcriptomics data was determined by comparing five different aspects of their behaviours:

1. The timepoint where the maximum concentration is reached
2. The relative gain (distance between the initial concentration and the peak)
3. The slope of increase (how fast the concentration reaches the peak)
4. The relative loss (distance between the peak and the subsequent minimum concentration)

5. The slope of decrease (how fast the concentration declines after the peak)

By calculating the values resulting from each of these criteria and their corresponding uncertainty (Table ??), probability distributions were generated which expressed our knowledge on the behaviours of these two species. These distributions were then used to compute the log-likelihood of each model for each one of the five criteria and to rank it based on how close it matched the experimental data, as per:

$$LL_i = \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right),$$

where $LL_i$ is the log-likelihood for one of the five criteria, $\mu$ is the maximum likelihood value for this criterion, $\sigma$ is the standard deviation expressing the uncertainty around the value and $x$ is the simulation results under comparison. Once the log-likelihood was determined for each of the criteria, the sum of the log-likelihoods was calculated for each species ($SLL_R$ and $SLL_A$). The total likelihood $TLL$ for each model was derived by adding the total likelihoods of the two species. The highest log likelihood profiles ($> -140$) are depicted in Figure 4.14. The details for the quantification of each comparative criterion and the complete likelihood profiles can be found in Appendix D.
Chapter 4. Modelling the γ-butyrolactone network in Streptomyces coelicolor

(a) scbR log-likelihood profiles for the 4 scenarios of the GBL model. Scenarios B and D have a clear prevalence, especially in the highest scoring area (log-likelihoods of -40 to -20).

(b) scbA log-likelihood profiles for the 4 scenarios of the GBL model. All scenarios follow a similar trend, however there is a higher number of models for scenarios B and D in the highest ranks (log-likelihoods of -40 to -20).

(c) Total log-likelihood profiles for the 4 scenarios of the GBL model. Similarly to the individual species’ log-likelihoods, scenarios B and D have the largest number of models whose behaviour approximates that of the experimental data.

Figure 4.14: Log-likelihood profiles (scbA, scbR and total) for the 4 scenarios of the GBL model. Scenarios B (antisense RNA interaction) and D (combination of all scenarios) show a clear predominance over the other two.

The log-likelihood analysis showed that the scenarios with the largest amount of models whose behaviour matches the experimental data were B (Transcriptional interference and antisense RNA interaction) and D (all three regulatory mechanisms combined) in which 116 models achieved a TLL > −140. However, the RA complex scenario (C) combined with the transcriptional interference had a very low score and its addition to the other two scenarios did not seem to affect the simulations’ outcome.
Similarly, transcriptional interference alone (scenario A) did not manage to achieve a large number of good matches with the experimental data (Figure 4.15). It therefore seems that the mechanism that is most influential for the model’s behaviour is the antisense RNA interaction.

![Figure 4.15: Summary of log-likelihood analysis of the improved GBL model for the 4 scenarios. Scenarios B and D achieve the same amount of optimal results, showing that scenario C does not seem to have any serious impact on the model.](image)

The plot of the models that achieved the highest log-likelihood score against the experimental data (Figure 4.16) shows that the defined criteria were successful in capturing the features of interest in the two species. Furthermore, the models that best match the experimental data, seem to be scenarios B and D. The fact that the curves for these two scenarios completely overlap, reinforces our belief that the RA complex and the transcriptional interference individually have a minimal effect in the model’s behaviour once they are acting alongside the antisense RNA mechanism. Another interesting point is that none of the models was able to explain the difference in the width of the scbR and scbA peaks; this could be due to some additional regulatory effect that is not considered in our models.
Comparison of scbR results from the best fit model of each scenario with the experimental data. The models managed to describe well the behaviour of scbR, with the scenarios B and D being the best matches.

Comparison of scbA results from the best fit model of each scenario with the experimental data. Again, scenarios B and D show the closest fit, however all models seem unable to explain why the two peaks are so different in width.

Figure 4.16: Comparison of the highest ranked models of the 4 scenarios with the transcriptomics data. The models capture the general behaviour of the system and scenarios B and D overlap completely, which reinforces our belief that scenario C does not really affect the model’s behaviour.

In order to investigate if any parameters are significantly affecting the model behaviour, the Kolmogorov–Smirnov test (K–S test) was performed on the models from each scenario that had a $SLL > -140$. The family-wise error rate was controlled by
was controlled by dividing the significance level by the number of tests performed (i.e., 44) to achieve a strict Bonferroni correction for multiple testing. The parameters for which we rejected the null hypothesis are summarised in Table 4.9. The complete K–S analysis results are included in Appendix D.

The parameters that seem to be most important are the degradation rate of ScbA protein and the degradation rate of $C_2 \cdot R_2$ (Figure 4.19). Those two parameters seemed to work on opposite directions in the cases where the highest ranked models were achieved. A faster ScbA degradation combined with a slower $C_2 \cdot R_2$ complex degradation appear to consistently produce models whose behaviour more closely matches the experimental data. In fact, further analysis of the ratio of these parameters, revealed that there is potentially compensation between them (K–S test rejects the null hypothesis with a p-value of $2.78 \cdot 10^{-33}$), as the highest ratios seem to be encountered in the highest ranked models (Figure 4.19e).

The two additional groups of simulations with the different promoter strengths also provided interesting results. The simulations from the first group ($k_{FR} < k_{FA}$) showed that the outputs of the scenarios A and C were not majorly affected by the change in the promoter strength. However, the results for scenarios B and D plummeted from 116 to 23 models that achieved a log-likelihood score above $140$ and from 875 models that had a measurable likelihood to 280 (Figures 4.17 and 4.18).

The K–S analysis showed that apart from the degradation of ScbA protein, only the heterogeneity factor significantly affected the behaviour of the models. The parameters of the optimal result models seem to cluster in the larger value region of $\chi$ (Figure 4.19c), meaning that the highest rankings in this group were achieved when the difference between the strengths of the two promoters was minimised.
Chapter 4. Modelling the \(\gamma\)-butyrolactone network in Streptomyces coelicolor

(a) Number of models that achieved high total log likelihoods (\(SLL > -140\))

(b) Number of models that achieved any quantifiable total log likelihood (\(SLL \neq NaN, SLL \neq \pm inf\))

Figure 4.17: Effect of promoter strength on the improved GBL model for the four scenarios. The blue bars represent the initial simulations with promoters of equal strength. The amount of successful models decreases when \(scbA\) promoter is stronger (red bars) and increases when \(scbR\) promoter is stronger (green bars). Scenarios A and C do not seem to become majorly affected.
Chapter 4. Modelling the γ-butyrolactone network in Streptomyces coelicolor

(a) $scbR$ log-likelihood profiles for scenario B of the GBL model with different strengths of promoters. A stronger $scbR$ promoter seems to significantly improve the behaviour of $scbA$.

(b) $scbA$ log-likelihood profiles for scenario B of the GBL model with different strengths of promoters. The simulations from the second group ($k_{FR} > k_{FA}$) followed a similar pattern, with the number of successful models ($SLL > -140$) from scenarios A and C not changing much. On the contrary, the amount of models that achieved a higher log-likelihood was increased for scenarios B and D (Figures 4.17 and 4.18). The K–S analysis revealed once again the degradation of ScbA protein, the degradation rate of $C_2 \cdot R_2$ and the heterogeneity factor consistently being significant for the behaviour of the models.
## Table 4.9: Summary of K–S test results for all scenarios and for different promoter strength combinations

$k_{FR} = k_{FA}$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>TI</th>
<th>TI-AR</th>
<th>TI-AS</th>
<th>TI-AR-AS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_A$</td>
<td>1</td>
<td>0.00021</td>
<td>1</td>
<td>5.91·10^{-7}</td>
</tr>
<tr>
<td>$d_{CR}$</td>
<td>1</td>
<td>0.00035</td>
<td>1</td>
<td>6.83·10^{-6}</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>8.64·10^{-5}</td>
<td>0</td>
<td>0.0087</td>
</tr>
</tbody>
</table>

$k_{FR} > k_{FA}$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>TI</th>
<th>TI-AR</th>
<th>TI-AS</th>
<th>TI-AR-AS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_A$</td>
<td>1</td>
<td>6.86·10^{-5}</td>
<td>1</td>
<td>0.00022</td>
</tr>
<tr>
<td>$d_{CR}$</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0.0030</td>
</tr>
<tr>
<td>$\chi$</td>
<td>1</td>
<td>2.50·10^{-7}</td>
<td>1</td>
<td>2.77·10^{-7}</td>
</tr>
</tbody>
</table>

$k_{FR} < k_{FA}$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>TI</th>
<th>TI-AR</th>
<th>TI-AS</th>
<th>TI-AR-AS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_A$</td>
<td>1</td>
<td>0.00017</td>
<td>1</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\chi$</td>
<td>1</td>
<td>1.27·10^{-14}</td>
<td>1</td>
<td>2.14·10^{-16}</td>
</tr>
</tbody>
</table>
(a) Comparison of the initial distribution for the degradation rate of ScbA protein ($d_A$) with the actual values of the highly ranked models. It seems that fastest degradation improves the behaviour of the models.

(b) Comparison of the initial distribution for the degradation rate of $C_2 \cdot R_2$ ($d_{CR}$) with the actual values of the highly ranked models. In this case, slower degradation of the complex leads to better results.

(c) Comparison of the initial distribution of the heterogeneity factor $\chi$ with the values of the highly ranked models, in the case where scbA promoter is stronger. The analysis shows that the highest ranked models were achieved with the largest values of $\chi$, meaning the cases where the difference between the firing rates of the two promoters was approaching the minimum.

(d) Comparison of the initial distribution of the heterogeneity factor $\chi$ with the values of the highly ranked models, in the case where scbR promoter is stronger. In this case, most of the $\chi$ producing good models were in the area of 2.3–3.2.

(e) Comparison of the expected distribution of the ratio of $d_A$ over $d_{CR}$ with the values of the highly ranked models. It appears that these two parameters could be compensating for each other, as highest ratios seem to be encountered in the best models.

**Figure 4.19:** Comparison between the expected parameter values according to the defined priors and the actual parameters that belonged to the models that had a $TLL > -140$. The purple bar plots represent the log-ratio of the actual parameter values over the expected ones and the light blue lines in the background show the original distribution.
4.5 Discussion

The recreation and analysis of the published QS model provided an extremely useful insight into the characteristics and behaviour of this system. Being able to understand its functionality and the output it yields, aids the comparison with the structurally similar GBL system. Additionally, the framework of analysis employed in the publication by Weber et al. formed the basis of the design of the improved GBL model. The QS model was quite stable and robust under varying parameters which allowed the exploration of the system’s inherent characteristics, such as the bistability. Furthermore, the importance of specific parameters (i.e., the diffusion rate and the doubling time) for the manifestation of these traits was brought into the spotlight. The update of the priors and the repetition of the simulations revealed that the bistability is enhanced as the parameter values converge towards the ones reported in the publication. Another intriguing outcome was the compensation between the parameters D and T and the fact that both extreme ends of their sampling distributions enhanced the bistable behaviour, but not the values in between them. Overall, the application of ensemble modelling on this system revealed areas where model behaviour is critically dependent on a tight selection of parameter values and demonstrated how to successfully update our beliefs about plausible parameter ranges based on comparison between experimental results and model simulations, resulting in a more informative final model.

On the other hand, the recreation and careful scrutiny of the published GBL models revealed their fragilities and the potential points for improvement. Both previously published models of the GBL circuit were focusing mostly on testing different parameter values and detecting the optimal combination for bistable behaviour. However, upon careful research into the literature (Appendix D), many of the optimal parameter values of these models may be biologically implausible or even infeasible. For instance, the degradation rates of the ScbR protein for which bistable behaviour was observed in the two published models by Mehra et al. and Chatterjee et al. were 0.132 – 0.402 min\(^{-1}\) and 5.4 – 21.6 min\(^{-1}\) respectively. This is in striking contrast to the information found in the literature which suggests that these proteins are quite stable and degrade in a very slow rate (5\(\times\)10\(^{-4}\) – 0.00578 min\(^{-1}\)). This creates doubt on whether the outcomes of these models are realistic or merely a successful proof-of-concept application under
Another point of interest is that the ensemble modelling performed on the published model by using the range of parameter values that the publication reported, could not replicate the published results. For the majority of the parameter sets the model crashed, and only in the very tight region of bistability the simulations were able to be completed. Even then, however, the average behaviour of the model’s species was quite far from the single parameter set simulation. This suggests that the optimal parameter model behaviour could be an outlier and not representative of the behaviour expected for this circuit topology. On top of all that, Mehra et al. reported that the behaviour of \textit{scbR} did not follow the experimental data. The improved GBL model seems to also support this finding, as the RA complex scenario had a very low score and did not seem to contribute at all towards matching the model’s behaviour with that of the transcriptomics data.

The simulation outputs of the GBL model revealed a system behaviour very different from the QS circuit. While QS produced consistent and robust results, the GBL model was very unpredictable and different sets of parameters could lead to completely different behaviours (e.g. a peak and decline or a smooth increase). This is one of the reasons why a large number of models achieved very low to non-calculable likelihood scores.

The investigation of the different scenarios and the model analysis elucidated some important features of the system and revealed behaviours that cannot be explained by any combination of plausible parameter values. The scenario with the existence of the putative ScbR–ScbA complex formation is rejected, as it only adds to the complexity of the model without actually contributing to its quality. Similarly, transcriptional interference by itself does not sufficiently explain the experimental results. However, when it is combined with the antisense RNA mechanism the number of successful models is approximately 5 times higher, suggesting that this mechanism is critical for the behaviour of the system. The promoter strength seems to also contribute to the mechanism, with the \textit{scbR} promoter seemingly being 4–9 times more aggressive in the most successful models than its \textit{scbA} counterpart.

The parameter analysis revealed that in contrast to the QS system, diffusion does not seem to majorly affect the GBL model. On the other hand, the importance of the
degradation of \textit{scbA} protein and of $C_2 \cdot R_2$ complex seemed to be a recurring issue in all groups of simulations, with $d_{CR}$ having small values and $d_A$ having large values in the optimal results. These two parameters seem to also compensate for each other as their high ratio seemingly improves the model behaviour. The heterogeneity factor $\chi$ is also a highly influential parameter for the model output. The analysis of the two different groups of simulations revealed that assigning higher strength to the \textit{scbA} promoter majorly impairs the quality of the results, with the highest rankings being achieved when the difference between the aggressiveness of the two promoters being minimised. On the contrary, the amount of high-likelihood-score models seems to increase when the \textit{scbR} promoter is stronger.

These findings suggest that the GBL system behaviour might not stem from an internal self regulation but from a growth-dependent response of the system to its external environment. If the diffusion of the SCBs is not an important factor for the model behaviour, it might be possible that bacterial communication is not actually involved and there is no coordination within the colony to trigger the antibiotic production, but the transition is performed by the cells individually once they reach their stationary phase. Furthermore, the fact that the model could not explain the difference in the width and maximum height of the peaks of \textit{scbR} and \textit{scbA}, suggests that there could be another agent involved in the system that is associated with regulating the \textit{scbA} gene and the ScbA protein. Additional experiments and simulations will need to be performed in order to fully clarify the role of the two genes and their interactions. An option to test the existence and significance of the antisense RNA and transcriptional interference mechanisms would be to conduct a series of experiments using synthetic genetic circuits with the \textit{scbR} and \textit{scbA} genes uncoupled and coupled, and with either both of the genes or only one of them being present each time. Finally, the existence of bistability in the system needs to be further investigated, as the evidence so far is inconclusive.

Based on the results obtained and on the apparent significance of growth for the system’s behaviour, an interesting potential expansion of the model would include the incorporation of uncertainty in the time dependent growth curve parameters. This could be achieved by introducing a lag phase with variable length in combination with a sampled $\mu_{\text{max}}$ (both sampled from log-normal distributions). The consistency of the final
growth curve so that it follows the pattern of the experimental data could be ensured by a multivariate distribution that would use the two priors as marginal distributions.

Furthermore, diffusion was modelled as an instantaneous reaction under the assumption that the concentration of the SCBs is homogeneous and that potential spatiotemporal patterns can be ignored for relatively small cell colonies and for long time scales such as the ones used in the simulations. However, in systems where the signalling molecules need to cover long distances, spatial effects are indeed relevant. It would, therefore, be interesting to model the diffusion mechanism as a spatial mass transfer following Fick’s first law \( J = -D \frac{d\phi}{dx} \) which would describe the \( \gamma \)-butyrolactone flux from a region of high concentration (cell) to a region of low concentration (environment) across a concentration gradient. This approach could be expanded to include the effect of diffusion on the concentration change over time (Fick’s second law: \( \frac{\partial \phi}{\partial t} = D \frac{\partial^2 \phi}{\partial x^2} \)) or a potential time dependency of the diffusion (e.g. a time delay between the production of the signal and its arrival to the receiving cells). In this way the role of diffusion in the system could be further explored and some of the intriguing issues regarding the function of SCBs might be clarified.

Unquestionably, the availability of more experimental data would also greatly assist in the further validation and improvement of the model. Additional quantitative transcriptomics results for \( scbA \) and \( scbR \) genes could validate the difference in the width of the two peaks, and more precise measurements on the degradation rates of the ScbA and ScbR proteins would help to fine-tune the probability distributions for these parameters and discern the biological plausibility of the previously published models. Finally, quantitative proteomics results from an experiment where cells do not produce \( \gamma \)-butyrolactones but are added externally in different concentrations would also be of interest, as it would assist with the validation of the model in a protein level additional to the mRNA level.

The improved GBL model clarified some aspects of the system but also raised some interesting questions. However, most importantly, it became clear that the model can now be used as a versatile and adaptable tool which will challenge and refine our understanding of the proposed functioning of this system, and perhaps even suggest a different biological role than originally envisaged. The developed framework of analysis with the
explicit consideration and documentation of uncertainty will now form the basis for a further extension of the model using alternative topologies and will allow us to quantify our posterior belief about the model’s parameters in the face of new experimental data. Finally, the model indicates key experiments which could completely elucidate the role of the system and the interactions of its components and potentially lead to the design of robust and sensitive systems with significant applications as orthologous regulatory circuits in synthetic biology and biotechnology.
Blank page
Chapter 5

Summary and future perspectives

Computational models are an extremely valuable tool for biological research, generating hypotheses that can be experimentally tested and validated. As the research focus in biology is turning to more complex projects, such as genome-scale modelling and engineering of microorganisms, the need for high-quality models with reliable predictions is also increasing. A prime target for the development of genetic engineering strategies are gene regulatory circuits such as the $\gamma$-butyrolactone system in *Streptomyces coelicolor*. These bacteria regulate their antibiotic production via a two-gene system whose exact mechanism of action is not fully elucidated.

The dual objective of this thesis was to contribute to the development of computational tools that will facilitate enhance the use of models in biology within an ensemble modelling framework, and at the same time apply these tools for the exploration of the GBL regulatory system.

Chapter 1 contained a review on the current status of modelling techniques, as well as the challenges that traditionally used methods, such as parameter fitting, are facing. The background for the mathematical methods that were used in this work was also introduced. In Chapter 2, the principles underlying the framework of analysis employed during this work were presented. This chapter discussed the benefits of explicit consideration of uncertainty from the early stages of model building and suggested adapting a more “respectful” attitude towards our models. This will ultimately lead to higher quality of models and promote collaborations between modellers and experimentalists.

In Chapter 3, the thesis expanded on these principles by putting forward a protocol that
addresses the challenge of defining informative priors for ensemble modelling. The protocol assesses the plausibility of parameter values retrieved from various sources and generates probability distributions, while ensuring the thermodynamic consistency of the model. Finally, in Chapter 4, a novel, adaptable and versatile model of the GBL circuit was designed by employing the methods discussed and developed in the previous chapters. This model aimed at testing the alternative mechanisms proposed for the GBL system in various combinations (transcriptional interference, antisense RNA interactions and formation of a putative complex between ScbR and ScbA proteins) and thus attempt to explain the system’s behaviour. The model development and analysis was also aided by the recreation and comprehensive refactoring of the previously published model on the GBL system by Mehra et al.\cite{mehra2006}, as well as a model on the quorum sensing system by Weber et al.\cite{weber2005} The results lead to the rejection of the relevance of the putative complex, but also suggests that transcriptional interference alone is not sufficient to explain the experimental results. The driving force behind the observed behaviour seems to be Antisense RNA interactions, combined with an aggressive $scbR$ promoter.

Whilst having produced very interesting results, the two key areas of focus during the thesis (the ensemble modelling protocol and the improved GBL model) also pave the way for future significant improvements (Figure 5.2).

The protocol presented in Chapter 3 of the thesis majorly automates and facilitates the definition of log-normal probability distributions which can be used as informative priors in an ensemble modelling framework. The next step is to make the protocol even more accessible and user-friendly, by creating a graphical user interface (GUI) which will allow the user to easily design probability distributions without a lot of computational effort. The user will be able to input the parameter values and assign weights by selecting options from a menu, as well as choose potential thermodynamic links between interdependent parameters. Additionally, the GUI will be able to retrieve data from the BRENDA database, to enrich parameter distributions or even define them in cases where no literature data can be found. A prototype which has been created for this purpose and which will be the base for future developments is presented in Figure 5.1.
Chapter 5. Summary and future perspectives

Figure 5.1: GUI prototype for the design of informative probability distributions. Features such as the weights and the thermodynamic links between parameters will be easily chosen from lists.

Additionally, the protocol part on imposition of thermodynamic constraints will be expanded to cover more complex systems (e.g., enzymatic reactions involving different numbers of substrates and products, enzyme inhibitions and more complex futile cycles). Finally, an interesting addition to the protocol would be the option to statistically analyse enzyme databases to infer the information content of related data (e.g., for related species) with respect to that of a target enzyme, and thus enable a rational parametrization of the weights through statistical learning.\textsuperscript{265} This could ultimately lead to a development of a fully Bayesian weighing scheme consistent with prior information on enzyme kinetic parameters.
On the other hand, the improved GBL model clarified some aspects of the system, but also raised some interesting questions. The interactions between \( \text{scbR} \) and \( \text{scbA} \) promoters and mRNAs are not yet fully clarified, although the model suggests that transcriptional interference due to promoter overlap as well as antisense RNA interactions are taking place. A potential future work would be to conduct a series of experiments and simulations on the two genes, inside a heterologous host (e.g. \( \text{E. coli} \)) and in different arrangements. For instance, the two promoters could be coupled and uncoupled and with one of the two genes at the time being knocked out. This will lead to a further elucidation on the importance of both genes being present in order to obtain the observed results and on how they function outside their native host.

Additionally, it would be possible to experimentally test the strength of the two promoters using constructs where the promoters \( O_R \) and \( O_A \) are alternatively employed upstream of suitable reporter genes in order to test the aggressiveness of each promoter.

Furthermore, the potential existence of another agent that is associated with regulating the \( \text{scbA} \) gene and the ScbA protein needs to be investigated. Another scenario that has been previously suggested is the potential action of ScbR as both an inhibitor and an activator of the gene expression.\(^{[117]}\) This mechanism could be further explored in an expansion of the current model and compared with the influence of the existing mechanisms.

Finally, previous studies have shown that in a system involving a small number of molecules, such as a regulatory or signalling system, stochasticity (fluctuations in transcription and translation or randomness in the autoinducer diffusion from the cell to the environment) can have a significant impact on the switch induction.\(^{[266,267]}\) Therefore, the small-size GBL system provides a good opportunity for stochastic modelling, in order to study the sensitivity of this system to internal or external fluctuations. Furthermore, the stochastic analysis could reveal more information on the type of communication (if any) that takes place within a \( \text{Streptomyces} \) colony. The stochastic model developed by Weber et al.\(^{[2]}\) on the quorum sensing system can therefore be employed to study the effects of noise in the GBL system. This model considers each bacterium as an individual cell carrying a copy of the GBL regulatory network. The resulting ensemble of all reactions in all cells is functioning as one global system. The cells are coupled
through the autoinducer diffusion reaction in a similar manner with the deterministic model: Each time a GBL molecule diffuses from out of one cell into the external environment it increases the number of molecules in the medium by one, thereby also increasing the probability of a GBL molecule to diffuse into another cell of the colony. The stochastic modelling should be able to represent the heterogeneity arising from intrinsic or extrinsic noise and thus achieve a more realistic description of key properties of the system, such as population-wide bet hedging.

![Diagram](image)

**Figure 5.2:** Future perspectives and contribution of the thesis. The methodological developments and the improved GBL model can promote the use of ensemble modelling, make models more accessible and guide future experimentation. Together they are a very useful addition to the synthetic biology toolbox.

It is therefore becoming obvious that the methodological developments described in this thesis can promote the use of ensemble modelling within the biological community and make models more accessible and adaptable. Furthermore, the improved GBL model which was designed by employing these methodologies, can be used to question and refine our understanding of the systems activity and guide future experimentation. Together, they can be extremely useful additions to the synthetic biology toolbox and
lead to the design of robust and sensitive systems with significant applications as orthologous regulatory circuits in synthetic biology and biotechnology.
Appendix A

Functions for the design of informative priors protocol

All the following scripts are written in MATLAB. The original MATLAB files are included as supplementary material in the publication.

A.1 Functions CalcModeSpread and weightedMedian

SCRIPT A.1: Function CalcModeSpread to calculate the Mode and Spread of the lognormal distribution from literature data

```
function [Mode, Spread] = CalcModeSpread(V,PredictInt)

% This function generates the mode and confidence interval ...
% factor for a lognormal
% distribution, given a matrix P which contains 4 columns. 1)The ...
% parameter values
% % retrieved from literature 2)the standard deviation of ...
% measurements in parameter values,
% % 3)the weights of the parameter values and 4) a definition of ...
% the SD as
% % multiplicative or additive. If the SD is additive (Value +/- ...
% SD) input 0
```
Appendix A. Functions for the design of informative priors protocol

8 \% and if it is multiplicative (Value */ $\text{div}\$ SD) input 1 in ...
9 \% the last column.
10 \% Sample format of the table:
11 \% V = [10000 NaN 2 0; 100 0 8 0; 0.01 5 10 0; 160 NaN 4 0; 1 0.1 4 ... 
12 \% 1];
13 \% The input can be given in two ways:
14 \% [Mode, Spread] = CalcModeSpread(V) \rightarrow only given table V.
15 \% In this case, the percentage is set to default value (0.6827).
16 \% [Mode, Spread] = CalcModeSpread(V, PredictInt) \rightarrow given table V ...
17 \% and percentage.
18
19 if nargin > 2
20 error('Too many inputs');
21 end
22
23 \% Assign default percentage if not given
24 if nargin == 1
25 PredictInt = 0.6827;
26 end
27
28 \% Find n for the inputted (or default) prediction interval
29 n = sqrt(2) * erfinv(PredictInt);
30
31 D = [ ];
32 W = [ ];
33
34 for i = 1:length(V(:, 1))
35 if V(i, 4) == 0 \% logtransform additive SD
36 lnE(i) = sqrt(log(1 + (V(i, 2).^2 ./ V(i, 1).^2)));  
37 \if isnan(V(i, 2)) \% if SD is NaN use 10\% multiplicative SD 
38 lnP(i) = log(V(i, 1)) - 1/2. * log(1.1)^2;
39 else 
40 lnP(i) = log(V(i, 1)) - 1/2. * lnE(i)^2;
41 end
42 else \% logtransform multiplicative SD
43 lnP(i) = log(V(i, 1));
44 lnE(i) = log(V(i, 2));
45 end
46
end

V(:,1)= lnP;
V(:,2)= lnE;

% Sort table from smallest to largest parameter value
A=sortrows(V,1);

% Split table columns into separate vectors
P=A(:,1);
E=A(:,2);
Wo=A(:,3);

if any(Wo < 0.0001)
    error('The weights cannot have values smaller than 0.0001.');
end

for i=1:length(P)
    if isnan(E(i)) %if SD is NaN assign default 10% SD
        mu=P(i);
        sigma=log(1.1);
        nbins = 1000; %generate bins within the parameter distribution ...
            range
        binEdges = linspace(mu-5*sigma,mu+5*sigma,nbins+1);
        aj = binEdges(1:end-1);
        aj=aj(:);
        bj = binEdges(2:end);
        bj=bj(:);
        cj = ( aj + bj ) ./ 2; %find centre of bins
        Pj = exp(-(cj-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
        Wj=Wo(i)*Pj.*(bj - aj); %calculate weight at the centre of bins
    elseif E(i)~0 %if SD is not NaN or 0 use logtransformed SD
        mu=P(i);
        sigma=E(i);
        nbins = 1000;
        binEdges = linspace(mu-5*sigma,mu+5*sigma,nbins+1);
Appendix A. Functions for the design of informative priors protocol

80  aj = binEdges(1:end-1);
81  aj=aj(:);
82  bj = binEdges(2:end);
83  bj=bj(:);
84  cj = ( aj + bj ) ./ 2;
85  Pj = exp(-(cj-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
86  Wj=Wo(i)*Pj.*(bj - aj);
87  else
88  cj=P(i); %if SD is 0 do not assign SD but keep only the single ...
     value
89  Wj=Wo(i);
90  end
91
92  %if the value is not the minimum, it's not a single value and it ...
     does not overlap
93  %with the previous value, generate additional bins between twice ...
     the distance of P(i) and
94  %P(i-1), otherwise do nothing
95  if P(i) ̸= min(P) && min(cj)>P(i-1) && length(cj) ̸= 1
96  nbins2 = 1000;
97  binEdges2 = ...
     linspace(min(cj)-2*abs(min(cj)-P(i-1)),min(cj),nbins2+1);
98  ajad = binEdges2(1:end-1);
99  ajad=ajad(:);
100  bjad = binEdges2(2:end);
101  bjad=bjad(:);
102  cjad = ( ajad + bjad ) ./ 2;
103  Pjad = exp(-(cjad-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
104  Wjad=Wo(i)*Pjad.*(bjad - ajad);
105  else
106  cjad=[];
107  Wjad=[];
108  end
109
110
% if the value is not the maximum, it's not a single value and it does not overlap
% with the next value, generate additional bins between twice the distance of P(i) and P(i+1), otherwise do nothing
if P(i) ≠ max(P) && max(cj) < P(i+1) && length(cj) ≠ 1
nbins3 = 1000;
binEdges3 = ...
linspace(max(cj),max(cj)+2*abs(P(i+1)-max(cj)),nbins3+1);
ajad2 = binEdges3(1:end-1);
ajad2 = ajad2(:);
bjad2 = binEdges3(2:end);
bjad2 = bjad2(:);
cjad2 = (ajad2 + bjad2) ./ 2;
Pjad2 = exp(-(cjad2-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
Wjad2 = Wo(i)*Pjad2.*(bjad2 - ajad2);
else
cjad2 =[];
Wjad2 =[];
end
D = [D ; cj ; cjad ; cjad2]; % add the centres of all bins in a vector
W = [W ; Wj ; Wjad ; Wjad2]; % add the weights of all bin centres in a vector
end
wMed = weightedMedian(D,W); % calculate the weighted median of values in vector D
S = std(D,W); % calculate the weighted standard deviation of values in vector D
Mode=exp(wMed); % calculate Mode
Spread=exp(n*S); % calculate Spread
SCRIPT A.2: Function \texttt{weightedMedian} to calculate the weighted median of a vector of values according to their weights

```matlab
function wMed = weightedMedian(D,W)

% The function calculates the weighted median from a vector of ...
% values based
% on their weights. The input is a vector with the values and a ...
% vector with
% their corresponding weights.

if nargin \neq 2
    error('weightedMedian:wrongNumberOfArguments', ...
          'Wrong number of arguments.');
end

if size(D) \neq size(W)
    error('weightedMedian:wrongMatrixDimension', ...
          'The dimensions of the input-matrices must match.');
end

wMed = [];

% (line by line) transformation of the input-matrices to ...
% line-vectors

d = reshape(D',1,[]);
w = reshape(W',1,[]);

% sort the vectors
A = [d' w'];
ASortzeros = sortrows(A,1);
ASort=ASortzeros;
ASort(any(ASortzeros(:,2) \leq (10^-14)\:2),:) = [];
dSort = ASort(:,1)';
wSort = ASort(:,2)';

% If there is only one value, it is the weighted median
if length(dSort)== 1
    wMed=dSort;
```

Appendix A. Functions for the design of informative priors protocol 154
% If there are only two values the weighted median is their mean ...
if they
% have equal weights, otherwise the weighted median is the value ...
with the
% largest weight.
if length(dSort)== 2
if wSort(1)== wSort(2)
 wMed= (dSort(1)+dSort(2))/2;
elseif wSort(1)> wSort(2)
 wMed= dSort(1);
else wMed= dSort(2);
end
end

% If there are more than two values, find the value which is the ...
50% weighted percentile:
if (length(dSort)\neq 1) \&\& (length(dSort)\neq 2))
i = 1;
j = length(wSort);
Start=wSort(i);
End=wSort(j);

while i < j-1
if Start-End > 10^-14
End= End+wSort(j-1);
j=j-1;
else
Start= Start+wSort(i+1);
i=i+1;
end
end

% If the 50% weighted percentile falls between two values, the ...
weighted
% median is the average of the two values if their individual ... weights are
Appendix A. Functions for the design of informative priors protocol

```plaintext
% equal, otherwise the weighted median is the value with the...
  largest
% individual weight
if abs(Start-End) < (10^-14)
  wMed=(dSort(i)+dSort(j))/2;
elseif Start-End > (10^-13)
  wMed=dSort(i);
else
  wMed=dSort(j);
end
end
```

A.2 Function CreateLogNormDist

**Script A.3**: Function *CreateLogNormDist* to calculate the parameters $\mu$ and $\sigma$ of the log-normal distribution by using the Mode and Spread

```plaintext
function [mu, sigma, Xmin, Xmax] = CreateLogNormDist(Mode, ...
  Spread, Percentage)

% This function generates the location and scale parameters for ...
% a lognormal
% distribution, given a mode, a Spread and a percentage of ...
% values that will
% be contained within the defined boundaries. The input can be ...
% given in two ways:
% [mu, sigma, Xmin, Xmax] = CreateLogNormDist(Mode, Spread)-->
% only given Mode and
% Spread. In this case, the percentage is set to default value ...
% (0.6827).
% [mu, sigma, Xmin, Xmax] = CreateLogNormDist(Mode, Spread, ...
% Percentage)-->
% given
% Mode, Spread and percentage. The function then generates the ...
% location and
% scale parameters $\mu$ and $\sigma$ of the lognormal distribution, ...
```
% the Xmin(=Mode/Spread) and Xmax(=Mode+Spread).

% Error checking
if nargin < 2
    error('Must have at least 2 input arguments');
end

if nargin > 3
    error('Too many inputs');
end

% Assign default percentage if not given
if nargin == 2
    Percentage=0.6827;
end

% Calculate Xmin and Xmax of the range of values from the Mode ...
    and Spread
Xmin=Mode/Spread;
Xmax=Mode*Spread;

% Calculate mu and sigma of the lognormal distribution
syms s
eqn = 1/2+1/2*erf((log(Xmax)-(log(Mode)+s^2))/(sqrt(2)*s))
    -(1/2+1/2*erf((log(Xmin)-(log(Mode)+s^2))/(sqrt(2)*s)))
    ==Percentage;
sigma = vpasolve(eqn,s);
mu=log(Mode)+sigma^2;

% Plot final distribution
mu=double(mu);
sigma=double(sigma);
pd=makedist('Lognormal',mu,sigma);
Dist=random(pd,10000,1);
figure()
histfit(Dist,80,'lognormal');
h = findobj(gca,'Type','patch');
48 set(h,'FaceColor','c','facealpha',0.3);
49 htitle = title('Distribution for generated parameter values'); ...
    hxlabel=xlabel('Parameter values'); hxlabel=ylabel('Kernel ... Density');
50 set(gca,'FontSize',18) % font size of axis values
51 set(htitle,'FontSize',26) % font size of title
52 set(hxlabel,'FontSize',20) % font size of x axis
53 set(hylabel,'FontSize',20) % font size of y axis
54 end

A.3 Function Multivariate3param

SCRIPT A.4: Function Multivariate3param to generate a multivariate distribution for a three-parameter mass action reaction

1 function [KD1, kon1, koff1] = Multivariate3param(muKD, sigmaKD, ...
   mukon, sigmakon, mukoff, sigmakoff, ParamNo)
2 % The function generates a multivariate distribution for mass ... action reactions
3 % with 3 parameters. The user must provide the mu and sigma for ... the three
4 % distributions in the correct order, as well as the number of ... parameter
5 % samples that are needed (ParamNo).
6
7 %Define lognormal distributions:
8 % KD
9 pd1=makedist('Lognormal',muKD,sigmaKD);
10 KD=random(pd1,1000000,1);
11
12 % kon
13 pd2=makedist('Lognormal',mukon,sigmakon);
14 kon=random(pd2,1000000,1);
15
16 % koff
17 pd3=makedist('Lognormal',mukoff,sigmakoff);
Appendix A. Functions for the design of informative priors protocol

\[ k_{off} = \text{random(pd3,1000000,1)}; \]

% Calculate coefficients of variation:
\[ G_{CV1} = \exp(\sigma_{KD}) - 1; \]
\[ G_{CV2} = \exp(\sigma_{kon}) - 1; \]
\[ G_{CV3} = \exp(\sigma_{koff}) - 1; \]

% Choose the parameter with the largest coefficient of variation:
\[ A = [G_{CV1} \ G_{CV2} \ G_{CV3}]; \]
\[ M = \max(A); \]

% Set dependent parameter based on the largest GCV and
% calculate new \( \mu \) and \( \sigma \) for the dependent distribution:

\textbf{if} \( M == G_{CV1} \) % KD dependent
\[ K_{D} = k_{off}/k_{on}; \]
\[ \mu_{KD} = \mu_{koff} - \mu_{kon}; \]
\[ \sigma_{KD} = \sqrt{(\sigma_{koff}^2 + \sigma_{kon}^2)}; \]
\textbf{elseif} \( M == G_{CV2} \) % kon dependent
\[ k_{on} = k_{off}/K_{D}; \]
\[ \mu_{kon} = \mu_{koff} - \mu_{KD}; \]
\[ \sigma_{kon} = \sqrt{(\sigma_{koff}^2 + \sigma_{KD}^2)}; \]
\textbf{elseif} \( M == G_{CV3} \) % koff dependent
\[ k_{off} = k_{on} \times K_{D}; \]
\[ \mu_{koff} = \mu_{kon} + \mu_{KD}; \]
\[ \sigma_{koff} = \sqrt{(\sigma_{kon}^2 + \sigma_{KD}^2)}; \]
\textbf{end}

\textbf{if} \( M == G_{CV2} \) % kon dependent
% Generate multivariate distribution:
\[ B = [k_{on} \ k_{off}]; \]
CorrMat = \text{corrcoef}(B);
\[ \text{Mu} = [\mu_{kon} \ \mu_{koff}]; \]
\[ \text{Sigma} = [\sigma_{kon} \ \sigma_{koff}]; \]
% Force column vectors
Mu = Mu(:,);
Sigma = Sigma(:,);

% Calculate the covariance structure
sigma_down = repmat( Sigma', numel(Sigma), 1 );
sigma_acrs = repmat( Sigma, 1, numel(Sigma) );
covv = log( CorrMat .* sqrt(exp(sigma_down.^2)-1) .* ...
          sqrt(exp(sigma_acrs.^2)-1) + 1 );

% The Simulation
y = exp( mvnrnd( Mu, covv, ParamNo ));
kon1=y(:,1);
koff1=y(:,2);
KD1=koff1./kon1;

else % koff or KD dependent
    % Generate multivariate distribution:
    B=[KD koff];
    CorrMat=corrcoef(B);
    Mu = [muKD mukoff];
    Sigma = [sigmaKD sigmakoff];
    % Force column vectors
    Mu = Mu(:,);
    Sigma = Sigma(:,);
    % Calculate the covariance structure
    sigma_down = repmat( Sigma', numel(Sigma), 1 );
sigma_acrs = repmat( Sigma, 1, numel(Sigma) );
covv = log( CorrMat .* sqrt(exp(sigma_down.^2)-1) .* ...
          sqrt(exp(sigma_acrs.^2)-1) + 1 );
    % The Simulation
    y = exp( mvnrnd( Mu, covv, ParamNo ));
    KD1=y(:,1);
koff1=y(:,2);
    kon1=koff1./KD1;
end

A.4 Function Multivariate5param
Appendix A. Functions for the design of informative priors protocol

SCRIPT A.5: Function `Multivariate5param` to generate a multivariate distribution for a five-parameter Michaelis–Menten reaction

```matlab
function [Keqfin, Kafin, Kfpin, V1fin, V2fin] = ...
    Multivariate5param(muKeq, sigmaKeq, mukmf, sigmakmf, mukmr, ...
        sigmakmr, muvmaxf, sigmavmaxf, muvmaxr, sigmavmaxr, ParamNo, Eo)

% The function generates a multivariate distribution for ...
% reactions with
% Michaelis-Menten kinetics with 5 parameters. The user must ...
% provide the
% mu and sigma for the 5 distributions in the correct order, as ...
% well as the
% number of parameter samples that are needed (ParamNo).

% k1  k3
% E + A <-> X <-> P + E
% k2  k4
% KD1  KD2

% STAGE 1: Parameter consistency
% This stage finds the parameter with the largest geometric ...
% coefficient of
% variation and assigns it as dependent on the remaining 4 ...
% parameters. The
% distribution for the dependent parameter is calculated via the ...
% Haldane
% equation from the distributions of the 4 independent parameters.

% Calculate geometric coefficients of variation:
GCV1=exp(sigmaKeq)-1;
GCV2=exp(sigmakmf)-1;
GCV3=exp(sigmakmr)-1;
GCV4=exp(sigmavmaxf)-1;
GCV5=exp(sigmavmaxr)-1;

% Choose the parameter with the largest coefficient of variation:
A=[GCV1 GCV2 GCV3 GCV4 GCV5];
M=max(A);
```
% Set dependent parameter based on the largest GCV and calculate new mu and sigma for the dependent distribution:

if M==GCV2  % kmf dependent
    mukmf=(muvmaxf+mukmr)-(muKeq+muvmaxr);
    sigmakmf=sqrt(sigmavmaxf^2+sigmakmr^2+sigmaKeq^2+sigmavmaxr^2);
elseif M==GCV1  % Keq dependent
    muKeq=(muvmaxf+mukmr)-(mukmf+muvmaxr);
    sigmaKeq=sqrt(sigmavmaxf^2+sigmakmr^2+sigmakmf^2+sigmavmaxr^2);
elseif M==GCV3  % kmr dependent
    mukmr=(muKeq+muvmaxr+mukmf)-muvmaxf;
    sigmakmr=sqrt(sigmaKeq^2+sigmavmaxr^2+sigmakmf^2+sigmavmaxf^2);
elseif M==GCV4  % vmaxf dependent
    muvmaxf=(muKeq+muvmaxr+mukmf)-mukmr;
    sigmavmaxf=sqrt(log(((exp(2*muvmaxf+sigmavmaxf^2)\*\(\exp(\text{sigmavmaxf}^2)\-1)\+\exp(2*\text{muvmaxr}+\text{sigmavmaxr}^2)\*\(\exp(\text{sigmavmaxr}^2)\-1\)/((\exp(\text{muvmaxf}+(\text{sigmavmaxf}^2)/2)\+\exp(\text{muvmaxr}+(\text{sigmavmaxr}^2)/2))\^2\)+1)+\text{sigmakmf}^2));
    muk1=log(\exp(\text{muvmaxf}\+(\text{sigmavmaxf}^2)/2)+\exp(\text{muvmaxr}+(\text{sigmavmaxr}^2)/2))\-\log(\(\exp(2*\text{muvmaxf}+\text{sigmavmaxf}^2)\)*

% STAGE 1: Parameter correlation
% This stage calculates the correlation between the external michaelis-menten parameters, by using the internal mass action kinetics.

% Design mass action kinetics' distributions:
% k1=(V1+V2)/(K1*Eo);

% sigmak1=sqrt(log((exp(2*\text{muvmaxf}+\text{sigmavmaxf}^2)\*\(\exp(\text{sigmavmaxf}^2)\-1)\+\exp(2*\text{muvmaxr}+\text{sigmavmaxr}^2)\*\(\exp(\text{sigmavmaxr}^2)\-1\)/((\exp(\text{muvmaxf}+(\text{sigmavmaxf}^2)/2)\+\exp(\text{muvmaxr}+(\text{sigmavmaxr}^2)/2))\^2\)+1)+\text{sigmakmf}^2));

% muk1=log(\exp(\text{muvmaxf}\+(\text{sigmavmaxf}^2)/2)+\exp(\text{muvmaxr}+(\text{sigmavmaxr}^2)/2))\-\log(\(\exp(2*\text{muvmaxf}+\text{sigmavmaxf}^2)\)*

}
Appendix A. Functions for the design of informative priors protocol

66 \( (\exp(\text{sigmavmaxf}^2)-1)+\exp(2\times\text{muvmaxr}+\text{sigmavmaxr}^2)\times(\exp(\text{sigmavmaxr}^2)\times(\exp(\text{sigmavmaxr}^2)-1))/((\exp(\text{muvmaxf}+(\text{sigmavmaxf}^2)/2)+\exp(\text{muvmaxr}+(\text{sigmavmaxr}^2)/2)^2)+1)))/2-(\text{mukmf}+\log(\text{Eo})); \)

67

68 \% k2=V2./Eo;
69 muk2=muvmaxr-log(\text{Eo});
70 sigmak2=\text{sigmavmaxr};

71 \% k3=V1./Eo;
72 muk3=muvmaxf-log(\text{Eo});
73 sigmak3=\text{sigmavmaxf};

74 \% k4=(V1+V2)./(\text{KP}.*\text{Eo});
75 sigmak4=sqrt(log(((\exp(2\times\text{muvmaxf}+\text{sigmavmaxf}^2)\times(\exp(\text{sigmavmaxf}^2)-1)+\exp(2\times\text{muvmaxr}+\text{sigmavmaxr}^2)\times(\exp(\text{sigmavmaxr}^2)-1))/((\exp(\text{muvmaxf}+(\text{sigmavmaxf}^2)/2)+\exp(\text{muvmaxr}+(\text{sigmavmaxr}^2)/2)^2)+1))/2-(\text{mukmr}+\log(\text{Eo}));

76 muk4=log(\exp(\text{muvmaxf}+(\text{sigmavmaxf}^2)/2)+\exp(\text{muvmaxr}+(\text{sigmavmaxr}^2)/2)-\log(((\exp(2\times\text{muvmaxf}+\text{sigmavmaxf}^2)\times(\exp(\text{sigmavmaxf}^2)-1)+\exp(2\times\text{muvmaxr}+\text{sigmavmaxr}^2)\times(\exp(\text{sigmavmaxr}^2)-1))/((\exp(\text{muvmaxf}+(\text{sigmavmaxf}^2)/2)+\exp(\text{muvmaxr}+(\text{sigmavmaxr}^2)/2)^2)+1))/2-(\text{mukmr}+\log(\text{Eo}));

77 \% KD1=k2./k1;
78 muKD1=muk2-muk1;
79 sigmakD1=sqrt(sigmak2^2+sigmak1^2);

80 \% KD2=k4./k3;
81 muKD2=muk4-muk3;
82 sigmakD2=sqrt(sigmak4^2+sigmak3^2);

83 \% Generate random values from the distributions of KD1, KD2, k2 ... and k3:
84 pd6=makedist('Lognormal',muKD1,sigmakD1);
85 KD1=random(pd6,1000000,1);
86 pd7=makedist('Lognormal',muKD2,sigmakD2);
87 KD2=random(pd7,1000000,1);
% Calculate the values for k1 and k3 through KD1 & k2 and KD2 & ... k4 respectively:
    k1 = k2 ./ KD1;
    k4 = k3 .* KD2;

% Calculate the new Keq through the mass action kinetics' ...
    Keqnew = (k1 .* k3) ./ (k2 .* k4);

% Calculate the new Michaelis-Menten parameters:
    V1new = k3 .* Eo;
    V2new = k2 .* Eo;
    KAnew = (k2 + k3) ./ k1;
    KPnew = (k2 + k3) ./ k4;

% Generate multivariate distribution by using the covariance ...
    B = [Keqnew KAnew KPnew V1new V2new];
    CorrMat = corrcoef(B);
    MuB = [muKeq mukmf mukmr muvmaxf muvmaxr];
    SigmaB = [sigmaKeq sigmakmf sigmakmr sigmavmaxf sigmavmaxr];

% Force column vectors
    MuB = MuB(:);
    SigmaB = SigmaB(:);

% Calculate the covariance structure
    sigma_down = repmat(SigmaB', numel(SigmaB), 1);
    sigma_acrs = repmat(SigmaB, 1, numel(SigmaB));
    covv = log(CorrMat .* sqrt(exp(sigma_down.^2)-1) .* ...
               sqrt(exp(sigma_acrs.^2)-1) + 1);

% The Simulation
    x = exp(mvnrnd(MuB, covv, ParamNo));
    Keqfin = x(:,1);
    KAfin = x(:,2);
A.5 Function MultivariateBiBi

SCRIPT A.6: Function MultivariateBiBi to generate a multivariate distribution for a
Bi–Bi Michaelis–Menten reaction

```
function [Keqfin, kAfin, KBfin, KPfin, KQfin, V1fin, V2fin] = ...
    MultivariateBiBi(muKeq, sigmaKeq, mukA, sigmakA, mukB, ...
        sigmakB, mukP, sigmakP, mukQ, sigmakQ, muvmaxf, sigmavmaxf, ...
        muvmaxr, sigmavmaxr, ParamNo)
% The function generates a multivariate distribution for Bi–Bi...
% reactions with
% Michaelis–Menten kinetics. The user must provide the
% mu and sigma for the 7 distributions in the correct order, as ...
% well as the
% number of parameter samples that are needed (ParamNo).

% Define lognormal distributions:
% Keq
pd1=makedist('Lognormal',muKeq,sigmaKeq);
Keq=random(pd1,1000000,1);

% kA
pd2=makedist('Lognormal',mukA, sigmakA);
kA=random(pd2,1000000,1);

% kB
pd3=makedist('Lognormal',mukB, sigmakB);
kB=random(pd3,1000000,1);

% kP
pd4=makedist('Lognormal',mukP, sigmakP);
```
Appendix A. Functions for the design of informative priors protocol

22 \( k_P = \text{random}(pd4, 1000000, 1) \);
23
24 \% k_Q
25 pd5 = \text{makedist}('\text{Lognormal}', muk_Q, sigmak_Q);
26 k_Q = \text{random}(pd5, 1000000, 1);
27
28 \% V_1
29 pd6 = \text{makedist}('\text{Lognormal}', muvmax_f, sigmavmax_f);
30 V_1 = \text{random}(pd6, 1000000, 1);
31
32 \% V_2
33 pd7 = \text{makedist}('\text{Lognormal}', muvmax_r, sigmavmax_r);
34 V_2 = \text{random}(pd7, 1000000, 1);
35
36 \% Calculate geometric coefficients of variation:
37 GCV1 = \exp(sigmak_{\text{Keq}}) - 1;
38 GCV2 = \exp(sigmak_A) - 1;
39 GCV3 = \exp(sigmak_B) - 1;
40 GCV4 = \exp(sigmak_P) - 1;
41 GCV5 = \exp(sigmak_Q) - 1;
42 GCV6 = \exp(sigmavmax_f) - 1;
43 GCV7 = \exp(sigmavmax_r) - 1;
44
45 \% Choose the parameter with the largest coefficient of variation:
46 A = [GCV1 GCV2 GCV3 GCV4 GCV5 GCV6 GCV7];
47 M = \max(A);
48
49 \% Set dependent parameter based on the largest GCV and
50 \% calculate new mu and sigma for the dependent distribution:
51
52 if M == GCV2 \% kA dependent
53 k_A = (k_P * k_Q * (V_1.^2)) / (K_{\text{eq}} * k_B * (V_2.^2));
54 muk_A = (muk_P + muk_Q + 2 * muvmax_f) - (mu_{\text{Keq}} + muk_B + 2 * muvmax_r);
55 sigmak_A = \sqrt{\text{sigma}_{\text{Keq}}^2 + \text{sigma}_{\text{A}}^2 + \text{sigma}_{\text{P}}^2 + \text{sigma}_{\text{Q}}^2 + (2 * \text{sigma}_{\text{max}}_f)^2 + (2 * \text{sigma}_{\text{max}}_r)^2};
56
57 elseif M == GCV1 \% Keq dependent
58 K_{\text{eq}} = (k_P * k_Q * (V_1.^2)) / (k_A * k_B * (V_2.^2));
59
muKeq = (2 * muvmaxf + mukP + mukQ) - (mukA + mukB + 2 * muvmaxr);

sigmaKeq = sqrt(sigmakA^2 + sigmakB^2 + sigmakP^2 + sigmakQ^2 +
           (2 * sigmavmaxr)^2 + (2 * sigmavmaxf)^2);

elseif M == GCV3 % kB dependent
kB = (kP .* kQ .* (V1.^2))./(kA .* Keq .* (V2.^2));
mukB = (muKeq + mukA + mukB + 2 * muvmaxr) - (muKQ + 2 * muvmaxf);
sigmakB = sqrt(sigmakA^2 + sigmaKeq^2 + sigmakP^2 + sigmakQ^2 +
               (2 * sigmavmaxr)^2 + (2 * sigmavmaxf)^2);

elseif M == GCV4 % kP dependent
kP = (kA .* kB .* Keq .* (V2.^2))./(kQ .* (V1.^2));
mukP = (muKeq + mukA + mukB + 2 * muvmaxr) - (muKQ + 2 * muvmaxf);
sigmakP = sqrt(sigmakA^2 + sigmaKeq^2 + sigmakB^2 + sigmakQ^2 +
               (2 * sigmavmaxr)^2 + (2 * sigmavmaxf)^2);

elseif M == GCV5 % kQ dependent
kQ = (kA .* kB .* Keq .* (V2.^2))./(kP .* (V1.^2));
mukQ = (muKeq + mukA + mukB + 2 * muvmaxr) - (muKQ + 2 * muvmaxf);
sigmakQ = sqrt(sigmakA^2 + sigmaKeq^2 + sigmakB^2 + sigmakP^2 +
               (2 * sigmavmaxr)^2 + (2 * sigmavmaxf)^2);

elseif M == GCV6 % vmaxf dependent
V1 = sqrt((kA .* kB .* Keq .* (V2.^2))./(kQ .* kP));
muvmaxf = (0.5 * muKeq + 0.5 * mukA + 0.5 * mukB + muvmaxr) - (0.5 * mukP + 0.5 * mukQ);
sigmavmaxf = sqrt(sigmakA^2 + sigmaKeq^2 + sigmakB^2 + sigmakP^2 +
                   sigmakQ^2 + (2 * sigmavmaxr)^2);

elseif M == GCV6 % vmaxr dependent
V2 = sqrt((kP .* kQ .* (V1.^2))./(kA .* Keq .* kB));
muvmaxr = (0.5 * mukP + 0.5 * mukQ + muvmaxf) - (0.5 * muKeq + 0.5 * mukA + 0.5 * mukB);
sigmavmaxr = sqrt(sigmakA^2 + sigmaKeq^2 + sigmakB^2 + sigmakP^2 +
                   sigmakQ^2 + (2 * sigmavmaxf)^2);
end

if M == GCV1 % Keq dependent
  % Generate multivariate distribution:
  B = [Keq kA kB kP kQ V1];
Appendix A. Functions for the design of informative priors protocol

CorrMat = corrcoef(B);
Mu = [muKeq mukA mukB mukP mukQ muvmaxf];
Sigma = [sigmaKeq sigmakA sigmakB sigmakP sigmakQ sigmavmaxf];

% Force column vectors
Mu = Mu(:,);
Sigma = Sigma(:,);

% Calculate the covariance structure
sigma_down = repmat( Sigma', numel(Sigma), 1 );
sigma_acrs = repmat( Sigma, 1, numel(Sigma) );
covv = log( CorrMat .* sqrt(exp(sigma_down.^2)-1) .* ...
sqrt(exp(sigma_acrs.^2)-1) + 1 );

% The Simulation
y = exp( mvnrnd( Mu, covv, ParamNo ));

Keqfin = y(:,1);
KAfin = y(:,2);
KBfin = y(:,3);
KPfin = y(:,4);
KQfin = y(:,5);
V1fin = y(:,6);
V2fin = sqrt((KPfin.*KQfin.*(V1fin.^2))./(KAfin.*Keqfin.*KBfin));

else

B = [kA kB kP kQ V1 V2];
CorrMat = corrcoef(B);
Mu = [mukA mukB mukP mukQ muvmaxf muvmaxr];
Sigma = [sigmakA sigmakB sigmakP sigmakQ sigmavmaxf sigmavmaxr];

% Force column vectors
Mu = Mu(:,);
Sigma = Sigma(:,);

% Calculate the covariance structure
sigma_down = repmat( Sigma', numel(Sigma), 1 );
sigma_acrs = repmat( Sigma, 1, numel(Sigma) );
covv = log( CorrMat .* sqrt(exp(sigma_down.^2)-1) .* ...
sqrt(exp(sigma_acrs.^2)-1) + 1 );

% The Simulation
y = exp( mvnrnd(Mu, covv, ParamNo ));

KAfin = y(:,1);
Appendix A. Functions for the design of informative priors protocol

136 KBfin=y(:,2);
137 KPfin=y(:,3);
138 KQfin=y(:,4);
139 V1fin=y(:,5);
140 V2fin=y(:,6);
141 Keqfin=(KPfin.*KQfin.*(V1fin.^2))./(KAfin.*KBfin.*(V2fin.^2));
142
143 end

A.6 Function MultivariateUniBi

SCRIPT A.7: Function MultivariateUniBi to generate a multivariate distribution for a Uni–Bi Michaelis–Menten reaction

1 function [Keqfin, KAfin, KPfin, KiQfin, V1fin, V2fin] = ...
   MultivariateUniBi(muKeq, sigmaKeq, mukA, sigmakA, mukP, ...
   sigmakP, mukiQ, sigmakiQ, muvmaxf, sigmavmaxf, muvmaxr, ...
   sigmavmaxr, ParamNo)
2   % The function generates a multivariate distribution for Uni-Bi ...
3   % reactions with
4   % Michaelis–Menten kinetics. The user must provide the
5   % mu and sigma for the 6 distributions in the correct order, as ...
6   % well as the
7   % number of parameter samples that are needed (ParamNo).
8
9   % Define lognormal distributions:
10   % Keq
11   pd1=makedist('Lognormal',muKeq,sigmaKeq);
12   Keq=random(pd1,1000000,1);
13
14   % kA
15   pd2=makedist('Lognormal',mukA, sigmakA);
16   kA=random(pd2,1000000,1);
17
18   % kP
19   pd3=makedist('Lognormal',mukP, sigmakP);
Appendix A. Functions for the design of informative priors protocol

18 \( kP = \text{random}(pd3, 1000000, 1); \)

19

20 \%
21 pd4 = \text{makedist('Lognormal', mukiQ, sigmakiQ)};
22 kiQ = \text{random}(pd4, 1000000, 1);

23

24 \%
25 pd5 = \text{makedist('Lognormal', muvmaxf, sigmavmaxf)};
26 V1 = \text{random}(pd5, 1000000, 1);

27

28 \%
29 pd6 = \text{makedist('Lognormal', muvmaxr, sigmavmaxr)};
30 V2 = \text{random}(pd6, 1000000, 1);

31 \%
32 \text{Calculate geometric coefficients of variation:}
33 GCV1 = \exp(sigmKeq) - 1;
34 GCV2 = \exp(sigmakA) - 1;
35 GCV3 = \exp(sigmakP) - 1;
36 GCV4 = \exp(sigmakiQ) - 1;
37 GCV5 = \exp(sigmavmaxr) - 1;
38 GCV6 = \exp(sigmavmaxf) - 1;

39 \%
40 \text{Choose the parameter with the largest coefficient of variation:}
41 A = [GCV1 GCV2 GCV3 GCV4 GCV5 GCV6];
42 M = \max(A);

43 \%
44 \text{Set dependent parameter based on the largest GCV and}
45 \text{calculate new mu and sigma for the dependent distribution:}
46
47 \text{if } M == \text{GCV2} \%
48 \text{KA dependent}
49 kA = (V1.*kiQ.*kP/)(V2.*Keq);
50 mukA = (mukP + mukiQ + muvmaxf) - (muKeq + muvmaxr);
51 sigmakA = \sqrt{\text{sigmaKeq}^2 + \text{sigmakP}^2 + \text{sigmakiQ}^2 + \\
52 \text{sigmavmaxr}^2 + \text{sigmavmaxf}^2};

53 \text{elseif } M == \text{GCV1} \%
54 \text{K eq dependent}
55 Keq = (kP.*kiQ.*V1/)(kA.*V2);
56 muKeq = (muvmaxf + mukP + mukiQ) - (mukA + muvmaxr);
56 \sigma_{Keq} = \sqrt{(\sigma_{kA}^2 + \sigma_{kP}^2 + \sigma_{kiQ}^2 + \\
57 \sigma_{vmaxr}^2 + \sigma_{vmaxf}^2)};

59 \text{elseif } M == \text{GCV3} \quad \% \, kP \text{ dependent }
60 \quad kP = (kA.*Keq.*V2)./(kiQ.*V1);
61 \quad mukP = (muKeq+mukA+muvmaxr)-(mukiQ+muvmaxf);
62 \quad \sigma_{kP} = \sqrt{(\sigma_{Keq}^2 + \sigma_{kA}^2 + \sigma_{kiQ}^2 + \\
63 \sigma_{vmaxr}^2 + \sigma_{vmaxf}^2)};

65 \text{elseif } M == \text{GCV4} \quad \% \, kiQ \text{ dependent }
66 \quad kiQ = (kA.*Keq.*V2)./(kP.*V1);
67 \quad mukiQ = (muKeq+mukA+muvmaxr)-(mukP+muvmaxf);
68 \quad \sigma_{kiQ} = \sqrt{(\sigma_{Keq}^2 + \sigma_{kA}^2 + \sigma_{kP}^2 + \\
69 \sigma_{vmaxr}^2 + \sigma_{vmaxf}^2)};

71 \text{elseif } M == \text{GCV5} \quad \% \, vmaxf \text{ dependent }
72 \quad V1 = (kA.*Keq.*V2)./(kP.*kP);
73 \quad muvmaxf = (muKeq+mukA+muvmaxr)-(mukiQ+mukP);
74 \quad \sigma_{vmaxf} = \sqrt{(\sigma_{Keq}^2 + \sigma_{kA}^2 + \sigma_{kiQ}^2 + \\
75 \sigma_{vmaxr}^2 + \sigma_{kP}^2)};

77 \text{elseif } M == \text{GCV6} \quad \% \, vmaxr \text{ dependent }
78 \quad V2 = (V1.*kiQ.*kP)./(kA.*Keq);
79 \quad muvmaxr = (muKeq+mukiQ+muvmaxf)-(muKeq+mukA);
80 \quad \sigma_{vmaxr} = \sqrt{(\sigma_{Keq}^2 + \sigma_{kP}^2 + \sigma_{kiQ}^2 + \\
81 \sigma_{kA}^2 + \sigma_{vmaxf}^2)};

83 \text{if } M == \text{GCV1} \quad \% \, Keq \text{ dependent }
84 \quad \% \, \text{Generate multivariate distribution:}
85 \quad B = [Keq \, kA \, kP \, kiQ \, V1];
86 \quad CorrMat = corrcoef(B);
87 \quad Mu = [muKeq \, mukA \, mukP \, mukiQ \, muvmaxf];
88 \quad Sigma = [\sigma_{Keq} \, \sigma_{kA} \, \sigma_{kP} \, \sigma_{kiQ} \, \sigma_{vmaxf}];
89 \quad \% \, \text{Force column vectors}
90 \quad Mu = Mu(:);
91 \quad Sigma = Sigma(:);
92 \quad \% \, \text{Calculate the covariance structure}
Appendix A. Functions for the design of informative priors protocol

```
sigma_down = repmat( Sigma', numel(Sigma), 1 );
sigma_acrs = repmat( Sigma, 1, numel(Sigma) );
covv = log( CorrMat .* sqrt(exp(sigma_down.^2)-1) .* sqrt(exp(sigma_acrs.^2)-1) + 1 );

% The Simulation
y = exp( mvnrnd( Mu , covv , ParamNo ));
Keqfin=y(:,1);
KAfin=y(:,2);
KPfin=y(:,3);
KiQfin=y(:,4);
V1fin=y(:,5);
V2fin=(V1fin.*KiQfin.*KPfin)./(KAfin.*Keqfin);

% Generate multivariate distribution:
B=[kA kP kiQ V1 V2];
CorrMat=corrcoef(B);
Mu = [mukA mukP mukiQ muvmaxf muvmaxr];
Sigma = [sigmakA sigmakP sigmakiQ sigmavmaxf sigmavmaxr];
% Force column vectors
Mu = Mu(:);
Sigma = Sigma(:);
% Calculate the covariance structure
sigma_down = repmat( Sigma', numel(Sigma), 1 );
sigma_acrs = repmat( Sigma, 1, numel(Sigma) );
covv = log( CorrMat .* sqrt(exp(sigma_down.^2)-1) .* sqrt(exp(sigma_acrs.^2)-1) + 1 );
% The Simulation
y = exp( mvnrnd( Mu , covv , ParamNo ));
KAfin=y(:,1);
KPfin=y(:,2);
KiQfin=y(:,3);
V1fin=y(:,4);
V2fin=y(:,5);
Keqfin=(KPfin.*KiQfin.*V1fin)./(KAfin.*V2fin);
end
```
Appendix B

Supplementary case study for the design of informative priors protocol

The step-by-step application of the protocol described in Chapter 3 is implemented to the model of the glycolysis pathway in the sleeping-sickness parasite, *Trypanosoma brucei*, of Helfert et al. (Biochem J. 2001, 357:117-25)\(^{268}\). The structure of the model was obtained from the BioModels database (Publication ID: 11415442, BioModels ID: BIOMD0000000071). The model simulates an *in vivo* experiment using *T. brucei* under physiological conditions where the pH is approximately 7 and the temperature is 37°C. This glycolysis pathway model has 14 reactions as illustrated in Figure B.1.
Figure B.1: Schematic representation of the model. Blue arrows correspond to the maintenance of the ATP/ADP ratio by direct assignment (in combination with reaction 13), rather than by differential equations, in the published model. Likewise, the cytosolic glycerol levels are kept at zero by direct assignment, corresponding to rapid export of glycerol.
Appendix B. Case study 2 for the design of informative priors protocol

TABLE B.1: Case study model reactions

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Model Reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r1)</td>
<td>Glucose external $\leftrightarrow$ Glucose</td>
</tr>
<tr>
<td>(r2)</td>
<td>Phosphates in Glycosome + Glucose $^{ATP_{gly}, ADP_{gly}}$ Glucose 6-phosphate</td>
</tr>
<tr>
<td>(r3)</td>
<td>Glucose 6-phosphate $\leftrightarrow$ Fructose 6-phosphate</td>
</tr>
<tr>
<td>(r4)</td>
<td>Phosphates in Glycosome + Fructose 6-phosphate $^{ATP_{gly}}$ Fructose 1,6-bisphosphate</td>
</tr>
<tr>
<td>(r5)</td>
<td>Fructose 1,6-bisphosphate $^{DHAP_{gly}, ATP_{gly}, ADP_{gly}}$ Glyceraldehyde 3-phosphate + Dihydroxyacetone phosphate</td>
</tr>
<tr>
<td>(r6)</td>
<td>Dihydroxyacetone phosphate $^{DHAP_{gly}}$ Glyceraldehyde 3-phosphate</td>
</tr>
<tr>
<td>(r7)</td>
<td>Glyceraldehyde 3-phosphate + NAD $\leftrightarrow$ NADH + 1,3-bisphosphoglycerate</td>
</tr>
<tr>
<td>(r8)</td>
<td>Dihydroxyacetone phosphate + NADH $^{DHAP_{gly}, Glycerol_{3-phosphate}}$ NAD + Glycerol 3-phosphate</td>
</tr>
<tr>
<td>(r9)</td>
<td>Glycerol 3-phosphate $^{Glycerol_{3-phosphate}}$ Dihydroxyacetone phosphate</td>
</tr>
<tr>
<td>(r10)</td>
<td>Pyruvate $\rightarrow$ Pyruvate external</td>
</tr>
<tr>
<td>(r11)</td>
<td>1,3-bisphosphoglycerate $^{3-PGA_{gly}, ATP_{gly}, ADP_{gly}}$ 3-PGA 2-PGA PEP + Phosphates in Glycosome</td>
</tr>
<tr>
<td>(r12)</td>
<td>3-PGA 2-PGA PEP $^{PEP_{c}, ADP_{c}, ATP_{c}}$ Phosphates cytosol + Pyruvate</td>
</tr>
<tr>
<td>(r13)</td>
<td>Phosphates cytosol $^{ADP_{c}, ATP_{c}}$ $\emptyset$</td>
</tr>
<tr>
<td>(r14)</td>
<td>Glycerol 3-phosphate $^{Glycerol_{3-phosphate}}$ Phosphates in Glycosome + Glycerol</td>
</tr>
</tbody>
</table>

Find and document parameter values

A range of values was found in the literature for each of the parameters in the model. The collection and documentation of the parameter values was performed in the same manner as in Step 1 of the “Anticipated Results” (Section 3.5) of Chapter 3. The documented parameter values are briefly summarised in Table B.2.

TABLE B.2: Summary of parameter values with their associated uncertainty retrieved from literature for the trypanosome case study

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Name*</th>
<th>Value</th>
<th>Units</th>
<th>SD</th>
<th>Uncertainty</th>
<th>Reference</th>
</tr>
</thead>
</table>
### Appendix B. Case study 2 for the design of informative priors protocol

<table>
<thead>
<tr>
<th>Vm1</th>
<th>106.2 nM/min</th>
<th>NaN</th>
<th>Multiplicative</th>
<th>Ter Kuile, B.H. et al. (1991) J. Biol. Chem. 266:857-62</th>
</tr>
</thead>
<tbody>
<tr>
<td>K2Glc1</td>
<td>0.1 mM</td>
<td>NaN</td>
<td>Multiplicative</td>
<td>Bakker, B. M. et al. (1997) J. Biol. Chem. 272:32073215</td>
</tr>
<tr>
<td>K3Fru6P</td>
<td>0.12 mM</td>
<td>0.045</td>
<td>Additive</td>
<td>Marchand, M., et al. (1989) Eur J Biochem. 184(2):455-64</td>
</tr>
</tbody>
</table>

r5
## Appendix B. Case study 2 for the design of informative priors protocol

<table>
<thead>
<tr>
<th>Substance</th>
<th>Concentration</th>
<th>Unit</th>
<th>Distance</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>K5GAP</td>
<td>0.067</td>
<td>mM</td>
<td>NaN</td>
<td>Multiplicative</td>
</tr>
<tr>
<td>K5GAPi</td>
<td>0.098</td>
<td>mM</td>
<td>NaN</td>
<td>Multiplicative</td>
</tr>
<tr>
<td>Vm5f</td>
<td>560</td>
<td>nM/min</td>
<td>153</td>
<td>Additive</td>
</tr>
<tr>
<td>Vm5r</td>
<td>219.555</td>
<td>nM/min</td>
<td>NaN</td>
<td>Multiplicative</td>
</tr>
<tr>
<td>K6DHAPg</td>
<td>1.2</td>
<td>mM</td>
<td>0.1</td>
<td>Additive</td>
</tr>
<tr>
<td>r6</td>
<td>K6DHAPg</td>
<td>0.62</td>
<td>mM</td>
<td>0.01</td>
</tr>
<tr>
<td>K6DHAPg</td>
<td>0.6</td>
<td>mM</td>
<td>0.02</td>
<td>Additive</td>
</tr>
<tr>
<td>K6GAP</td>
<td>0.25</td>
<td>mM</td>
<td>0.05</td>
<td>Additive</td>
</tr>
<tr>
<td>K6GAP</td>
<td>0.32</td>
<td>mM</td>
<td>0.02</td>
<td>Additive</td>
</tr>
<tr>
<td>K6GAP</td>
<td>0.32</td>
<td>mM</td>
<td>0.01</td>
<td>Additive</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Vm6</th>
<th>1000</th>
<th>nM/min</th>
<th>NaN</th>
<th>Multiplicative</th>
</tr>
</thead>
<tbody>
<tr>
<td>r7</td>
<td>K7BPGA13</td>
<td>0.10</td>
<td>mM</td>
<td>0.06</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>K7GAP</td>
<td>0.15</td>
<td>mM</td>
<td>0.06</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>K7GAP</td>
<td>0.17</td>
<td>mM</td>
<td>0.01</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>K7NAD</td>
<td>0.45</td>
<td>mM</td>
<td>0.18</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>K7NADH</td>
<td>0.02</td>
<td>mM</td>
<td>0.009</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>Vm7r</td>
<td>161</td>
<td>nM/min</td>
<td>27.7</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>Albert, M.A et al. (2005), J. Biol. Chem. 280:28306-28315</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vm7f</td>
<td>710</td>
<td>nM/min</td>
<td>288</td>
<td>Additive</td>
</tr>
<tr>
<td></td>
<td>Albert, M.A et al. (2005), J. Biol. Chem. 280:28306-28315</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>K8DHAPg</td>
<td>0.1</td>
<td>mM</td>
<td>NaN</td>
<td>Multiplicative</td>
</tr>
<tr>
<td></td>
<td>K8DHAPg</td>
<td>0.12</td>
<td>mM</td>
<td>NaN</td>
<td>Multiplicative</td>
</tr>
<tr>
<td></td>
<td>K8DHAPg</td>
<td>0.22</td>
<td>mM</td>
<td>NaN</td>
<td>Multiplicative</td>
</tr>
<tr>
<td></td>
<td>K8Gly3Pg</td>
<td>2.25</td>
<td>mM</td>
<td>NaN</td>
<td>Multiplicative</td>
</tr>
<tr>
<td></td>
<td>K8Gly3Pg</td>
<td>1.7</td>
<td>mM</td>
<td>NaN</td>
<td>Multiplicative</td>
</tr>
</tbody>
</table>
Appendix B. Case study 2 for the design of informative priors protocol

<table>
<thead>
<tr>
<th>Substance</th>
<th>Concentration</th>
<th>Enzyme</th>
<th>P/N</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>K8Gly3Pg</td>
<td>0.18 mM NaN</td>
<td></td>
<td></td>
<td>Multiplicative</td>
</tr>
<tr>
<td>K8NAD</td>
<td>0.32 mM NaN</td>
<td></td>
<td></td>
<td>Multiplicative</td>
</tr>
<tr>
<td>K8NAD</td>
<td>0.42 mM NaN</td>
<td></td>
<td></td>
<td>Multiplicative</td>
</tr>
<tr>
<td>K8NAD</td>
<td>0.012 mM NaN</td>
<td></td>
<td></td>
<td>Multiplicative</td>
</tr>
<tr>
<td>K8NADH</td>
<td>0.01 mM NaN</td>
<td></td>
<td></td>
<td>Multiplicative</td>
</tr>
<tr>
<td>K8NADH</td>
<td>0.024 mM NaN</td>
<td></td>
<td></td>
<td>Multiplicative</td>
</tr>
<tr>
<td>K8NADH</td>
<td>0.008 mM NaN</td>
<td></td>
<td></td>
<td>Multiplicative</td>
</tr>
<tr>
<td>V8f</td>
<td>465 nM/min</td>
<td>21</td>
<td></td>
<td>Additive</td>
</tr>
<tr>
<td>V8r</td>
<td>130.2 nM/min</td>
<td>NaN</td>
<td></td>
<td>Multiplicative</td>
</tr>
<tr>
<td>r9</td>
<td>Vm9 3.68E+02 nM/min</td>
<td>NaN</td>
<td></td>
<td>Multiplicative</td>
</tr>
<tr>
<td>K9Gly3Pc</td>
<td>1.72 mM 0.08</td>
<td></td>
<td></td>
<td>Additive</td>
</tr>
</tbody>
</table>

Unpublished data from P.A.Michels
<table>
<thead>
<tr>
<th>r10</th>
<th>K10Pyr</th>
<th>1.96 mM</th>
<th>0.28 Additive</th>
<th>Wiemer, E.A. et al. (1992) Biochem Biophys Res Commun. 184(2):1028-34.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K11ADPg</td>
<td>0.1 mM</td>
<td>NaN Multiplicative</td>
<td>Bakker, B. M. et al. (1997) J. Biol. Chem. 272:32073215</td>
</tr>
<tr>
<td></td>
<td>K11ATPg</td>
<td>0.29 mM</td>
<td>0.15 Additive</td>
<td>Misset, O. et al. (1987) Eur J Biochem. 162(3):493-500.</td>
</tr>
<tr>
<td></td>
<td>Vm11r</td>
<td>1358 nM/min</td>
<td>254 Additive</td>
<td>Albert, M.A. et al. (2005) J. Biol. Chem. 280:28306-28315</td>
</tr>
<tr>
<td></td>
<td>Vm12</td>
<td>1020 nM/min</td>
<td>221 Additive</td>
<td>Albert, M.A. et al. (2005) J. Biol. Chem. 280:28306-28315</td>
</tr>
</tbody>
</table>
### Judge plausibility of values and assign weights

Upon the detailed documentation of the parameter sources, values and uncertainties, the next step is to assign the appropriate weight to each parameter observation. For this case study, the weighting scheme presented in Table B.3 was employed. For parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
<th>Weight</th>
<th>Type</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>K14ADPg</td>
<td>0.56</td>
<td>mM</td>
<td>0.02</td>
<td>Additive</td>
<td>Králová, I. et al. (2000) Eur J Biochem. 267(8):2323-33.</td>
</tr>
<tr>
<td>K14ATPg</td>
<td>0.24</td>
<td>mM</td>
<td>0.09</td>
<td>Additive</td>
<td>Králová, I. et al. (2000) Eur J Biochem. 267(8):2323-33.</td>
</tr>
</tbody>
</table>

* Parameter names are as used in the published model.
for which only a single value was found, this step was omitted and a 10% relative standard deviation applied as described in Step 1B in the “Procedure” (Section 3.3) in the main manuscript. Parameters with more than one value are weighted according to the proposed scheme, and the result is shown in Table B.4.
### Table B.3: Weighting scheme used to assess the plausibility of parameter values found for reactions in glycolysis pathway

<table>
<thead>
<tr>
<th>Categories</th>
<th>Subcategories</th>
<th>Description</th>
<th>Weight value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>In vivo</td>
<td>Parameter value was measured <em>in vivo</em> the ideal condition for our modelling system</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>In vitro</td>
<td>Parameter value was measured <em>in vitro</em>, and does not match our modelling system, but the values are still useful.</td>
<td>1</td>
</tr>
<tr>
<td>Organism</td>
<td>Identical</td>
<td>In the ideal case, the characterized enzyme was obtained from <em>Trypanosoma brucei</em>, our model organism.</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Related</td>
<td>Enzymes from other protozoan organisms (i.e. from the same phylum)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Unrelated</td>
<td>Enzymes from all other organisms</td>
<td>1</td>
</tr>
<tr>
<td>Protein/Enzyme</td>
<td>Identical</td>
<td>Parameter value is measured for a protein/enzyme catalyzing the same reaction as in the model</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Related</td>
<td>Parameter value is measured for a reaction from a related enzyme class (shared three leading digits of the EC number)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Different</td>
<td>Parameter value is measured for a different protein/enzyme class</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Identical</td>
<td>Ideally, the pH and temperature of the assay matches the modelling conditions pH 7.0 and 37°C.</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Similar or within range</td>
<td>Only partial information is available, but the reported information matches the ideal modelling condition (e.g., temperature information missing, but pH 7.0 is reported, or pH information is missing, but 37°C is reported). Also applied when the pH is within ± 1 of pH 7 and temperature reported is within ± 2°C of 37°C.</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Experimental Conditions</td>
<td>The parameter value is measured in conditions that differ more extremely from the modelling conditions.</td>
<td>1</td>
</tr>
</tbody>
</table>
### TABLE B.4: Summary of parameter values and assignment of weights for case study

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Parameter</th>
<th>Value</th>
<th>SD</th>
<th>Method</th>
<th>Organism</th>
<th>Protein</th>
<th>Condition</th>
<th>Total weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>r6</td>
<td>K6DHAPg</td>
<td>1.2</td>
<td>0.1</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>K6DHAPg</td>
<td>0.62</td>
<td>0.01</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8/2*</td>
</tr>
<tr>
<td></td>
<td>K6DHAPg</td>
<td>0.6</td>
<td>0.02</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8/2*</td>
</tr>
<tr>
<td></td>
<td>K6GAP</td>
<td>0.25</td>
<td>0.05</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>K6GAP</td>
<td>0.32</td>
<td>0.02</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8/2*</td>
</tr>
<tr>
<td></td>
<td>K6GAP</td>
<td>0.32</td>
<td>0.01</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8/2*</td>
</tr>
<tr>
<td></td>
<td>K7GAP</td>
<td>0.17</td>
<td>0.01</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>K7GAP</td>
<td>0.15</td>
<td>0.06</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>32/2*</td>
</tr>
<tr>
<td>r7</td>
<td>K8DHAPg</td>
<td>0.22</td>
<td>NaN</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>K8DHAPg</td>
<td>0.1</td>
<td>NaN</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>K8DHAPg</td>
<td>0.12</td>
<td>NaN</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>K8Gly3Pg</td>
<td>2.25</td>
<td>NaN</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>K8Gly3Pg</td>
<td>1.7</td>
<td>NaN</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>K8Gly3Pg</td>
<td>0.18</td>
<td>NaN</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>r8</td>
<td>K8NAD</td>
<td>0.32</td>
<td>NaN</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>K8NAD</td>
<td>0.42</td>
<td>NaN</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>K8NAD</td>
<td>0.012</td>
<td>NaN</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>K8NADH</td>
<td>0.01</td>
<td>NaN</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>K8NADH</td>
<td>0.024</td>
<td>NaN</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>K8NADH</td>
<td>0.008</td>
<td>NaN</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>

* If more than one parameter value (for the same parameter) is retrieved from the same publication, the total weight of each of the values is divided by the total number of values used from this specific publication. This avoids inflating the influence of a single study.
Calculate the Mode and Spread and location and scale parameters \((\mu \text{ and } \sigma)\) of the log-normal distribution

Using the input matrices in Table B.5 as separate inputs for the `CalcModeSpread` and `CreateLogNormDist` functions for each parameter produces the results summarised in Table B.6. Uncertainty type 0 refers to values with additive uncertainty, type 1 to multiplicative uncertainty.
### Table B.5: Input matrices

<table>
<thead>
<tr>
<th>Input matrix</th>
<th>Parameter</th>
<th>Value</th>
<th>Uncertainty</th>
<th>Weight</th>
<th>Uncertainty type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (r6)</td>
<td>K6DHAPg</td>
<td>1.2</td>
<td>0.1</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>K6DHAPg</td>
<td>0.62</td>
<td>0.01</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>K6DHAPg</td>
<td>0.6</td>
<td>0.02</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>K6GAP</td>
<td>0.25</td>
<td>0.05</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>2 (r6)</td>
<td>K6GAP</td>
<td>0.32</td>
<td>0.02</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>K6GAP</td>
<td>0.32</td>
<td>0.01</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>3 (r7)</td>
<td>K7BPGA13</td>
<td>0.013</td>
<td>0.0035</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>K7BPGA13</td>
<td>0.01</td>
<td>0.06</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>4 (r7)</td>
<td>K7GAP</td>
<td>0.17</td>
<td>0.01</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>K7GAP</td>
<td>0.15</td>
<td>0.06</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>K8DHAPg</td>
<td>0.1</td>
<td>NaN</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>5 (r8)</td>
<td>K8DHAPg</td>
<td>0.12</td>
<td>NaN</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>K8DHAPg</td>
<td>0.22</td>
<td>NaN</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>K8Gly3Pg</td>
<td>2.25</td>
<td>NaN</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>6 (r8)</td>
<td>K8Gly3Pg</td>
<td>1.7</td>
<td>NaN</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>K8Gly3Pg</td>
<td>0.18</td>
<td>NaN</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>K8NAD</td>
<td>0.32</td>
<td>NaN</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>7 (r8)</td>
<td>K8NAD</td>
<td>0.42</td>
<td>NaN</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>K8NAD</td>
<td>0.012</td>
<td>NaN</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>K8NADH</td>
<td>0.01</td>
<td>NaN</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>8 (r8)</td>
<td>K8NADH</td>
<td>0.024</td>
<td>NaN</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>K8NADH</td>
<td>0.008</td>
<td>NaN</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table B.6: Summary of log-normal distribution properties for parameters in the trypanosome glycolysis pathway

<table>
<thead>
<tr>
<th>Reaction Name</th>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$x_{\text{min}}$</th>
<th>$x_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>K1Glc</td>
<td>1.9909</td>
<td>1.21</td>
<td>0.72369</td>
<td>0.18732</td>
<td>1.6454</td>
<td>2.409</td>
</tr>
</tbody>
</table>
Appendix B. Case study 2 for the design of informative priors protocol

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>r1</td>
<td>Vm1</td>
<td>105.72</td>
<td>1.21</td>
<td>4.6959</td>
<td>0.18732</td>
</tr>
<tr>
<td>r2</td>
<td>K2ADPpg</td>
<td>0.12543</td>
<td>1.21</td>
<td>-2.0409</td>
<td>0.18732</td>
</tr>
<tr>
<td>r2</td>
<td>K2ATPpg</td>
<td>0.11547</td>
<td>1.21</td>
<td>-2.1236</td>
<td>0.18732</td>
</tr>
<tr>
<td>r2</td>
<td>K2Glc6P</td>
<td>11.946</td>
<td>1.21</td>
<td>2.5155</td>
<td>0.18732</td>
</tr>
<tr>
<td>r2</td>
<td>K2GlcI</td>
<td>0.099547</td>
<td>1.21</td>
<td>-2.272</td>
<td>0.18732</td>
</tr>
<tr>
<td>r2</td>
<td>Vm2</td>
<td>1925.5</td>
<td>1.1277</td>
<td>7.5772</td>
<td>0.11931</td>
</tr>
<tr>
<td>r3</td>
<td>K3Fru6P</td>
<td>0.11236</td>
<td>2.0657</td>
<td>-1.8138</td>
<td>0.61011</td>
</tr>
<tr>
<td>r3</td>
<td>K3Glc6P</td>
<td>0.39819</td>
<td>1.21</td>
<td>-0.88574</td>
<td>0.18732</td>
</tr>
<tr>
<td>r3</td>
<td>Vm3</td>
<td>1300</td>
<td>1.1923</td>
<td>7.2001</td>
<td>0.17329</td>
</tr>
<tr>
<td>r4</td>
<td>K4ATPpg</td>
<td>0.02587</td>
<td>1.21</td>
<td>-3.6196</td>
<td>0.18732</td>
</tr>
<tr>
<td>r4</td>
<td>K4Fru6P</td>
<td>0.81628</td>
<td>1.21</td>
<td>-0.1679</td>
<td>0.18732</td>
</tr>
<tr>
<td>r4</td>
<td>K4i1Fru16BP</td>
<td>15.728</td>
<td>1.21</td>
<td>2.7906</td>
<td>0.18732</td>
</tr>
<tr>
<td>r4</td>
<td>K4i2Fru16BP</td>
<td>10.652</td>
<td>1.21</td>
<td>2.4008</td>
<td>0.18732</td>
</tr>
<tr>
<td>r4</td>
<td>Vm4</td>
<td>1683.9</td>
<td>1.4155</td>
<td>7.5374</td>
<td>0.32945</td>
</tr>
<tr>
<td>r5</td>
<td>K5DHAP</td>
<td>0.014932</td>
<td>1.21</td>
<td>-4.1692</td>
<td>0.18732</td>
</tr>
<tr>
<td>r5</td>
<td>K5GAP</td>
<td>0.066696</td>
<td>1.21</td>
<td>-2.6725</td>
<td>0.18732</td>
</tr>
<tr>
<td>r5</td>
<td>K5GAPi</td>
<td>0.097556</td>
<td>1.21</td>
<td>-2.2922</td>
<td>0.18732</td>
</tr>
<tr>
<td>r5</td>
<td>Vm5f</td>
<td>540.2</td>
<td>1.7102</td>
<td>6.5228</td>
<td>0.48043</td>
</tr>
<tr>
<td>r5</td>
<td>Vm5r</td>
<td>218.56</td>
<td>1.21</td>
<td>5.4221</td>
<td>0.18732</td>
</tr>
<tr>
<td>r6</td>
<td>K6DHAPpg</td>
<td>1.1352</td>
<td>1.35</td>
<td>6.20821</td>
<td>0.28537</td>
</tr>
<tr>
<td>r6</td>
<td>K6GAP</td>
<td>0.2618</td>
<td>1.39</td>
<td>-1.2435</td>
<td>0.31118</td>
</tr>
<tr>
<td>r6</td>
<td>Vm6</td>
<td>1E-03</td>
<td>1.1</td>
<td>-6.8988</td>
<td>0.0949</td>
</tr>
<tr>
<td>r7</td>
<td>K7BPGA13</td>
<td>0.1</td>
<td>1.67</td>
<td>-2.0885</td>
<td>0.4627</td>
</tr>
<tr>
<td>r7</td>
<td>K7GAP</td>
<td>0.15</td>
<td>2.5</td>
<td>-1.373</td>
<td>0.7239</td>
</tr>
<tr>
<td>r7</td>
<td>K7NAD</td>
<td>0.45</td>
<td>2.5</td>
<td>-0.2744</td>
<td>0.7239</td>
</tr>
<tr>
<td>r7</td>
<td>K7NADH</td>
<td>0.02</td>
<td>2.22</td>
<td>-3.4832</td>
<td>0.6549</td>
</tr>
<tr>
<td>r7</td>
<td>Vm7f</td>
<td>236.7</td>
<td>2.47</td>
<td>5.9811</td>
<td>0.7172</td>
</tr>
<tr>
<td>r7</td>
<td>Vm7r</td>
<td>710</td>
<td>2.47</td>
<td>7.0796</td>
<td>0.7172</td>
</tr>
<tr>
<td>r8</td>
<td>K8DHAPpg</td>
<td>0.1055</td>
<td>1.355</td>
<td>-2.1648</td>
<td>0.291</td>
</tr>
<tr>
<td>r8</td>
<td>K8Gly3Pg</td>
<td>2.11</td>
<td>2.54</td>
<td>1.279</td>
<td>0.73209</td>
</tr>
<tr>
<td>r8</td>
<td>K8NAD</td>
<td>0.3184</td>
<td>3.5</td>
<td>-0.34533</td>
<td>0.89394</td>
</tr>
</tbody>
</table>
Account for thermodynamic consistency

The model contains 8 reversible reactions that should take thermodynamic consistency into account.

1. **Glucose transport (r1): Uni–Uni reaction**

   The parameters used for this reaction are K1Glc (K_{M}^+ equal to K_{M}^-; the justification of this and other assumptions in the model could be debated, but we are
following the original publication for all steps of this illustrative case study, to facilitate the comparison of the results) and $V_{m1} (V_{m1}^+ \text{ equal to } V_{m1}^-)$. We therefore assume that $K_M^+$ and $K_M^-$ are sampled from the same distributions; the same applies for $V_{m1}^+$ and $V_{m1}^-$. The equilibrium constant $K_{eq}$ is assumed to be equal to 1 with a very small uncertainty (Mode: 1, Spread: 1.01, $\mu = 9.8995 \times 10^{-05}$, $\sigma = 0.0099$), as the transport reaction is considered as facilitated diffusion, not coupled to any source of energy that would lead to a shift of the equilibrium. The total concentration of transporter, $E_0$, is another required input parameter; this was not provided in the original model, and for illustration purposes we used a value of 1 mM here and for the subsequent enzymes (usually, the approximate concentrations would be known, and their uncertainty would be represented by the uncertainty of the $V_{max}$ values). Therefore, using the information from Table B.6, the input for the multivariate distribution function is:

1. $[K_{eq}, K_{1Glc}, K_{1GlcR}, V_{m1}, V_{m1R}] = ...$
   Multivariate5param($9.8995 \times 10^{-05}$, 0.0099, 0.72369, 0.18732, ...
   0.72369, 0.18732, 4.6959, 0.18732, 4.6959, 0.18732, ...
   ParamNo, 1)

2. Glucose-phosphate isomerase (r3): Uni–Uni reaction

The parameters used for this reaction are $K_{3Fru6P} (K_{3Fru6P}^+)$, $K_{3Glc6P} (K_{3Glc6P}^-)$ and $V_{m3} (V_{m3}^+)$. Based on the rate law of this reaction of the model, we assume that $V_{m3}^-$ is sampled from the same distribution as $V_{m3}^+$. The equilibrium constant $K_{eq}$ is assumed to be equal to 1 with a relatively narrow range of uncertainty (Mode: 1, Spread: 1.2, $\mu = 0.0322$, $\sigma = 0.1794$), as the $\Delta G^0$ of an isomerization reaction is expected to be close to 0, and $E_0 = 1$. Therefore, the input for the multivariate distribution function is:
Appendix B. Case study 2 for the design of informative priors protocol

3. Aldolase (r5): Uni–Bi reaction

The parameters used for this reaction are K5DHAP (\(K_P\)), K5GAPi (\(K_iQ\)), Vm5f (\(V_{max}^+\)) and Vm5r (\(V_{max}^-\)). The \(K_m\) for Fructose (\(K_A\)) is not used as a parameter in the model, but as a constant with a value of 0.009, and will be used in the Haldane equation to restrict the sampling of the other parameters. The equilibrium constant \(K_{eq}\) can be set as the dependent parameter and calculated via the Haldane equation, \(K_{eq} = \frac{K_P \cdot K_iQ \cdot V_1}{K_A \cdot V_2}\) (\(\mu = -0.6506\) and \(\sigma = 0.5797\)). Therefore, the input for the multivariate distribution function is:

```plaintext
[Keq, K5DHAP, K5GAPi, Vm5f, Vm5r] = ... Multivariate5param(0.0322, 0.1794, 0.20821, 0.28537, ... -1.2435, 0.31118, 7.2001, 0.17329, -6.8988, 0.0949, ... ParamNo, 1)
```

4. Triosephosphate isomerase (r6): Uni–Uni reaction

The parameters used for this reaction are K6DHAPg (\(K_{M}^+\)), K6GAP (\(K_M^-\)) and Vm6 (\(V_{max}^+\)). From the rate law of this reaction of the model, we assume that \(V_{max}^-\) is sampled from the same distribution as \(V_{max}^+\). The equilibrium constant \(K_{eq}\) for this isomerization reaction is again assumed to be equal to 1, with a narrow range of uncertainty (Mode: 1, Spread: 1.2, \(\mu = 0.0322\), \(\sigma = 0.1794\)), and \(E_0 = 1\). Therefore, the input for the multivariate distribution function is:

```plaintext
[Keq, K6DHAPg, K6GAP, Vm6, Vm6R] = ... Multivariate5param(0.0322, 0.1794, 0.20821, 0.28537, ... -1.2435, 0.31118, 7.2001, 0.17329, -6.8988, 0.0949, ... ParamNo, 1)
```
5. Glyceraldehyde 3-phosphate dehydrogenase (r7): Bi–Bi reaction

The parameters used for this reaction are K7GAP (K_A), K7NAD (K_B), K7BPGA13 (K_P), K7NADH (K_Q), Vm7f (V_{max}^+) and Vm7r (V_{max}^-). The equilibrium constant \( K_{eq} \) can be set as the dependent parameter and calculated via the Haldane equation, \( K_{eq} = \frac{K_P \cdot K_Q \cdot V_{max}^2}{K_A \cdot K_B \cdot V_{max}^-} \) (\( \mu = -6.1213 \) and \( \sigma = 3.139 \)). Therefore, the input for the multivariate distribution function is:

\[
[Keq, K7GAP, K7NAD, K7BPGA13, K7NADH, Vm7f, Vm7r] = ... \\
\text{MultivariateBiBi}(-6.1213, 3.139, -1.373, 0.7239, ... \\
-0.2744, 0.7239, -2.0885, 0.4627, -3.4832, 0.6549, ... \\
5.9811, 0.7172, 7.0796, 0.7172, \text{ParamNo})
\]

6. Glyceraldehyde 3-phosphate dehydrogenase (r8): Bi–Bi reaction

The parameters used for this reaction are K8DHAPg (K_A), K8NADH (K_B), K8Gly3Pg (K_P), K8NAD (K_Q), V8mf (V_{max}^+) and V8mr (V_{max}^-). The equilibrium constant \( K_{eq} \) can be set as the dependent parameter and calculated via the Haldane equation (\( \mu = 11.7601 \) and \( \sigma = 2.2064 \)). Therefore, the input for the multivariate distribution function is:

\[
[Keq, K8DHAPg, K8NADH, K8Gly3Pg, K8NAD, V8mf, V8mr] = ... \\
\text{MultivariateBiBi}(11.7601, 2.2064, -2.1648, 0.291, ... \\
-4.4868, 0.3441, 1.279, 0.732, -0.3453, 0.8939, 6.9655, ... \\
0.9075, 4.8781, 0.0949, \text{ParamNo})
\]

7. Glyceraldehyde 3-phosphate dehydrogenase (r11): Bi–Bi reaction

The parameters used for this reaction are K11ADPg (K_A), K11BPGA13 (K_B), K11ATPg (K_P), K11PGA3 (K_Q), Vm11f (V_{max}^+) and Vm11r (V_{max}^-). The equilibrium constant \( K_{eq} \) can be set as the dependent parameter and calculated via
the Haldane equation ($\mu=9.63928$ and $\sigma=1.3925$). Therefore, the input for the multivariate distribution function is:

\[
[\text{Keq}, \text{K11ADPg}, \text{K11BPGA13}, \text{K11ATPg}, \text{K11PGA3}, \text{Vm11f}, ... \\
\text{Vm11r}] = \text{MultivariateBiBi}(9.63928, 1.3925, -2.272, ... \\
0.18732, -5.7786, 0.18732, -0.78322, 0.75549, 1.0785, ... \\
0.88462, 7.9653, 0.082836, 7.3186, 0.34936, \text{ParamNo})
\]

8. Glycerol kinase (r14): Bi–Bi reaction

The parameters used for this reaction are $\text{K14ADPg}$ ($K_A$), $\text{K14Gly3Pg}$ ($K_B$), $\text{K14ATPg}$ ($K_P$), $\text{K14Gly}$ ($K_Q$), $\text{Vm14f}$ ($V_{max}^+$) and $\text{Vm14r}$ ($V_{max}^-$). The equilibrium constant $K_{eq}$ can be set as the dependent parameter and calculated via the Haldane equation ($\mu=13.0459$ and $\sigma=1.0336$). Therefore, the input for the multivariate distribution function is:

\[
[\text{Keq}, \text{K14ADPg}, \text{K14Gly3Pg}, \text{K14ATPg}, \text{K14Gly}, \text{Vm14f}, ... \\
\text{Vm14r}] = \text{MultivariateBiBi}(-13.0459, 1.0336, -0.57538, ... \\
0.071223, 1.5674, 0.51736, -1.1189, 0.61011, -0.69882, ... \\
0.3777, 5.3289, 0.18732, 10.447, 0.18732, \text{ParamNo})
\]

Results

The simulations of the published model were repeated by using parameter sets sampled from the generated priors. A representative example of the results is shown in Figure B.2, which reproduces results shown in one of the figures in the original publication, but with confidence intervals added by the ensemble modelling strategy.
Figure B.2: Effect of reducing TPI on the steady-state flux of glucose, pyruvate and glycerol. Replicated Figure 3B of the published model (Helfert et al., Biochem. J. (2001))
Blank page
Appendix C

Parameter information for recreated \(\gamma\)-butyrolactone and quorum sensing systems

Parameter distributions for the GBL model

The log-normal distributions created for each parameter in the published GBL model are found below, along with the relevant information for the total range tested, the reported values for bistability (and the corresponding range of bistability), the mode of the distributions, the Spreads and the location and scale parameters. The grey area in the plots represents the bistability area and the blue line is the mode of the distribution.
Appendix C. Parameter information for the GBL and QS systems models

Range tested: $10^{-5} - 10$ nM.
Published value for bistability: 7.44 nM (range: 3.86-4.39). Mode: 4 nM, Spread: 2.5.

Range tested: $10^{-4} - 10$ s$^{-1}$.
Published value for bistability: 0.18 s$^{-1}$ (range: 0.13-0.2). Mode: 0.18 s$^{-1}$, Spread: 50, $\mu$: -0.0098, $\sigma$: 1.3057

Range tested: $10^{-4} - 10$ s$^{-1}$.
Published value for bistability: 0.45 s$^{-1}$ (range: 0.44-0.5). Mode: 0.18 s$^{-1}$, Spread: 0.0017 s$^{-1}$

Range tested: $6.6 \cdot 10^{-4} - 1.2 \cdot 10^{-2}$ s$^{-1}$.
Published value for bistability: 0.0018 s$^{-1}$ (range: 0.0016-0.0019). Mode: 0.0018 s$^{-1}$, Spread: 10, $\mu$: -5.5259, $\sigma$: 0.8911

Range tested: $6.6 \cdot 10^{-4} - 1.2 \cdot 10^{-2}$ s$^{-1}$.
Published value for bistability: 0.0018 s$^{-1}$ (range: 0.25-0.39). Mode: 0.18 s$^{-1}$, Spread: 50, $\mu$: -0.0098, $\sigma$: 1.3057

Range tested: $10^{-4} - 10$ s$^{-1}$.
Published value for bistability: 0.36 s$^{-1}$ (range: 0.0016-0.0019). Mode: 0.0018 s$^{-1}$, Spread: 10, $\mu$: -5.5259, $\sigma$: 0.8911

Range tested: $6.6 \cdot 10^{-4} - 1.2 \cdot 10^{-2}$ s$^{-1}$.
Published value for bistability: 0.0018 s$^{-1}$ (range: 0.25-0.39). Mode: 0.18 s$^{-1}$, Spread: 50, $\mu$: -0.0098, $\sigma$: 1.3057

Range tested: $10^{-4} - 10$ s$^{-1}$.
Published value for bistability: 0.36 s$^{-1}$ (range: 0.0016-0.0019). Mode: 0.0018 s$^{-1}$, Spread: 10, $\mu$: -5.5259, $\sigma$: 0.8911

Range tested: $6.6 \cdot 10^{-4} - 1.2 \cdot 10^{-2}$ s$^{-1}$.
Published value for bistability: 0.0018 s$^{-1}$ (range: 0.25-0.39). Mode: 0.18 s$^{-1}$, Spread: 50, $\mu$: -0.0098, $\sigma$: 1.3057
Appendix C. Parameter information for the GBL and QS systems models

Range tested: $10^{-4} - 10 \text{s}^{-1}$. Published value for bistability: $0.066 \text{s}^{-1}$ (range: $0.064-0.073$). Mode: $0.18 \text{s}^{-1}$, Spread: 50, $\mu$: -0.0098, $\sigma$: 1.3057

Range tested: $10^{-7} - 10^{-1} \text{s}^{-1}$. Published value for bistability: $0.004 \text{s}^{-1}$ (range: $0.0022-0.0067$). Mode: $0.001 \text{s}^{-1}$, Spread: 100, $\mu$: -4.7724, $\sigma$: 1.4613

Range tested: $10^{-7} - 10^{-1} \text{s}^{-1}$. Published value for bistability: $0.0018 \text{s}^{-1}$ (range: $0.0013-0.0018$). Mode: $0.001 \text{s}^{-1}$, Spread: 100, $\mu$: -4.7724, $\sigma$: 1.4613

Range tested: $0 - 10^{-4} \text{s}^{-1}$. Published value for bistability: $6.7 \times 10^{-5} \text{s}^{-1}$ (range: $0.0001-0.0001$). Mode: $0.0001 \text{s}^{-1}$, Spread: 10, $\mu$: -8.4163, $\sigma$: 0.8911

Range tested: $0 - 1.7 \text{s}^{-1}$. Published value for bistability: $0.74 \text{s}^{-1}$ (range: $0.67-1.7$). Mode: $0.3 \text{s}^{-1}$, Spread: 4, $\mu$: -0.84369, $\sigma$: 0.60023

Range tested: $0 - 10^{-4} \text{s}^{-1}$. Published value for bistability: $6.7 \times 10^{-5} \text{s}^{-1}$ (range: $0.0001-0.0001$). Mode: $0.0001 \text{s}^{-1}$, Spread: 10, $\mu$: -8.4163, $\sigma$: 0.8911
Appendix C. Parameter information for the GBL and QS systems models

Range tested: $10^{-7} - 10^{-1} \text{nM}^{-1} \text{s}^{-1}$.
Published value for bistability: 0.083 $\text{nM}^{-1} \text{s}^{-1}$ (range: 0.077-0.17). Mode: 0.001 $\text{nM}^{-1} \text{s}^{-1}$, Spread: 100, $\mu$: -4.7724, $\sigma$: 1.4613

Range tested: $0 - 10^3 \text{s}^{-1}$.
Published value for bistability: 170 $\text{s}^{-1}$ (range: 24-180). Mode: 150 $\text{s}^{-1}$, Spread: 6.5, $\mu$: 5.5904, $\sigma$: 0.76142

Range tested: $10^{-7} - 10^{-1} \text{nM}^{-1} \text{s}^{-1}$.
Published value for bistability: 0.083 $\text{nM}^{-1} \text{s}^{-1}$ (range: 0.083-0.12). Mode: 0.001 $\text{nM}^{-1} \text{s}^{-1}$, Spread: 100, $\mu$: -4.7724, $\sigma$: 1.4613

Range tested: $0 - 10^3 \text{s}^{-1}$.
Published value for bistability: 630 $\text{s}^{-1}$ (range: 460-630). Mode: 150 $\text{s}^{-1}$, Spread: 6.5, $\mu$: 5.5904, $\sigma$: 0.76142

Range tested: $0 - 4 \times 10^{-1} \text{s}^{-1}$.
Published value for bistability: 0.083 $\text{s}^{-1}$ (range: 0.02-0.097). Mode: 0.08 $\text{s}^{-1}$, Spread: (range: 0.061-0.17). Mode: 0.01 $\text{s}^{-1}$, Spread: 3, $\mu$: -2.28096, $\sigma$: 0.49474

Range tested: $10^{-7} - 10^{-1} \text{s}^{-1}$.
Published value for bistability: 0.062 $\text{s}^{-1}$ (range: 0.02-0.097). Mode: 0.08 $\text{s}^{-1}$, Spread: (range: 0.061-0.17). Mode: 0.01 $\text{s}^{-1}$, Spread: 8, $\mu$: -3.9241, $\sigma$: 0.8253
Appendix C. Parameter information for the GBL and QS systems models

Parameter distributions for the QS model

The log-normal distributions created for each parameter in the published QS model are found below, along with the relevant information for the reported values, the mode of the distributions, the Spreads and the location and scale parameters.
Appendix C. Parameter information for the GBL and QS systems models

Parameter information for the GBL and QS systems models

Published value: 20 nM.
Mode: 20 nM, Spread: 2.5, $\mu$:3.1747, $\sigma$: 0.42306

Published value: 1 min$^{-1}$.
Mode: 1 min$^{-1}$, Spread: 2.5, $\mu$:0.17898, $\sigma$: 0.42306

Published value: 0.04 min$^{-1}$.
Mode: 0.04 min$^{-1}$, Spread: 2.5, $\mu$:3.0399, $\sigma$: 0.4231

Published value: 200 nM (range:0.0015-0.0025).
Mode: 200 nM, Spread: 2.5, $\mu$:5.4773, $\sigma$: 0.42306

Published value: 10 min$^{-1}$.
Mode: 10 min$^{-1}$, Spread: 2.5, $\mu$:2.48156, $\sigma$: 0.42306

Published value: 20.
Mode: 20, Spread: 2.5, $\mu$:3.1747, $\sigma$: 0.42306
Appendix C. Parameter information for the GBL and QS systems models

### Published value: 0.001
- **Mode:** 0.001, **Spread:** 2.5, **μ:** -6.7287, **σ:** 0.42306

### Published value: 0.01
- **Mode:** 0.001, **Spread:** 2.5, **μ:** -4.4262, **σ:** 0.42306

### Published value: 0.002 min⁻¹
- **Mode:** 0.002 min⁻¹, **Spread:** 2.5, **μ:** -6.0356, **σ:** 0.42306
Appendix C. Parameter information for the GBL and QS systems models

### Published values

- **Parameter**: $d_i$ (min$^{-1}$)
  - **Value**: 0.002 min$^{-1}$
  - **Mode**: 0.002 min$^{-1}$
  - **Spread**: 2.5
  - **Mode**: $\mu$: 0.0356, $\sigma$: 0.42306

- **Parameter**: $d_{mi}$ (min$^{-1}$)
  - **Value**: 0.347 min$^{-1}$
  - **Mode**: 0.347 min$^{-1}$
  - **Spread**: 2.5
  - **Mode**: $\mu$: 0.8794, $\sigma$: 0.42306

- **Parameter**: $d_{mr}$ (min$^{-1}$)
  - **Value**: 10 min$^{-1}$
  - **Mode**: 0.45 min$^{-1}$
  - **Spread**: 45
  - **Mode**: $\mu$: 0.84274, $\sigma$: 1.2811

- **Parameter**: $d_{GFP}$ (min$^{-1}$)
  - **Value**: 0.01 min$^{-1}$
  - **Mode**: 0.01 min$^{-1}$
  - **Spread**: 2.5
  - **Mode**: $\mu$: 4.4262, $\sigma$: 0.42306

### Probability Distributions

- **Probability Distribution for $d_i$ values**
- **Probability Distribution for $d_{mi}$ values**
- **Probability Distribution for $d_{mr}$ values**
- **Probability Distribution for $d_{GFP}$ values**
Appendix D

Parameter information and uncertainty for the model of the γ-butyrolactone System

The following content is extracted from the website Systems Biology Models Wiki (http://www.systemsbiology.ls.manchester.ac.uk/wiki/index.php), which documents the literature sources and calculations underlying the model parameter values and is intended for future publication to make the full information on the model construction publicly available. It has been slightly edited for conformity with a printed format, but includes the full information contained in the website version.

D.1 Binding of R₂ to O₉R operator

The ScbR homo-dimer (R₂) binds to its own gene operator (O₉R) and represses its mRNA transcription. As reported by Bhukya et al.[1] two ScbR homo-dimers can bind to the promoter. When one homo-dimer is bound, the mRNA transcription is already being repressed. As the concentration of ScbR rises, a second homo-dimer may bind to the already suppressed promoter and further enhance the suppression of the transcription. Therefore, this behaviour is represented by two separate reactions that may take place independently.
Appendix D. Parameter information for the GBL system model

**Chemical equation**

\[ O_R + R_2 \rightleftharpoons O_R \cdot R_2 \]
\[ O_R \cdot R_2 + R_2 \rightleftharpoons O_R \cdot 2R_2 \]

**Rate equation**

\[ r1 = \frac{k_1}{K_{d1}} \cdot [O_R] \cdot [R_2] - k_1^- \cdot [O_R \cdot R_2] \]
\[ r2 = \frac{k_7^-}{K_{d7}} \cdot [O_R \cdot R_2] \cdot [R_2] - k_7^- \cdot [O_R \cdot 2R_2] \]

**Parameters**

The parameters of this reaction are the dissociation constant for binding of ScbR to O\(R\) \((K_{d1} \text{ and } K_{d7})\) and the dissociation rate for binding of ScbR to O\(R\) \((k_1^- \text{ and } k_7^-)\). \(K_{d1}\) and \(K_{d7}\) will have the same distributions, as they refer to a similar binding process and by taking uncertainty into account, the range of parameter values can be considered the same. The same applies for \(k_1^-\) and \(k_7^-\). \(scbR\) is a member of the TetR family of repressors, named after the member of this group which is the most completely characterized, the TetR repressor protein. TetR binds to the operator tetO, repressing its own expression and that of the efflux determinant tetA in a similar way as ScbR binding to O\(R\) and O\(A\) and repressing its own expression and the expression of ScbA. Therefore parameter values were derived from published data on the TetR–tetO interaction and on tetR-like proteins binding to their corresponding operators.
Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models [2] [3]</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d1}$</td>
<td>0.005 – 5.8</td>
<td>nM</td>
<td>8.82 nM (Range tested: $10^{-5} – 10 nM$)</td>
<td>An early publication based on stopped-flow measurements at various salt concentrations reports a $K_A$ of $2 \cdot 10^{11} M^{-1}$ [4], therefore a $K_d \left( \frac{1}{K_A} \right)$ of 0.005 nM. The Tet repressor and TetO operator were derived by using an overproducing E. coli strain.</td>
</tr>
<tr>
<td>$K_{d2}$</td>
<td></td>
<td></td>
<td>(Bistability range: 5.80 – 9.5 nM)</td>
<td>Final results should be discussed with respect to the conditions in vivo. The salt concentration in a bacterial cell is estimated to be 160 mM NaCl and 3 mM MgCl$<em>2$. When the contribution of Mg$^{2+}$ is neglected, the binding constants for the Tet repressor are: $K</em>{d1}^{(160 \text{ mM} \ \text{NaCl})} = 2 \times 10^{11} \text{ M}^{-1}$, $K_{d2}^{(160 \text{ mM} \ \text{NaCl})} = 1 \times 10^8 \text{ M}^{-1}$. The ratio of the specific to the nonspecific binding constant ($7 \times 10^9$) is even slightly higher than the corresponding ratio obtained for the lac repressor ($3 \times 10^9$). These specificity differences were observed in Kleinschmidt et al. 1988 [4].</td>
</tr>
</tbody>
</table>

Data derived from equilibrium SPR analysis of TetR-tetO interaction report a $K_d$ of $5.6 \pm 2 nM^{-1}$[5], therefore a $K_d \left( \frac{1}{K_A} \right)$ of $0.179 \pm 0.064$ nM. E. coli was used as a host for the expression of proteins and the experiments were conducted with synthetic tetO-containing fragments.

Table 1. Affinity constants for TetR and revTetR

<table>
<thead>
<tr>
<th>Equilibrium</th>
<th>Binding constant</th>
<th>TetR</th>
<th>RevTetR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TetR + [αMg$^{2+}$] $\rightleftharpoons$ Tetr [αMg$^{2+}$]</td>
<td>$K_t$ [M$^{-1}$]</td>
<td>$1.26 \pm 0.36 \times 10^8$ (7)</td>
<td>$1.38 \pm 0.38 \times 10^8$</td>
</tr>
<tr>
<td>TetR + $\alpha$CcpA $\rightleftharpoons$ Tetr [αCcpA]</td>
<td>$K_{d1}$ [M$^{-1}$]</td>
<td>$5.6 \pm 2 \times 10^8$</td>
<td>$K_{d1} = 4 \pm 1 \times 10^8$</td>
</tr>
<tr>
<td>TetR + [αMg$^{2+}$] $\rightleftharpoons$ Tetr [αMg$^{2+}$] + αCcpA</td>
<td>$K_{d2}$ [M$^{-1}$]</td>
<td>$4.2 \pm 0.2 \times 10^6$</td>
<td>$K_{d2} = 1 \pm 0.2 \times 10^6$</td>
</tr>
</tbody>
</table>

$K_{d1}$: Mg$^{2+}$-dependent binding for TetR and revTetR. Constants for TetR and revTetR binding to αCcpA. $K_{d1}$: effector-binding pockets are occupied, $K_{d2}$: free effector-binding pockets.

Kamionka et al. 2004 [5]

Additionally, a study on the TetR-like protein Rv3066 binding to the mmr operon in M. tuberculosis suggests a $K_d$ of $4.4 \pm 0.3 nM$.[6]

Preumably, Rv3066 suppresses the expression of the Mmr multidrug efflux pump by directly binding to its target DNA. Fluorescence polarization-based assay was carried out to study the interaction between Rv3066 and the 30-bp DNA containing the IR1 sequence. Figure 7T illustrates the binding isotherms of Rv3066 in the presence of 1 nM fluorophore-labeled DNA. The titration experiment indicated that this regulator binds the 30-bp IR1 operator with a dissociation constant, $K_d$ of $4.4 \pm 0.3$ nM. This value is similar to that of the QcrR regulator where it binds DNA with the $K_d$ of 5.7 nM (42). The binding data also indicate that Rv3066 binds its cognate DNA with a stoichiometry of one Rv3066 dimer per IR1.

Bolla et al. 2012 [6]

Finally, in vitro studies on TetR-like protein ActR in S. coelicolor suggest a $K_d$ in the range of $0.1 – 5.8 nM$.[7]

Ahn et al. 2007 [7]
Appendix D. Parameter information for the GBL system model

Parameters with uncertainty

When deciding how to describe the uncertainty for each parameter there are a few points to be taken into consideration. Firstly, the values reported in literature are spread in a relatively large range and correspond to different protein–operator binding reactions, although all are part of the same family of repressors (TetR). Additionally, the values were acquired by in vitro testing, although in some of the publications a strong correlation between in vitro and in vivo measurements is noted. This means that there might be a notable difference between actual parameter values and the ones reported in literature. These facts influence the weight assigned to each parameter value, the quantification of the parameter uncertainty and therefore the shape of the corresponding distributions.

<table>
<thead>
<tr>
<th>$k_1$</th>
<th>$k_7$</th>
<th>$0.09 - 0.846$ [4] [8]</th>
<th>$min^{-1}$</th>
<th>N/A</th>
</tr>
</thead>
</table>

According to the study by Kleinschmidt et al. [4] mentioned above, the maximal association rate constant was $k_a = 3 \cdot 10^8 M^{-1}s^{-1}$. By taking into account the equilibrium association constant reported above ($K_A = 2 \cdot 10^{13}M^{-1}$), the dissociation rate constant can be calculated as per $k_1 = \frac{k_a}{k_d} = \frac{3 \cdot 10^8 M^{-1}s^{-1}}{2 \cdot 10^{13}M^{-1}} = 0.0015 s^{-1} = 0.09 min^{-1}$.

Additionally, a study on a TetR-like protein (RolR) which binds to its operator rolO and blocks the transcription of rolHMD and of its own gene, in Corynebacterium glutamicum (Gram positive and GC content ~ 50-60% bacteria) reports the dissociation rates $1.41 \cdot 10^{-2} s^{-1}$ ($0.846 min^{-1}$) and $7.34 \cdot 10^{-3} s^{-1}$ ($0.44 min^{-1}$) [8] measured by surface plasmon resonance (SPR).

As less information was available for $k_1$, most of which was acquired by in vitro measurements and none was derived from Streptomyces, the uncertainty associated with this parameter is larger and should be taken into account while designing the corresponding distributions.
Therefore, by assigning the appropriate weights to the parameter values and using the protocol described in Chapter 3, the appropriate probability distributions were designed.

More specifically, with regards to the $K_d$, the value that is considered as the most accurate in different publications is $0.005\,nm$ measured by Kleinschmidt et al. However, the $K_d$s measured for the ActR protein in S. coelicolor cover the range of $0.1 - 5.8\,nm$. Therefore, the mode of the log-normal distribution calculated for the $K_d$ is $0.299\,nm$ which is within the range of the values reported for S. coelicolor and the Spread is 6.8 in order to be able to explore a wide range of values and thus take into account the uncertainty caused by both the in vitro testing and the measurements in different species and proteins. Thus, the range where 68.27% of the values are found is between 0.044 and 2.048 nM.

Similarly, the value calculated for the $k_1$ by the measurements of Kleinschmidt et al. is $0.09\,min^{-1}$; however other studies reported values of up to $0.85\,min^{-1}$. After considering all the factors and assigning the relevant weights to the parameter values, the mode of the log-normal distribution calculated for $k_1$ was $0.489\,min^{-1}$ and the Spread was 1.98. This means that the range where 68.27% of the values are found is between 0.247 and 0.968 min$^{-1}$.

Since the two parameters are interdependent, thermodynamic consistency also needs to be taken into account. This is achieved by creating a bivariate system as described in Chapter 3. Since $k_{on1}$ is the parameter with the largest geometric coefficient of variation, as only one reported value was retrieved from literature, this is set as the dependent parameter as per: $k_{on1} = \frac{k_1}{K_d}$. The location and scale parameters of $k_{on1}$ ($\mu=-0.53113$ and $\sigma=1.3034$) were calculated from those of $K_d$ and $k_1$.

The probability distributions (plotted on a log-scale) for the three parameters are the following:
The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines). The correlation matrix which is necessary to define the relationship between the two marginal distributions \(k_{on1}, k^-_1\) of the bivariate system is derived by employing random values generated by the two distributions.

The parameter information of the distributions of the multivariate system is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>(\mu)</th>
<th>(\sigma)</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_{d1})</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
<td></td>
</tr>
</tbody>
</table>
| \(k^-_1\) | 0.489 | 1.98   | \(-0.37642\) | 0.58225     | \[
\begin{pmatrix}
1 & 0.3045 \\
0.3045 & 1
\end{pmatrix}
\]|
| \(k_{on1}\) | N/A | N/A    | \(-0.53113\) | 1.3034      |                   |

The multivariate system of the normal distributions \((ln(k^-_1)\) and \((ln(k_{on1}))\) and the resulting samples of values are presented in the following figure:

![Bivariate system of parameters](image)
In this way, a system of distributions is created where each distribution is described and constrained by the other two. Therefore, the parameters will be sampled by the two marginal distributions in a way consistent with our beliefs and with the relevant thermodynamic constraints. However, since the model’s reaction rate requires the parameters $K_{d1}$ and $k_{r1}$, and not the $k_{on1}$, the value for $K_{d1}$ is calculated by the parameters sampled from the other two distributions in an additional step, as per $k_{d1} = \frac{k_{r1}}{k_{on1}}$. All the above apply as well for the parameters $K_{d7}$ and $k_{r7}$.

References


D.2 Binding of $R_2$ to $O_A$ operator

The ScbR homo-dimer ($R_2$) binds to the ScbA gene operator ($O_A$) and represses its mRNA transcription. As reported by Bhukya et al.\textsuperscript{[1]} two ScbR homo-dimers can bind to the promoter. When one homo-dimer is bound, the mRNA transcription is already being repressed. As the concentration of ScbR rises, a second homo-dimer may bind to the already suppressed promoter and further enhance the suppression of the transcription. Therefore, this behaviour is represented by two separate reactions that may take place in different times.

**Chemical equation**

\[
O_A + R_2 \rightleftharpoons O_A \cdot R_2 \\
O_A \cdot R_2 + R_2 \rightleftharpoons O_A \cdot 2R_2
\]

**Rate equation**

\[
r_1 = \frac{k_2^{-}}{K_{d2}} \cdot [O_R] \cdot [R_2] - k_2^{-} \cdot [O_R \cdot R_2] \\
r_2 = \frac{k_8^{-}}{K_{d8}} \cdot [O_R \cdot R_2] \cdot [R_2] - k_8^{-} \cdot [O_R \cdot 2R_2]
\]

**Parameters**

The parameters of this reaction are the dissociation constant for binding of ScbR to $O_A$ ($K_{d2}$ and $K_{d8}$) and the dissociation rate for binding of ScbR to $O_A$ ($k_2^{-}$ and $k_8^{-}$). $K_{d2}$ and $K_{d8}$ will have the same distributions, as they refer to a similar binding process and by taking uncertainty into account, the range of parameter values can be considered the same. The same applies for $k_2^{-}$ and $k_8^{-}$. \textit{scbR} is a member of the TetR family of repressors, named after the member of this group which is the most completely characterized, the TetR repressor protein. TetR binds to the operator tetO, repressing its own expression and that of the efflux determinant tetA in a similar way as ScbR binding to $O_R$ and $O_A$ and repressing its own expression and the expression of ScbA. Therefore parameter values were derived from published data on the TetR–tetO interaction and on tetR-like proteins binding to their corresponding operators.
<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Remarks/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d3}$</td>
<td>0.005 - 5.8</td>
<td>nM</td>
<td>An early publication based on stopped-flow measurements at various salt concentrations reports a $K_A$ of $2 \times 10^{-13} M^{-1}$ [4], therefore a $K_d (\frac{1}{K_A})$ of 0.005nM. The Tet repressor and TetO operator were derived by using an overproducing <em>E. coli</em> strain.</td>
</tr>
<tr>
<td>$K_{d8}$</td>
<td>0.005 - 5.8</td>
<td>nM</td>
<td>(Range tested: $10^{-5} - 10nM$) (Bistability range: 7.30 - 8.30nM)</td>
</tr>
</tbody>
</table>

Data derived from equilibrium SPR analysis of TetR-tetO interaction report a $K_A$ of $5.6 \pm 2nM^{-1}$[5], therefore a $K_d (\frac{1}{K_A})$ of $0.179 \pm 0.064nM$. *E. coli* was used as a host for the expression of proteins and the experiments were conducted with synthetic tetO-containing fragments.

| Table 1. Affinity constants for TetR and revTetR |
|---|---|---|---|
| Equilibrium | Binding constant | TetR | RevTetR |
| TetR + tetO | $K_{d3} (nM)$ | $1.26 \pm 0.36 \times 10^2$ (7) | $1.38 \pm 0.38 \times 10^2$ |
| TetR + revTetR | $K_{d8} (nM)$ | $5.6 \pm 2 \times 10^2$ | $2 \pm 0 \times 10^2$ |

Kamionka et al. 2004 [5]

Additionally, a study on the TetR-like protein Rv3066 binding to the mmr operon in *M. tuberculosis* suggests a $K_d$ of $4.4 \pm 0.3nM$ [6].

Presumably, Rv3066 suppresses the expression of the Mmr multidrug efflux pump by directly binding to its target DNA. Fluorescence polarization-based assay was carried out to study the interaction between Rv3066 and the 36-bp DNA containing the IR1 sequence. Figure 7a illustrates the binding isotherms of Rv3066 in the presence of 1nM fluorescein-labeled DNA. The titration experiment indicated that this regulator binds the 36-bp IR1 operator with a dissociation constant, $K_d$, of $4.4 \pm 0.3nM$. This value is similar to that of the QsrR regulator where it binds DNA with the $K_d$ of 5.7nM (42). The binding data also indicate that Rv3066 binds its cognate DNA with a stoichiometry of one Rv3066 dimer per IR1.

Bolla et al. 2012 [6]

Finally, *in vitro* studies on TetR-like protein ActR in *S. coelicolor* suggest a $K_d$ in the range of $0.1 - 5.8nM$ [7].

protein (Hill's coefficient of $>3$) to its three binding sites. The interaction of ActR with the actI and ctcB intergenic sequence was also characterized by strong binding ($K_d$ of 0.1 nM, 0.3 nM, and 5.8 nM), but with a Hill's coefficient of $<1$ (data not shown), there was no evidence of cooperativity.

Ahn et al. 2007 [7]
Parameters with uncertainty

When deciding how to describe the uncertainty for each parameter there are a few points to be taken into consideration. Firstly, the values reported in literature are spread in a relatively large range and correspond to different protein–operator binding reactions, although all are part of the same family of repressors (TetR). Additionally, the values were acquired by in vitro testing, although in some of the publications a strong correlation between in vitro and in vivo measurements is noted. This means that there might be a notable difference between actual parameter values and the ones reported in literature. These facts influence the weight assigned to each parameter value, the quantification of the parameter uncertainty and therefore the shape of the corresponding distributions.
Appendix D. Parameter information for the GBL system model

Therefore, by assigning the appropriate weights to the parameter values and using the protocol described in Chapter 3, the appropriate probability distributions were designed.

More specifically, with regards to the $K_{d2}$, the value that is considered as the most accurate in different publications is 0.005$nM$ measured by Kleinschmidt et al. However, the $K_D$s measured for the ActR protein in *S. coelicolor* cover the range of $0.1 - 5.8nM$. Therefore, the mode of the log-normal distribution calculated for the $K_{d1}$ is 0.299$nM$ which is within the range of the values reported for *S. coelicolor* and the Spread is 6.8 in order to be able to explore a wide range of values and thus take into account the uncertainty caused by both the *in vitro* testing and the measurements in different species and proteins. Thus, the range where 68.27% of the values are found is between 0.044 and 2.048 nM.

Similarly, the value calculated for the $k_{2-}$ by the measurements of Kleinschmidt et al. is $0.09min^{-1}$; however other studies reported values of up to $0.85min^{-1}$. After considering all the factors and assigning the relevant weights to the parameter values, the mode of the log-normal distribution calculated for $k_{2-}$ was $0.489min^{-1}$ and the Spread was 1.98. This means that the range where 68.27% of the values are found is between 0.247 and 0.968 $min^{-1}$.

Since the two parameters are interdependent, thermodynamic consistency also needs to be taken into account. This is achieved by creating a bivariate system as described in Chapter 3. Since $k_{on2}$ is the parameter with the largest geometric coefficient of variation, as only one reported value was retrieved from literature, this is set as the dependent parameter as per: $k_{on2} = \frac{k_2}{K_{d2}}$. The location and scale parameters of $k_{on2}$ ($\mu=-0.53113$ and $\sigma=1.3034$) were calculated from those of $K_{d2}$ and $k_{2-}$.

The probability distributions (plotted on a log-scale) for the three parameters are the following:
Appendix D. Parameter information for the GBL system model

Distribution for $K_{d2}$ values

Distribution for $k_2^*$ values

Distribution for $k_{on2}$ values
Appendix D. Parameter information for the GBL system model

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines). The correlation matrix which is necessary to define the relationship between the two marginal distributions \((k_{on2}, k_2)\) of the bivariate system is derived by employing random values generated by the two distributions.

The parameter information of the distributions of the multivariate system is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>(\mu)</th>
<th>(\sigma)</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_{d2})</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
<td></td>
</tr>
</tbody>
</table>
| \(k_2^-\) | 0.489 | 1.98   | -0.37642| 0.58225| \[
\begin{pmatrix}
1 & 0.3045 \\
0.3045 & 1
\end{pmatrix}
\] |
| \(k_{on2}\) | N/A  | N/A    | -0.53113| 1.3034 |

The multivariate system of the normal distributions \((ln(k_2^-)\) and \(ln(k_{on2})\)\) and the resulting samples of values are presented in the following figure:
Appendix D. Parameter information for the GBL system model

In this way, a system of distributions is created where each distribution is described and constrained by the other two. Therefore, the parameters will be sampled by the two marginal distributions in a way consistent with our beliefs and with the relevant thermodynamic constraints. However, since the model’s reaction rate requires the parameters $K_{d2}$ and $k_{2}^{-}$, and not the $k_{on2}$, the value for $K_{d2}$ is calculated by the parameters sampled from the other two distributions in an additional step, as per $k_{d2} = \frac{k_{2}^{-}}{k_{on2}}$. All the above apply as well for the parameters $K_{d8}$ and $k_{8}^{-}$.

References


Appendix D. Parameter information for the GBL system model

D.3 Binding of R\textsubscript{2} to A

The ScbR homo-dimer (R\textsubscript{2}) forms a complex with ScbA (A).

### Chemical equation

\[ A + R_2 \rightleftharpoons AR_2 \]

### Rate equation

\[ r = \frac{k_3^-}{K_{d3}} \cdot [A] \cdot [R_2] - k_3^- \cdot [AR_2] \]

### Parameters

The parameters of this reaction are the dissociation constant for binding of ScbR to ScbA (\(K_{d3}\)) and the dissociation rate for binding of ScbR to ScbA (\(k_3^-\)). Since there is no concrete evidence of the existence of the ScbA-ScbR complex so far, it is possible that the interaction between the two proteins is unstable/ transient and therefore the parameter values reflect this belief. The values of such complexes according to the literature \[1\], lie in the millimolar or micromolar scale.

(a) Ozbabacan et al. 2011 \[1\]

(b) Perkins et al. 2010 \[2\]
Appendix D. Parameter information for the GBL system model

Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL model</th>
<th>Remarks/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d3}$</td>
<td>$10^4 - 10^6$ [1]</td>
<td>nM</td>
<td>$0.083 nM^{-1}s^{-1}$</td>
<td>According to the Ozbabacan et al. association constants for transient protein-protein interactions lie in the millimolar or micromolar range. $(10^6(10^4) - 10^6 nM)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(4.98 nM^{-1}min^{-1})$</td>
<td>affinity with a low ratio of bound to free form. Thus, the equilibrium dissociation constants of the transient complexes may be observed in the range of millimolar $(10^{-3} M)$ to micromolar $(10^{-6} M)$, as their constituents associate and dissociate rapidly, whereas the constants of permanent complexes may be in the range of micromolar to femtomolar $(10^{-12} M)$ (Wallis et al., 1995). The knowledge of strong Ozbabacan et al. 201[1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$10^{-7} - 10^{-1} nM^{-1} s^{-1}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(6 \cdot 10^{-6} - 6nM^{-1}min^{-1})$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bistability range: $0.083 - 0.12nM^{-1} s^{-1}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(4.98 - 7.2nM^{-1}min^{-1})$</td>
<td></td>
</tr>
<tr>
<td>$k_3$</td>
<td>$38.3 - 4.46 \cdot 10^6$ [4][5]</td>
<td>min$^{-1}$</td>
<td>$630s^{-1}(37800min^{-1})$</td>
<td>According to Northrup et al. the association rate of protein-protein bond formations occurs in the order of $10^6 M^{-1}s^{-1}$.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$0 - 10^3 s^{-1}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(0 - 6 \cdot 10^4 min^{-1})$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bistability range: $460 - 630s^{-1}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$(27600 - 37800min^{-1})$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameters with uncertainty

Since the values we are using for this parameter correspond to generic association constant values of a wide range of protein-protein interactions and not specifically to GBL
or related systems, we wish to explore the whole range of values and investigate the conditions under which the ScbR-ScbA complex formation would be feasible. Therefore, by assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distributions were designed. The mode of the log-normal distribution for $K_{d3}$ is $9.95 \cdot 10^4 nM$ and the Spread to 10. Thus, the range where 68.27% of the values are found is between $9.93 \cdot 10^3 nM$ and $9.97 \cdot 10^5 nM$.

Similarly, by following the same reasoning, the mode of the log-normal distribution for $k_{on3}$ is set to $0.018 nM^{-1} min^{-1}$ and the Spread to 3.17. This means that the range where 68.27% of the values are found is between $0.0059 nM^{-1} min^{-1}$ and $0.06 nM^{-1} min^{-1}$.

Since the two parameters are interdependent, thermodynamic consistency also needs to be taken into account. This is achieved by creating a bivariate system as described in Chapter 3. Since no information was retrieved for $k_3^-$ and therefore is the parameter with the largest geometric coefficient of variation, this is set as the dependent parameter as per: $k_3^- = k_{on3} \cdot K_{d3}$. The location and scale parameters of $k_3^-$ ($\mu=9.9423$ and $\sigma=1.5503$) were calculated from those of $K_{d3}$ and $k_{on3}$.

The probability distributions for the three parameters, adjusted accordingly in order to reflect the above values, are the following:
The correlation matrix which is necessary to define the relationship between the two marginal distributions \((k_{d3}, k_{3}^-)\) of the bivariate system is derived by employing random values generated by the two distributions.

The parameter information of the distributions of the multivariate system is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>(\mu)</th>
<th>(\sigma)</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k_{on3})</td>
<td>0.0188</td>
<td>3.17</td>
<td>-3.2497</td>
<td>0.84825</td>
<td>N/A</td>
</tr>
</tbody>
</table>
| \(K_{d3}\) | 99547 | 10.02 | 13.192 | 1.2977 | \[
\begin{pmatrix}
1 & 0.6942 \\
0.6942 & 1
\end{pmatrix}
\] |
| \(k_{3}^-\) | N/A | N/A | 9.9423 | 1.5503 | |

The distribution for \(k_{3}^-\) values and \(k_{on3}\) values are shown in the images above.
The multivariate system of the normal distributions \((\ln(k^{-3})\) and \(\ln(k_{\text{on}})\)) and the resulting samples of values are presented in the following figure:

In this way, a system of distributions is created where each distribution is described and constrained by the other two. Therefore, the parameters will be sampled by the two marginal distributions in a way consistent with our beliefs and with the relevant thermodynamic constraints.

D.4 Binding of $R_2$ to C

SCBs (C) bind to ScbR homo-dimer ($R_2$) and inactivate its repression of the $scbR$ and $scbA$ promoters.

**Chemical equation**

The exact mechanism is still unclear, however in our model we assumed that two SCBs bind to the ScbR homo-dimer.

$$2C + R_2 \rightleftharpoons C_2 \cdot R_2$$

**Rate equation**

$$r = \frac{k_4^{-}}{K_{d4}} \cdot [C]^2 \cdot [R_2] - k_4^{-} \cdot [C_2 \cdot R_2]$$

**Parameters**

The parameters of this reaction are the dissociation constant for binding of SCB to ScbR ($K_{d4}$) and the dissociation rate for binding of SCB to ScbR ($k_4^{-}$). ScbR is a member of the TetR family of repressors, named after the member of this group which is the most completely characterized, the TetR repressor protein. TetR binds to the operator tetO, repressing its own expression and that of the efflux determinant tetA. However, $[\text{MgTc}]^+$ binds to TetR and thus the affinity of the later for the operator tetO is 9-fold reduced. This procedure is similar to ScbR binding to $O_R$ and $O_A$ and repressing its own expression and the expression of ScbA, while binding to SCBs reduces its affinity for the two operators. Therefore parameter values were derived from published data on the TetR–$[\text{MgTc}]^+$ and TetR–Tc (without Mg$^+$) interactions.
Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models [1] [2]</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d4}$</td>
<td>0.0011 – 10400 [3] [4] [5] [6]</td>
<td>nM</td>
<td>0.083 nM$^{-1}$s$^{-1}$ (4.98 nM$^{-1}$min$^{-1}$)</td>
<td>According to Hillen et al. and Orth et al. the association constant for TetR-Tc binding is in the $\sim 10^2$M$^{-1}$ range in presence of Mg$^{2+}$, as determined by in vitro measurements, therefore a $K_d = 1 \times 10^{-4}$M = 1 nM</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Range tested: $10^{-7} – 0.1 nM^{-1}s^{-1}$ (6 $\times$ 10$^{-6}$ – 6nM$^{-1}$min$^{-1}$)</td>
<td>Ortli et al. 2011[4]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bistability range: 0.077 – 0.17 nM$^{-1}$s$^{-1}$[1]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(4.62 – 10.2nM$^{-1}$min$^{-1}$) and 0.0042 – 0.253nM$^{-1}$s$^{-1}$[2]</td>
<td>On the other hand, Kedrakca-Krok et al. conducted stopped-flow measurements using TetR overproduced in Escherichia coli strain RB 791. They consequently reported an association constant $K_d$=0.96 $\times$ 10$^6$M$^{-1}$ in absence of Mg$^{2+}$ (therefore a $K_d$=1.04 $\times$ 10$^{-5}$M = 1.04 $\times$ nM) and an association constant $K_d$=6.3 $\times$ 10$^6$M$^{-1}$ for binding of TetR to [Tc-Mg]$^+$ (therefore a $K_d$=0.16 $\times$ 10$^{-6}$M = 160nM).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.252 – 15.18nM$^{-1}$min$^{-1}$)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Rate Constants for Association and Dissociation of TetR and Tetracycline in the Absence of Magnesium Ions. Obtained in the Experiment Mimicking the Changes in Fluorescence Emission of TetR(Tc) and the Disappearance of Tetracycline Emission in Photoreaction of Tetracycline.  

<table>
<thead>
<tr>
<th>TetR or Tet(Tc) emission</th>
<th>$k_{on}$ (nM$^{-1}$s$^{-1}$)</th>
<th>$k_{off}$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TetR(Tc) emission</td>
<td>$2.2 \pm 0.92$</td>
<td>$1.99 \pm 0.39$</td>
</tr>
<tr>
<td>$\Delta A_{450}$</td>
<td>$2.860 \pm 0.002$</td>
<td>$2.130 \pm 0.004$</td>
</tr>
<tr>
<td>Tet(Tc) emission</td>
<td>$1.72 \pm 0.36$</td>
<td>$1.42 \pm 0.76$</td>
</tr>
<tr>
<td>$\Delta A_{330}$</td>
<td>$1.72 \pm 0.36$</td>
<td>$1.42 \pm 0.76$</td>
</tr>
<tr>
<td>$\Delta A_{390}$</td>
<td>$1.38 \pm 0.08$</td>
<td>$1.04 \pm 0.65$</td>
</tr>
<tr>
<td>Tet(Tc) emission</td>
<td>$1.38 \pm 0.08$</td>
<td>$1.04 \pm 0.65$</td>
</tr>
<tr>
<td>$\Delta A_{390}$</td>
<td>$1.77 \pm 0.33$</td>
<td>$1.42 \pm 0.76$</td>
</tr>
</tbody>
</table>

[1] The measurements were performed in TetR, pH 6.0, at 23°C  
[2] The values of association constant obtained by analysis of transient emission under photochemical conditions. The values of association rate constants obtained by analysis of transients obtained from the Orli program.  

Kedrakca-Krok et al. 2005[3]  

The N-terminal domain and the C-terminal inducer binding domain. Binding of tetacycline (Tc) to TetR in the absence of Mg$^{2+}$ can be described by a simple kinetic model, which is limited to the first step-association without any unidirectional conformational change upon Tc binding. The rate constants for the process are equal to $2.0 \times 10^6$M$^{-1}$s$^{-1}$ and 2.0 s$^{-1}$ for the forward and backward processes, respectively. hillen et al. 2010

Kedrakca-Krok et al. 2005[3]  

Finally, Schubert et al. reported Mg$^{2+}$-independent $K_d$ of $1.95 \pm 0.11 \times 10^7$ M$^{-1}$ and $10 \pm 0.31 \times 10^4$ M$^{-1}$ ($K_d$ = 51.28 $\pm$ 2.89 nM and $K_d$ = 10 $\pm$ 0.31 nM) and Mg$^{2+}$ dependent $K_d$ of $2.3 \pm 0.21 \times 10^1$ M$^{-1}$ and $9.11 \pm 0.3 \times 10^1$ M$^{-1}$ ($K_d$ = 0.00435 $\pm$ 0.0004 nM and $K_d$ = 0.00111 $\pm$ 0.000036 nM). The data was derived from in vitro and in vivo measurements in E. coli K12, strains DH5a and WH207.

Table 5. Mg$^{2+}$-independent (a) and Mg$^{2+}$-dependent (b) binding constants, association and dissociation dissociation of $K_d$-values.  

<table>
<thead>
<tr>
<th>$K_d$ (M$^{-1}$)</th>
<th>$k_{on}$ (nM$^{-1}$s$^{-1}$)</th>
<th>$k_{off}$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TetR(Tc) emission</td>
<td>$1.95 \pm 0.11$</td>
<td>$10 \pm 0.31$</td>
</tr>
<tr>
<td>$\Delta A_{450}$</td>
<td>$2.860 \pm 0.002$</td>
<td>$2.130 \pm 0.004$</td>
</tr>
<tr>
<td>Tet(Tc) emission</td>
<td>$1.72 \pm 0.36$</td>
<td>$1.42 \pm 0.76$</td>
</tr>
<tr>
<td>$\Delta A_{330}$</td>
<td>$1.72 \pm 0.36$</td>
<td>$1.42 \pm 0.76$</td>
</tr>
<tr>
<td>$\Delta A_{390}$</td>
<td>$1.38 \pm 0.08$</td>
<td>$1.04 \pm 0.65$</td>
</tr>
<tr>
<td>Tet(Tc) emission</td>
<td>$1.38 \pm 0.08$</td>
<td>$1.04 \pm 0.65$</td>
</tr>
<tr>
<td>$\Delta A_{390}$</td>
<td>$1.77 \pm 0.33$</td>
<td>$1.42 \pm 0.76$</td>
</tr>
</tbody>
</table>

*Calculated as $k_{on} = k_{off}$.  

Schubert et al. 2004[6]
Parameters with uncertainty

When deciding how to describe the uncertainty for each parameter there are a few points to be taken into consideration. Firstly, the values reported in literature are spread in a relatively large range and correspond to TetR and TetR mutant proteins. Additionally, most of the values were acquired by “in vitro” testing. This means that there might be a notable difference between actual parameter values and the ones reported in literature. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distributions. By keeping all these factors into account and by assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distributions were designed.

With regards to the $K_{d4}$ the value that is mostly reported in different publications is $1nM$, therefore we put the weight of the distribution in the range $1 - 10nM$ and we consider as least likely the larger values. Therefore, the mode of the log-normal distribution is $1nM$ and the Spread is 318.77. Thus the range where 68.27% of the values are found is between $3 \cdot 10^{-3}$ and $317.5nM$. (Note: This distribution represents our initial beliefs about the system but is redefined as explained below, in order to account for the thermodynamic consistency of the system.)
The smaller values reported by Schubert et al. \((10^{-4}\text{min}^{-1})\) are also considered least likely for \(k_4^-\). The mode of the log-normal distribution for \(k_4^-\) is set as \(0.816\text{min}^{-1}\) and the Spread is 244. In this way the range where 68.27% of the values are found is between \(3.3 \cdot 10^{-3}\) and \(199.2\ \text{min}^{-1}\).

Finally, the probability distribution for \(k_{on4}\) is defined accordingly, in order to allow the exploration of the full range of the values retrieved from literature. Therefore, the mode is set to \(0.071\text{nM}^{-1}\text{min}^{-1}\) and the Spread is 4.4. In this way the range where 68.27% of the values are found is between 0.016 and 0.313 \(\text{nM}^{-1}\text{min}^{-1}\).

Since the three parameters are interdependent, thermodynamic consistency also needs to be taken into account. This is achieved by creating a bivariate system as described in Chapter 3. Since \(K_{d4}\) is the parameter with the largest geometric coefficient of variation, this is set as the dependent parameter as per: \(K_{d4} = \frac{k_4^-}{k_{on4}}\), and an updated probability distribution is defined. The location and scale parameters of \(K_{d4}\) \((\mu=5.9389\) and \(\sigma=2.3412)\) were calculated from those of \(k_{on4}\) and \(k_4^-\). The probability distributions for the three parameters, adjusted accordingly in order to reflect the above values, are the following:

![Distribution for \(K_{d4}\) values](image)
The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines). The correlation matrix which is necessary to define the relationship between the two marginal distributions \( (K_{d4}, k^-_4) \) of the bivariate system is derived by employing random values generated by the two distributions.

The parameter information of the distributions of the multivariate system is:
Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{on4}$</td>
<td>0.071</td>
<td>4.4</td>
<td>−1.6517</td>
<td>0.99557</td>
<td>N/A</td>
</tr>
<tr>
<td>$K_{d4}$</td>
<td>N/A</td>
<td>N/A</td>
<td>5.9389</td>
<td>2.3412</td>
<td>$\begin{pmatrix} 1 &amp; 0.9326 \ 0.9326 &amp; 1 \end{pmatrix}$</td>
</tr>
<tr>
<td>$k_{4^-}$</td>
<td>0.816</td>
<td>244</td>
<td>4.2872</td>
<td>2.119</td>
<td></td>
</tr>
</tbody>
</table>

The multivariate system of the normal distributions ($\ln(k_{d4})$ and $\ln(k_{4^-})$) and the resulting samples of values are presented in the following figure:

![Bivariate system of parameters](image)

In this way, a system of distributions is created where each distribution is described and constrained by the other two. Therefore, the parameters will be sampled by the two marginal distributions in a way consistent with our beliefs and with the relevant thermodynamic constraints.

References


Appendix D. Parameter information for the GBL system model

D.5 Binding of AR2 to O_A' operator

The ScbA–ScbR complex (AR2) binds to the putative ScbA gene operator (O_A') and activates its maximum mRNA transcription. In the model, O_A and O_A' are considered together under the general name O_A which can be found at different states. When only AR2 is bound, maximum transcription is activated. When both AR2 and R2 or two R2's are bound, the promoter is repressed. The different statuses of the promoter are represented by the set of reactions below.

Chemical equation

\[ O_A + AR_2 \rightleftharpoons O_A \cdot AR_2 \]
\[ O_A \cdot R_2 + AR_2 \rightleftharpoons O_A \cdot R_2 \cdot AR_2 \]
\[ O_A \cdot 2R_2 + AR_2 \rightleftharpoons O_A \cdot 2R_2 \cdot AR_2 \]
\[ O_A \cdot AR_2 + R_2 \rightleftharpoons O_A \cdot R_2 \cdot AR_2 \]
\[ O_A \cdot R_2 \cdot AR_2 + R_2 \rightleftharpoons O_A \cdot 2R_2 \cdot AR_2 \]

Rate equation

\[ r1 = \frac{k_5^-}{K_{d5}} \cdot [O_A] \cdot [AR_2] - k_5^- \cdot [O_A \cdot AR_2] \]
\[ r2 = \frac{k_9^-}{K_{d9}} \cdot [O_A \cdot R_2] \cdot [AR_2] - k_9^- \cdot [O_A \cdot R_2 \cdot AR_2] \]
\[ r3 = \frac{k_{10}^-}{K_{d10}} \cdot [O_A \cdot 2R_2] \cdot [AR_2] - k_{10}^- \cdot [O_A \cdot 2R_2 \cdot AR_2] \]
\[ r4 = \frac{k_{11}^-}{K_{d11}} \cdot [O_A \cdot AR_2] \cdot [R_2] - k_{11}^- \cdot [O_A \cdot R_2 \cdot AR_2] \]
\[ r5 = \frac{k_{12}^-}{K_{d12}} \cdot [O_A \cdot R_2 \cdot AR_2] \cdot [R_2] - k_{12}^- \cdot [O_A \cdot 2R_2 \cdot AR_2] \]

Parameters

The parameters of the reaction of the AR2 complex binding to the putative promoter are the dissociation constant for binding of ScbA-ScbR to O_A' (K_d5) and the dissociation rate for binding of ScbA-ScbR to O_A' (k_5^-). The parameters for the other reactions will have the same distributions, as they refer to a similar binding process and by taking uncertainty into account, the range of parameter values can be considered the same.
Therefore, $K_{d9}$, $K_{d10}$, $K_{d11}$ and $K_{d12}$ will follow the same distribution as $K_{d5}$ and $k_9$, $k_{10}$, $k_{11}$ and $k_{12}$ will follow the same distribution as $k_5$.

Since there is no evidence of the actual existence of the ScbA–ScbR complex, we assumed that its binding affinity for the $O_A$ operator would follow a similar behaviour to the binding of ScbR to $O_R$. ScbR is a member of the TetR family of repressors, named after the member of this group which is the most completely characterized, the TetR repressor protein. TetR binds to the operator $tetO$, repressing its own expression and that of the efflux determinant $tetA$ in a similar way as ScbR binding to $O_R$ and $O_A$ and repressing its own expression and the expression of ScbA. Therefore parameter values were derived from published data on the TetR–tetO interaction and on tetR-like proteins binding to their corresponding operators.

**Parameters with uncertainty**

When deciding how to describe the uncertainty for each parameter there are a few points to be taken into consideration. Firstly, the values reported in literature are spread in a relatively large range and correspond to different protein-operator binding reactions, although all are part of the same family of repressors (TetR). Additionally, the values were acquired by in vitro testing, although in some of the publications a strong correlation between in vitro and in vivo measurements is noted. This means that there might be a notable difference between actual parameter values and the ones reported in literature. These facts influence the weight assigned to each parameter value, the quantification of the parameter uncertainty and therefore the shape of the corresponding distributions. Therefore, by assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distributions were designed.
Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Remarks/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d5}$</td>
<td>0.005 - 5.8</td>
<td>nM</td>
<td>4.35nM</td>
</tr>
<tr>
<td>$K_{d9}$</td>
<td></td>
<td></td>
<td>(Range tested: 10^-5 - 10nM)</td>
</tr>
<tr>
<td>$K_{d10}$</td>
<td></td>
<td></td>
<td>(Bistability range: 3.86 - 4.39nM)</td>
</tr>
<tr>
<td>$K_{d11}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{d12}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An early publication based on stopped-flow measurements at various salt concentrations reports a $K_A$ of $2 \cdot 10^{13} M^{-1}$ [4], therefore a $K_d \left( \frac{1}{K_A} \right)$ of 0.005nM. The Tet repressor and TetO operator were derived by using an overproducing *E. coli* strain.

Finally, our results should be discussed with respect to the conditions in vivo. The salt concentration in a bacterial cell is estimated to be 160 nM NaCl and 3 mM MgCl₂. When the contribution of Mg²⁺ is neglected, the binding constants for the Tet repressor are:

\[
K_{rP}(160 \text{ mM NaCl}) = 1 \times 10^{11} M^{-1}
\]

\[
K_{rO}(160 \text{ mM NaCl}) = 3 \times 10^{5} M^{-1}
\]

The ratio of the specific to the nonspecific binding constant (7 x 10⁵) is even slightly higher than the corresponding ratio obtained for the lac repressor (3 x 10⁵). These specificity

Kleinschmidt et al. 1988 [4]

Data derived from equilibrium SPR analysis of TetR-tetO interaction report a $K_B$ of $5.6 \pm 2nM^{-1}$ [5], therefore a $K_d \left( \frac{1}{K_B} \right)$ of $0.179 \pm 0.064nM$. *E. coli* was used as a host for the expression of proteins and the experiments were conducted with synthetic tetO-containing fragments.

<table>
<thead>
<tr>
<th>Table 1. Affinity constants for TetR and revTetR</th>
<th>Equilibrium</th>
<th>Binding constant</th>
<th>TetR</th>
<th>revTetR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d5}$ $K_{d9}$</td>
<td>$K_{d10}$</td>
<td>$K_{d11}$</td>
<td>$K_{d12}$</td>
<td></td>
</tr>
<tr>
<td>$K_{rP}$ $revTetR + TetR \rightarrow TetR + revTetR$</td>
<td>$K_{rP}$ $[M^{-1}]$</td>
<td>$5.6 \pm 2 \times 10^5$ (7)</td>
<td>$1.26 \pm 3.36 \times 10^5$ (7)</td>
<td></td>
</tr>
<tr>
<td>$K_{rO}$ $revTetR + TetO \rightarrow TetR + revTetR$</td>
<td>$K_{rO}$ $[M^{-1}]$</td>
<td>$4.2 \pm 0.2 \times 10^4$ (7)</td>
<td>$1.38 \pm 0.38 \times 10^4$ (7)</td>
<td></td>
</tr>
</tbody>
</table>

$K_{d5}$ Mg²⁺-dependent arc binding for TetR and revTetR. Constants for TetR and revTetR binding to tetO. $K_{d5}$ effecter binding pockets are occupied, $K_{d9}$ free effector binding pockets.

Kamionka et al. 2004 [5]

Additionally, a study on the TetR-like protein Rv3066 binding to the mnr operon in *M. tuberculosis* suggests a $K_d$ of $4.4 \pm 0.3nM$ [6].

Presumably, Rv3066 suppresses the expression of the Mnr multihap efflux pump by directly binding to its target DNA. Fluorescence polarization-based assay was carried out to study the interaction between Rv3066 and the 38-bp DNA containing the IR1 sequence. Figure 7a illustrates the binding isotherms of Rv3066 in the presence of 38-bp fluoresceinated DNA. The titration experiment indicated that this regulator binds the 38-bp IR1 operator with a dissociation constant, $K_d$ of $4.4 \pm 0.3nM$. This value is similar to that of the QnrR regulator where it binds DNA with the $K_d$ of 5.7 nM [12]. The binding data also indicate that Rv3066 binds its cognate DNA with a stoichiometry of one Rv3066 dimer per IR1.

Bolla et al. 2012 [6]

Finally, in vitro studies on TetR-like protein ActR in *S. coelicolor* suggest a $K_d$ in the range of $0.1 - 5.8nM$ [7].

| Protein (Hill's coefficient of >3) to its three binding sites. The interaction of ActR with the actCactB intergenic sequence was also characterized by strong binding ($K_a$ of 0.1 nM, 0.3 nM, and 5.8 nM), but with a Hill's coefficient of ~1 (data not shown), there was no evidence of cooperativity. |
|-------------------------------------------------|-------------|-----------------|------|---------|
| Ahn et al. 2007 [7]                              |-------------|-----------------|------|---------|
Appendix D. Parameter information for the GBL system model

More specifically, with regards to the $K_d$ parameter, the value that is considered as the most accurate in different publications is $0.005\text{mM}$ measured by Kleinschmidt et al. However, the $K_d$ s measured for the ActR protein in *S. coelicolor* cover the range of $0.1 - 5.8\text{mM}$. Therefore, the mode of the log-normal distribution calculated for the $K_d$ is $0.299\text{mM}$ which is within the range of the values reported for *S. coelicolor* and the Spread is 6.8 in order to be able to explore a wide range of values and thus take into account the uncertainty caused by both the *in vitro* testing and the measurements in different species and proteins. Thus, the range where 68.27% of the values are found is between 0.044 and 2.048 nM.

Similarly, the value calculated for $k_5^-$ by the measurements of Kleinschmidt et al. is $0.09\text{min}^{-1}$, however other studies reported values of up to $0.85\text{min}^{-1}$. After considering all the factors and assigning the relevant weights to the parameter values,
the mode of the log-normal distribution calculated for $k_5^-$ was $0.489 \text{min}^{-1}$ and the Spread was 1.98. This means that the range where 68.27% of the values are found is between 0.247 and 0.968 $\text{min}^{-1}$.

Since the two parameters are interdependent, thermodynamic consistency also needs to be taken into account. This is achieved by creating a bivariate system as described in Chapter 3. Since $k_{on5}$ is the parameter with the largest geometric coefficient of variation, as only one reported value was retrieved from literature, this is set as the dependent parameter as per: $k_{on5} = \frac{k_5^-}{K_{d5}}$. The location and scale parameters of $k_{on5}$ ($\mu=-0.53113$ and $\sigma=1.3034$) were calculated from those of $K_{d5}$ and $k_5^-$. The probability distributions for the three parameters, adjusted accordingly in order to reflect the above values, are the following:
The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines). The correlation matrix which is necessary to define the relationship between the two marginal distributions \((k_{on5}, k_{5})\) of the bivariate system is derived by employing random values generated by the two distributions.

The parameter information of the distributions of the multivariate system is:
Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d5}$</td>
<td>0.299</td>
<td>6.8</td>
<td>0.15471</td>
<td>1.1661</td>
<td>N/A</td>
</tr>
</tbody>
</table>
| $k_5^-$    | 0.489| 1.98   | -0.37642| 0.58225 | \[
\begin{pmatrix}
1 & 0.3045 \\
0.3045 & 1
\end{pmatrix}
\] |
| $k_{on5}$  | N/A  | N/A    | -0.53113| 1.3034  |                    |

The multivariate system of the normal distributions ($\ln(k_5^-)$ and $\ln(k_{on5})$) and the resulting samples of values are presented in the following figure:

In this way, a system of distributions is created where each distribution is described and constrained by the other two. Therefore, the parameters will be sampled by the two marginal distributions in a way consistent with our beliefs and with the relevant thermodynamic constraints. However, since the model’s reaction rate requires the parameters $K_{d5}$ and $k_5^-$, and not the $k_{on5}$, the value for $K_{d5}$ is calculated by the parameters sampled from the other two distributions in an additional step, as per $K_{d5} = \frac{k_5^-}{k_{on5}}$. All the above apply as well for the parameters $K_{d9}$, $K_{d10}$, $K_{d11}$, $K_{d12}$, $k_9^-$, $k_{10}^-$, $k_{11}^-$ and $k_{12}^-$. 
Appendix D. Parameter information for the GBL system model

References


Appendix D. Parameter information for the GBL system model

D.6 Synthesis of C

SCB (C) is synthesized by glycerol derivative and β-keto acid derivative precursors in a reaction catalyzed by ScbA (A).

Chemical equation

\[ A \overset{S}{\rightarrow} A + C \]

Rate equation

In our model we assume that there is an infinite amount of substrate S for the SCB synthesis and that the rate of the reaction is dependent only on the concentration of the catalyst A. Therefore, the rate of the reaction is:

\[ r = K_C \cdot [A] \]

Parameters

The parameter of this reaction is the synthesis rate of SCB \((K_C)\). The parameter values were chosen with the assumption that GBL production involves a one rate limiting step involving the Afsa-like protein and then a spontaneous cyclization.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_C)</td>
<td>0.00315 – 85.5</td>
<td>(\text{min}^{-1})</td>
<td>0.74s(^{-1}) (44A-4min(^{-1}))</td>
<td>(\text{Range tested:} 0 - 1.7s^{-1}[1]) (\text{Bistability range:} 0.67 - 1.7s^{-1}[1]) Addition of Garcia-Ojalvo et al. modelled repressors coupled by quorum sensing and reported a value for the synthesis rate of Acyl homoserine lactone (AHL) in E. coli of 0.01min(^{-1}).</td>
</tr>
</tbody>
</table>

Chen et al. calculated a synthesis rate of 0.189h\(^{-1}\) (0.00315min\(^{-1}\)) for the autoinducer PAI2 in *Pseudomonas aeruginosa* at stationary phase, by subtracting the \(k_{\text{deg}}\) from the net decay constant (\(k_{\text{deg}} - k_{\text{syn}}\)).
Parameters with uncertainty

When deciding how to describe the uncertainty for each parameter we must take into consideration that the reported values are either calculated or derived with approximation from experiments. Additionally, they correspond to synthesis rates of autoinducers in other bacteria (E. coli, P. aeruginosa) and not to butyrolactones. This means that there might be a notable difference between actual parameter values and the ones reported in literature. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distributions. By assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distribution was designed.

Therefore, the weight of the distribution is kept at 0.09 4 min\(^{-1}\) which is set as the mode of the log-normal distribution for \(K_C\) and the Spread is set to 60.49. In this way, the range where 68.27% of the values are found is between 0.00155 and 5.67 4 min\(^{-1}\).

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:
Appendix D. Parameter information for the GBL system model

The parameter information of the distribution is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>(\mu)</th>
<th>(\sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_C)</td>
<td>0.094</td>
<td>60.49</td>
<td>0.87914</td>
<td>1.8017</td>
</tr>
</tbody>
</table>

References


D.7 Transcription of $r$

The $scbR$ gene is transcribed into $scbR$ mRNA ($r$).

**Chemical equation**

$$O_R \rightarrow O_R + r$$

**Rate equation**

$$r = T_R \cdot [O_R] \text{ (for overlapping promoters)}$$
$$r = \Omega_R \cdot [O_R] \text{ (for isolated promoters)}$$

**Parameters**

The parameters of this reaction are the transcription rate of $scbR$ ($T_R$) and the firing rate $k_F$. The transcription rate is derived by the strength of the promoter ($\Omega_R$) but also taking into account the transcriptional interference by the $scbA$ promoter. In this model, we have assumed that the isolated promoter strength is equal to the number of transcripts produced per unit of time. Therefore, the parameter $\Omega_R$ is assumed to be equal to the transcription rate constant of the isolated promoter. These parameter values were derived from published data on *E.coli* mRNA transcription rate and calculations based on genomic properties of *Streptomyces coelicolor* A3(2). Additionally, the firing rate (elongation initiation rate) constant ($k_F$) is needed to test the strength of each promoter. This parameter is also derived from literature and is sampled from the same distribution for both $scbR$ and $scbA$ promoters, but is then multiplied by a heterogeneity factor $\chi$ to calculate the final $k_{FR}$ for the $scbR$ promoter. The heterogeneity factor for each promoter is sampled from a log-normal distribution.
### Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models [1] [2]</th>
<th>Remarks/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Omega_R$</td>
<td>0.16 – 8.24 min$^{-1}$</td>
<td>$10.8$ min$^{-1}$ $[1]$</td>
<td>0.18 s$^{-1}$</td>
<td>[10.8 \text{min}^{-1}] [9 \text{min}^{-1}] ( \Omega_R = 0.16 - 8.24 \text{min}^{-1} ) ( \Omega_R = 10.8 \text{min}^{-1} ) ( \Omega_R = 9 \text{min}^{-1} ) ( \Omega_R = 0.18 \text{s}^{-1} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.15 \times 10^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range tested:</td>
<td>$10^{-4} - 10$ s$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.006 - 600 \text{min}^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.013 - 0.25$ s$^{-1}$ ( [1] )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bitability range:</td>
<td>$7.8 - 12 \text{min}^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.0075 - 0.3$ s$^{-1}$ ( [2] )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$0.45 - 18 \text{min}^{-1}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In a recent publication by R.A. Cox, genomic properties and macromolecular compositions of *Streptomyces coelicolor* A3(2) and *E. coli* were reported, along with equations that connect these properties. For *S. coelicolor*, the polypeptide elongation rate $\xi_{\text{pol}}$ is reported to be in the range between 0.59 and 3.17 amino acids s$^{-1}$, from which the mRNA elongation rate can be calculated according to Cox from the equation $\xi_{\text{mRNA}} = 3 \xi_{\text{pol}}$ (factor 3 reflects the number of nucleotides per codon), therefore resulting in 1.77 and 9.51 nucleotides s$^{-1}$ as minimum and maximum values respectively. As *S. coelicolor* has 648 pb, the transcription rate constant can be calculated as per 648 bp/gene, $\Omega_R = 1.77 \text{bp/s}$ = 366 bp/gene = 0.1 min/gene and $\Omega_R = 9.51 \text{bp/s}$ = 68.14 s/gene = 1.14 min/gene, thus resulting in the final values of 0.16 and 0.88 min$^{-1}$.

### Table 4. Macromolecular properties of *E. coli* (strain AB1157) grown at 30 °C

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_{\text{pol}}$</td>
<td>Rate of polypeptide elongation (amino acids min$^{-1}$)</td>
</tr>
<tr>
<td>$\xi_{\text{mRNA}}$</td>
<td>Rate of mRNA elongation (nucleotides min$^{-1}$)</td>
</tr>
<tr>
<td>$\xi_{\text{tRNA}}$</td>
<td>Rate of tRNA elongation (nucleotides min$^{-1}$)</td>
</tr>
<tr>
<td>$\xi_{\text{rRNA}}$</td>
<td>Rate of rRNA elongation (nucleotides min$^{-1}$)</td>
</tr>
<tr>
<td>$\xi_{\text{DNA}}$</td>
<td>Rate of DNA elongation (nucleotides min$^{-1}$)</td>
</tr>
</tbody>
</table>

### Table 2. Definitions of variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_{\text{pol}}$</td>
<td>Rate of polypeptide elongation (amino acids min$^{-1}$)</td>
</tr>
<tr>
<td>$\xi_{\text{mRNA}}$</td>
<td>Rate of mRNA elongation (nucleotides min$^{-1}$)</td>
</tr>
<tr>
<td>$\xi_{\text{tRNA}}$</td>
<td>Rate of tRNA elongation (nucleotides min$^{-1}$)</td>
</tr>
<tr>
<td>$\xi_{\text{rRNA}}$</td>
<td>Rate of rRNA elongation (nucleotides min$^{-1}$)</td>
</tr>
<tr>
<td>$\xi_{\text{DNA}}$</td>
<td>Rate of DNA elongation (nucleotides min$^{-1}$)</td>
</tr>
</tbody>
</table>

### Additional Information

Additionally, Bremer et al. have reported an mRNA transcription rate of 55 nucleotides/s for *E. coli*, a value which is also shared by R.A. Cox, while Vogel et al. have published a range of mRNA transcription rates in the range of 28–89 nucleotides/s, depending on different growth rates of *E. coli*. By the same calculations, the corresponding transcription rate constants are 5.09 min$^{-1}$ and 2.59 – 8.24 min$^{-1}$.
Appendix D. Parameter information for the GBL system model

$R_F$ 18.2 – 33 $min^{-1}$ N/A

Pai et al. reported a typical transcription initiation rate in QS systems to be $20 min^{-1}$.

The synthesis rate of signal peptides in QS systems has not been measured. A typical transcription initiation rate ($\kappa_{Kemp}$) & $K_{Thio}$) of 20 min$^{-1}$ for a gene with an average mRNA half-life ($\kappa_{Kemp}$ & $K_{Thio}$) of 1 min gives a steady-state estimate of 10 mRNA per cell. We note that signal peptides (including peptide precursors) are typically less than 1000 nucleotides.

This value is also supported by Kennell et al. who calculated the transcription initiation rates from experimental data derived from in vitro experiments using E. coli. The results showed one initiation every 3.3 sec (therefore transcription rate 18.2 $min^{-1}$).

Finally, Tadmor et al. reported a maximum transcription initiation rate of 33 $min^{-1}$ in E. coli based on observational data.
In order to include the strength of the promoter in the transcription rate, the following formula is used: 
\[
\frac{1}{k_{onR}} = \frac{1}{\Omega_R} - \frac{1}{k_{FR}}. \quad [9]
\]
In this way, a distribution with the values for parameter \(k_{onR}\) (rate of RNA polymerase binding the promoter) is obtained, which is necessary to link the firing rate with the total transcription rate according to the formulas suggested by Bendtsen et al. [9] From the \(k_{FR}\) and the \(k_{onR}\), the maximal occupancy for the isolated \(scbR\) promoter can be calculated by using the formula: 
\[
\theta_R^o = \frac{k_{onR}}{k_{FR} + k_{onR}}.
\]
The promoter aspect ratio \(\alpha_R\) is equal to \(\frac{k_{onR}}{k_{FR}}\), which leads to the equivalent equation \(\alpha_R = \frac{\theta_R^o}{1 - \theta_R^o}\). The total transcription rate in the case of the isolated (uncoupled) promoters is calculated through the formula: \(\Omega_R = k_{FR} \cdot \theta_R^o\). Similarly, the maximal occupancy of the isolated \(scbA\) promoter \(\theta_A^o\) is calculated as described in the Section D.8. In the case of the overlapping (coupled) promoters the maximal occupancy for the \(scbR\) promoter is \(\theta_R = \frac{\alpha_R + 1}{\alpha_R + \alpha_A + 1}\), where \(\alpha_A\) is the promoter aspect ratio for the isolated \(scbA\) promoter. Therefore, the final transcription rate constant is calculated as per \(T_R = k_{FR} \cdot \theta_R\).

**Parameters with uncertainty**

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are either calculated or derived with approximation from experiments and from other macromolecular properties. Additionally, some of the values correspond to mRNA transcription rates of different bacteria species \((E. coli)\). This means that there might be a notable difference between actual parameter values and the ones reported in literature. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distributions. Therefore, by assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distributions were designed.

Although the weight of the distribution is put on the values calculated for \(S. coelicolor\) by setting \(0.8346 min^{-1}\) as the mode of the log-normal distribution for the \(\Omega_R\), we wish to explore the full range of reported values. Thus, the Spread is set to \(3.55\) and the range where \(68.27\%\) of the values are found is between \(0.235\) and \(2.96 \, min^{-1}\).

With regards to the firing rate \(k_{FR}\), the reported values are within the range of \(18 - 33 \, min^{-1}\) with the most probable values being \(18 - 20 \, min^{-1}\). Since these values are reported as being the average rates (and \(33 \, min^{-1}\) being the maximum), we
will also sample lower values, so the final sampling range will be around the values \(10 - 33\, \text{min}^{-1}\). The mode of the distribution is set to \(20.7\, \text{min}^{-1}\) and the Spread is set to 1.34. Therefore, the range where 68.27% of the values are found is between 15.4 and 27.8 \(\text{min}^{-1}\).

From the literature values of \(\Omega_R\) and \(k_F\), the values for \(k_{onR}\) are calculated. By assigning the same weights to each of the \(k_{onR}\) values as of the \(\Omega_R\) values, a distribution with a mode of 0.867 \(\text{min}^{-1}\) and a Spread of 4.1 is generated. The range where 68.27% of the values are found is between 0.212 and 3.546 \(\text{min}^{-1}\).

Finally, the heterogeneity factor can be set so that \(k_{FR} > k_{FA}\) or \(k_{FR} < k_{FA}\). In the first case, \(\chi\) has values within the range 1-10 and in the second within the range 0.1-0.9. In order for the two promoters to have equal strength, \(\chi\) is set to be equal to 1. Therefore, the mode of the heterogeneity factor \(\chi\) is 3 and the Spread is 1.2 (68.27% of the values are found between 2.5 and 3.6), in the case where \(scbR\) promoter is stronger. In the opposite case the mode is 0.4 and the Spread is 1.3 (68.27% of the values are found between 0.3 and 0.52).

The probability distributions for the parameters, adjusted accordingly in order to reflect the above values, are the following:

![Distribution for \(\Omega_R\) values](image)
Appendix D. Parameter information for the GBL system model

1. Distribution for $k_{\text{onR}}$ values

2. Distribution for $k_F$ values

3. Distribution for $\chi$ values ($k_F^R > k_F^A$)
The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines).

The parameter information of the distributions is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>( \mu )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Omega_R )</td>
<td>0.8346</td>
<td>3.55</td>
<td>0.63056</td>
<td>0.90076</td>
</tr>
<tr>
<td>( k_{onR} )</td>
<td>0.867</td>
<td>4.09</td>
<td>0.78694</td>
<td>0.96428</td>
</tr>
<tr>
<td>( k_F )</td>
<td>20.7</td>
<td>1.34</td>
<td>3.1107</td>
<td>0.28276</td>
</tr>
<tr>
<td>( \chi ) ( (k_{FR} &gt; k_{FA}) )</td>
<td>3</td>
<td>1.2</td>
<td>1.1308</td>
<td>0.1794</td>
</tr>
</tbody>
</table>

References


D.8 Transcription of a

The \textit{scbA} gene is transcribed into \textit{scbA} mRNA (a).

**Chemical equation**

\[ O_A \rightarrow O_A + a \]

(maximum transcription in the scenario where the activating complex \( AR_2 \) does not exist or basal transcription in the scenario where the activating complex \( AR_2 \) exists)

\[ O_A \cdot AR_2 \rightarrow O_A \cdot AR_2 + a \]

(maximum transcription in the scenario where the activating complex \( AR_2 \) exists)

**Rate equation**

\[ r = T_A \cdot [O_A] \]

(for overlapping promoters in the scenario where the activating complex \( AR_2 \) does not exist)

\[ r = \Omega_A \cdot [O_A] \]

(for isolated promoters in the scenario where the activating complex \( AR_2 \) does not exist)

\[ r = T_A \cdot [O_A - AR_2] \]

(maximal transcription for overlapping promoters in the scenario which includes the \( AR_2 \) complex)

\[ r = \Omega_A \cdot [O_A - AR_2] \]

(maximal transcription for isolated promoters in the scenario which includes the \( AR_2 \) complex)

\[ r = T_{\text{basal}} \cdot [O_A] \]

(basal transcription for overlapping promoters in the scenario which includes the \( AR_2 \) complex)

\[ r = \Omega_{\text{basal}} \cdot [O_A] \]

(basal transcription for isolated promoters in the scenario which includes the \( AR_2 \) complex)

**Parameters**

The parameters of this reaction are the basal and maximum transcription rate of I \( (T_A \) and \( T_{\text{basal}}) \). These parameters are derived by the strength of the promoter \( (\Omega_A \) and
\( \Omega_{\text{A}_{\text{basal}}} \) but also taking into account the transcriptional interference by the \textit{scbR} promoter. In this model, we have assumed that the isolated promoter strength is equal to the number of transcripts produced per unit of time. Therefore, the parameters \( \Omega_A \) and \( \Omega_{\text{A}_{\text{basal}}} \) are assumed to be equal to the transcription rate constant of the isolated promoter. These parameter values were derived from published data on \textit{E.coli} mRNA transcription rate and calculations based on genomic properties of \textit{Streptomyces coelicolor} A3(2). Additionally, the firing rate (elongation initiation rate) constant (\( k_F \)) is needed. This parameter is also derived from literature and is sampled from the same distribution for both \textit{scbR} and \textit{scbA} promoters, but is then divided by a heterogeneity factor \( \chi \) to calculate the final \( k_{FA} \) for the \textit{scbA} promoter. The heterogeneity factor for each promoter is sampled from a log-normal distribution.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models [1] [2]</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Omega_{\text{A}_{\text{basal}}} )</td>
<td>0.001 – 0.2 [2] [3] [4] [5]</td>
<td>( \text{min}^{-1} )</td>
<td>( 8 \times 10^{-4} \text{s}^{-1} (0.048 \text{min}^{-1}) )[2] ( 4 \times 10^{-5} – 2.8 \times 10^{-3} \text{s}^{-1}[1] ) ( (0.0024 – 0.168 \text{min}^{-1}) )</td>
<td>In the model investigating the quorum sensing switch in \textit{Vibrio fischeri}, Weber et al. [3] have used the ratio between basal and maximal transcription rates of 0.01 and 0.001, based on the characteristics shown in the experiments by Williams et al. [5]. Through these ratios, an estimation on the range of basal transcription of our system can be made from the maximum transcription rate (( \Omega_A )), i.e., ( 0.00011 – 0.0565 \text{min}^{-1} ). Additionally, in the GBL model by Chatterjee et al. [2], the basal transcription rates employed are within the range ( 4 \times 10^{-5} – 2.8 \times 10^{-3} \text{s}^{-1} (0.0024 – 0.168 \text{min}^{-1}) ). The values on which their parameters were used, were derived from \textit{in vitro} studies in \textit{Escherichia coli} by Nam et al. [4].</td>
</tr>
</tbody>
</table>
Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models [1][2]</th>
<th>Remarks/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Omega_A$</td>
<td>0.16 – 8.24</td>
<td>min$^{-1}$</td>
<td>0.45$^{[3]}$ (27 min$^{-1}$)</td>
<td>In a recent publication by R.A. Cox, genomic properties and macromolecular compositions of <em>Streptomyces coelicolor</em> A3(2) and <em>E. coli</em> were reported, along with equations that connect these properties. For <em>S. coelicolor</em>, the polypeptide elongation rate $\varepsilon_{pol}$ is reported to be in the range between 0.59 and 3.17 amino acids s$^{-1}$, from which the mRNA elongation rate can be calculated according to Cox from the equation $\varepsilon_{mRNA} = 3\varepsilon_{pol}$ (factor 3 reflects the number of nucleotides per codon), therefore resulting in 1.77 and 9.51 nucleotides $\cdot$ s$^{-1}$ as minimum and maximum values respectively. As ScaB has 945 pb, the transcription rate constant can be calculated as per $\Omega_A = \frac{945pb}{gene} = 533.9s/gene = 8.89min/gene \quad \Omega_A = \frac{945bp/gene}{0.95bp/gene} = 99.37s/gene = 1.66min/gene,$ \text{ thus resulting in final values of } 0.11min^{-1} \text{ and } 0.60min^{-1}.</td>
</tr>
<tr>
<td></td>
<td>0.45$^{[3]}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(27 min$^{-1}$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range tested:</td>
<td>$10^{-4}$ – $10s^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bistability range:</td>
<td>0.44 – 0.5s$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(26.4 – 30 min$^{-1}$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0225 – 0.9s$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.35 – 54 min$^{-1}$)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 1. Macromolecular properties of average cells of *S. coelicolor* A3(2) grown at 30 °C

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon_{pol}$</td>
<td>0.59</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>1.77</td>
</tr>
<tr>
<td>$\varepsilon_{pol}$</td>
<td>3.17</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>9.51</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>533.9</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>8.89</td>
</tr>
</tbody>
</table>

Cox et al. 2004[3]

### Table 2. Definitions of variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition of variable (units in parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$n_{pol}$</td>
<td>Average number of polypeptides per cell (number)</td>
</tr>
<tr>
<td>$p$</td>
<td>Specific growth rate (h$^{-1}$)</td>
</tr>
<tr>
<td>$\varepsilon_{pol}$</td>
<td>Average number of ribosomes per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of ribosomes per cell</td>
</tr>
<tr>
<td>$\varepsilon_{mRNA}$</td>
<td>Average number of mRNA per cell (number)</td>
</tr>
</tbody>
</table>

Additionaly, Brenner et al. have reported an mRNA transcription rate of 55 nucleotides/s for *E. coli*, a value which is also shared by R.A. Cox, while Vogel et al. have published a range of mRNA transcription rates in the range of 28-89 nucleotides/s, depending on different growth rates of *E. coli*. By the same calculations, the corresponding transcription rate constants are 5.09min$^{-1}$ and 2.59 – 8.24min$^{-1}$.
Appendix D. Parameter information for the GBL system model

chain-growth has been calculated. By comparison of the growth with the observed changes, obtained from the changing sedimentation patterns of radioactive RNA, the size of RNA chain-growth for chloroplast and mRNA has been determined to be 35 to 40 nucleotides per second at 37°C. This rate is independent of the length of the growing RNA molecules (in the range 200 to 3000 nucleotides) of the entire cell in the medium (0.05 to 5 μg/ml), and of the doubling time of the bacteria (30 to 48 min).

Bremer et al. 2004[4]

<table>
<thead>
<tr>
<th>Table 3. Rates of transcription elongation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>DNAjoining</td>
</tr>
</tbody>
</table>

Pai et al. reported a typical transcription initiation rate in QS systems to be 20 min$^{-1}$.

The synthesis rate of 


This value is also supported by Kennell et al. who calculated the transcription initiation rates from experimental data derived from in vitro experiments using E. coli. The results showed one initiation every 3.3 sec (therefore transcription rate 18.2 min$^{-1}$).

Kennell et al. 1977[8]

Finally, Tadmor et al. reported a maximum transcription initiation rate of 33 min$^{-1}$ in E. coli based on observational data.

Table 5. Genetic parameters for E. coli growing at 1 and 2.5 doublings/h, 37°C. Values not in parentheses are for 1.5 doublings/h. Values in parentheses are for 1.5 doublings/h; with the rest of the values being identical to those in 2.5 doublings/h. Note that all values are for the same culture grown at 37°C. Cell volumes are given in Table 5. Using these data for $N_{Ecoli}$ and $M_{Ecoli}$ when setting the separation of mass for a genetically perturbed cell populations concentrations (e.g., $N_{perturbed}$, $N_{wild}$, etc.) also in these units, i.e. concentrations atm$^{-1}$.

Tadmor et al. 2008[7]
Appendix D. Parameter information for the GBL system model

In order to include the strength of the promoter in the transcription rate, the following formula is used: $\frac{1}{k_{onA}} = \frac{1}{\Omega_A} - \frac{1}{k_{FA}}$.[12] In this way, a distribution with the values for parameter $k_{onA}$ is obtained (rate of RNA polymerase binding the promoter), which is necessary to link the firing rate with the total transcription rate according to the formulas suggested by Bendtsen et al.[12] From the $k_{FA}$ and the $k_{onA}$, the maximal occupancy for the isolated $scbA$ promoter can be calculated by using the formula: $\theta_A^o = \frac{k_{onA}}{k_{onA} + k_{FA}}$. The promoter aspect ratio $\alpha_A$ is equal to $\frac{k_{onA}}{k_{FA}}$, which leads to the equivalent equation $\alpha_A = \frac{\theta_A^o}{1 - \theta_A^o}$. The total transcription rate in the case of the isolated (uncoupled) promoters is calculated through the formula: $\Omega_A = k_{FA} \cdot \theta_A^o$. Similarly, the effective occupancy of the isolated $scbR$ promoter ($\theta_R^o$) is calculated as described in Section D.7. In the case of the overlapping (coupled) promoters the maximal occupancy for the $scbA$ promoter is $\theta_A = \frac{\alpha_A + 1}{\alpha_R + \alpha_A + 1}$, where $\alpha_R$ is the promoter aspect ratio for the isolated $scbR$ promoter. Therefore, the final transcription rate constant is calculated as per $T_A = k_{FA} \cdot \theta_A$.

Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are either calculated or derived with approximation from experiments and from other macromolecular properties. Additionally, some of the values correspond to mRNA transcription rates of different bacteria species (E. coli). This means that there might be a notable difference between actual parameter values and the ones reported in literature. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distributions. Therefore, by assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distributions were designed.

Therefore, although the weight of the distribution is put on the values calculated for S. coelicolor by setting 0.57 min\(^{-1}\) as the mode of the log-normal distribution for the $\Omega_A$, we wish to explore the full range of reported values. Thus, the Spread is set to 3.5 and the range where 68.27% of the values are found is between 0.162 and 2.026 min\(^{-1}\).

With regards to the firing rate $k_F$, the reported values are within the range of $18 - 33$ min\(^{-1}\) with the most probable values being $18 - 20$ min\(^{-1}\). Since these values are reported as being the average rates (and 33 min\(^{-1}\) being the maximum), we
will also sample lower values, so the final sampling range will be around the values $10 - 33 \text{min}^{-1}$. The mode of the distribution is set to $20.7 \text{min}^{-1}$ and the Spread is set to 1.34. Therefore, the range where 68.27% of the values are found is between 15.4 and $27.8 \text{min}^{-1}$.

From the literature values of $\Omega_A$ and $k_F$, the values for $k_{on,A}$ are calculated. By assigning the same weights to each of the $k_{on,A}$ values as of the $\Omega_A$ values, a distribution with a mode of $0.587 \text{min}^{-1}$ and a Spread of 3.875 is generated. The range where 68.27% of the values are found is between 0.151 and 2.28.

With regards to the basal transcription rate $\Omega_{A_{basal}}$, in order to explore the full range of the literature values, the mode was set to $0.0037 \text{min}^{-1}$ and the Spread to 10.93. Therefore the range where 68.27% of the values are found is between 0.00034 and 0.041.

As performed for $k_{on,A}$, the values for $k_{on,A_{basal}}$ are calculated from the literature values of $\Omega_{A_{basal}}$ and $k_F$. By assigning the same weights to each of the $k_{on,A_{basal}}$ values as of the $\Omega_{A_{basal}}$ values, a distribution with a mode of $0.00373 \text{min}^{-1}$ and a Spread of 10.97 is generated. The range where 68.27% of the values are found is between 0.00034 and 0.0409.

Finally, the heterogeneity factor can be set so that $k_{F_R} > k_{F_A}$ or $k_{F_R} < k_{F_A}$. In the first case, $\chi$ has values within the range 1-10 and in the second within the range 0.1-0.9. In order for the two promoters to have equal strength, $\chi$ is set to be equal to 1. Therefore, the mode of the heterogeneity factor $\chi$ is 3 and the Spread is 1.2 (68.27% of the values are found between 2.5 and 3.6), in the case where $scbR$ promoter is stronger. In the opposite case the mode is 0.4 and the Spread is 1.3 (68.27% of the values are found between 0.3 and 0.52).

The probability distributions for the parameters, adjusted accordingly in order to reflect the above values, are the following:
Appendix D. Parameter information for the GBL system model

Distribution for $k_{\text{onA}_{\text{basal}}}$ values

Distribution for $k_F$ values

Distribution for $\chi$ values ($k_{F_R} < k_{F_A}$)
Appendix D. Parameter information for the GBL system model

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines).

The parameter information of the distributions is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>( \mu )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Omega_A )</td>
<td>0.572</td>
<td>3.5</td>
<td>0.2514</td>
<td>0.89956</td>
</tr>
<tr>
<td>( \Omega_{A_{basal}} )</td>
<td>0.0037</td>
<td>10.93</td>
<td>-3.8324</td>
<td>1.3259</td>
</tr>
<tr>
<td>( k_{on_A} )</td>
<td>0.587</td>
<td>3.875</td>
<td>0.35255</td>
<td>0.94054</td>
</tr>
<tr>
<td>( k_{on_{A_{basal}}} )</td>
<td>0.00373</td>
<td>10.97</td>
<td>-3.8297</td>
<td>1.3269</td>
</tr>
<tr>
<td>( k_F )</td>
<td>20.7</td>
<td>1.34</td>
<td>3.1107</td>
<td>0.28276</td>
</tr>
<tr>
<td>( \chi(k_{F_R} &lt; k_{F_A}) )</td>
<td>0.4</td>
<td>1.3</td>
<td>-0.8517</td>
<td>0.2541</td>
</tr>
</tbody>
</table>

References


D.9 Antisense interaction between r and a

The full length mRNA of scbR (r) binds to the full length mRNA of scbA (a) and form a complex which prevents further translation of both mRNAs.

**Chemical equation**

\[ a + r \rightleftharpoons r \cdot a \]

**Rate equation**

\[ r = \frac{k_{ar}^-}{K_{ar}} \cdot [r] \cdot [a] - k_{ar}^- \cdot [r \cdot a] \]

**Parameters**

The parameters of this reaction is the binding and unbinding rate constant for the interaction of scbR mRNA with scbA mRNA \((K_{ar})\) and \((k_{ar}^-)\). The parameter values are derived from various RNA/RNA interaction studies in bacteria.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL model (^{[1]})</th>
<th>Remarks-Reference</th>
</tr>
</thead>
</table>
| \(K_{ar}\) | 0.4 – 89 \([2] [3] [4] [5] [6]\) | nM | 0.001nM\(^{-1}\)s\(^{-1}\) \((0.06nM\(^{-1}\)min\(^{-1}\))\(^{[1]}\) | Bistability range: 
\(6.5 \cdot 10^{-4} - 0.16nM\(^{-1}\)s\(^{-1}\) \((0.030 - 9.6nM\(^{-1}\)min\(^{-1}\))\) |
|       |           |              |                                       | In a 1991 review, Y. Eguchi reported association rates for double stranded RNA binding in the range of \(1 \cdot 10^5 - 3 \cdot 10^6M\(^{-1}\)s\(^{-1}\) \((0.006 - 0.18nM\(^{-1}\)min\(^{-1}\))\). A similar range \((10^5M^{-1}s^{-1})\) is published by S. Brantl in a 2007 review on regulatory mechanisms employed by antisense RNAs. |

For many systems, antisense/sense RNA binding pathways have been studied in detail and binding kinetics are assisted. Analysis of pairing rate constants usually yielded values of \(~10^5M^{-1}s^{-1}\) . The initial contact between antisense and sense RNAs that often form secondary structures, can either occur between two complementary loops (many replication control systems, reviewed in [2]) or between a loop and a single-stranded region (e.g. RNA-RNA-OCT [19] and holc/sok [10]). In the first case, simple helixs |

S. Brantl 2007 \([3]\) |

These values are also in agreement with an *in vitro* study by Franch et al. on the effect of a U-turn loop structure in RNA/RNA interactions, which employed *E. coli* hok-RNAs and Sok-RNAs.
Appendix D. Parameter information for the GBL system model

The study reported binding rate constants $0.15 \cdot 10^6 - 5 \cdot 10^6 M^{-1}s^{-1}$ (0.009 - 0.3 M$^{-1}$ min$^{-1}$) and suggested that $10^6 M^{-1}s^{-1}$ is the upper limit for this interaction.

Table 1. The observed in vitro binding rate constants ($k_{on}$) for pairing of mutant self-RNAs and homologous self-RNAs.

<table>
<thead>
<tr>
<th>mRNA</th>
<th>Loop sequence</th>
<th>Binding rate constant ($k_{on}$)</th>
<th>Selecto in vitro</th>
<th>YNKR sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wildtype</td>
<td>GCGGCCG</td>
<td>20.0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>L1</td>
<td>AGGGCG</td>
<td>7.6</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>L2</td>
<td>AGGGCG</td>
<td>1.9</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>L3</td>
<td>AGGGCG</td>
<td>0.5</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>L4</td>
<td>AGGGCG</td>
<td>0.2</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>L5</td>
<td>AGGGCG</td>
<td>5.0</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>L6</td>
<td>AGGGCG</td>
<td>1.5</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>L7</td>
<td>AGGGCG</td>
<td>0.1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>L8</td>
<td>AGGGCG</td>
<td>0.1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>L9</td>
<td>AGGGCG</td>
<td>0.1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>L10</td>
<td>AGGGCG</td>
<td>0.1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The sequence corresponding to the YNKR motif of the wild-type sequence is shown in bold. Mutations are underlined. 11 site binding was also monitored using 2-methyl-3-ethyl-5-nitrobenzoyl mRNA for all loop mutations. The observed second-order binding-rate-constants showed the average of at least four experiments.

Franch et al. 1999[5]

Interestingly, these alternative pairing mechanisms yield similar second-order binding rate constants.

Franch et al. 1999[5]

Furthermore, Nordgren et al. published a study on the kinetics of antisense RNA interactions by using Surface Plasmon Resonance (in wild-type and mutant RNA pairs), where the values reported for the association rate constants are in the range of $4.5 \cdot 10^4 - 9.2 \cdot 10^5 M^{-1}s^{-1}$ (0.0027 - 0.0552 M$^{-1}$ min$^{-1}$) and the resulting equilibrium dissociation constant (calculated from the other parameters and measured by gel shift assay) are in the range of $0.4 \cdot 10^{-10} - 8.9 \cdot 10^{-8} M$ (0.4 - 89 nM). The experiments measured in vitro the rate of complex formation and in vivo the efficiency of control.

Table 1. Association rate constants, dissociation rate constants and equilibrium dissociation constants of wild-type and mutant patatin RNA pairs.

<table>
<thead>
<tr>
<th>Component</th>
<th>Wild-type</th>
<th>Mutant</th>
<th>25°C</th>
<th>32°C</th>
<th>37°C</th>
<th>25°C</th>
<th>32°C</th>
<th>37°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{on}$</td>
<td>1.1$\times$10$^6$</td>
<td>0.8$\times$10$^6$</td>
<td>0.5$\times$10$^6$</td>
<td>0.1$\times$10$^6$</td>
<td>0.05$\times$10$^6$</td>
<td>0.02$\times$10$^6$</td>
<td>0.01$\times$10$^6$</td>
<td>0.005$\times$10$^6$</td>
</tr>
<tr>
<td>$k_{off}$</td>
<td>1.0$\times$10$^{-4}$</td>
<td>0.8$\times$10$^{-4}$</td>
<td>0.5$\times$10$^{-4}$</td>
<td>0.1$\times$10$^{-4}$</td>
<td>0.05$\times$10$^{-4}$</td>
<td>0.02$\times$10$^{-4}$</td>
<td>0.01$\times$10$^{-4}$</td>
<td>0.005$\times$10$^{-4}$</td>
</tr>
<tr>
<td>$K_{d}$</td>
<td>4.0$\times$10$^{-6}$</td>
<td>3.2$\times$10$^{-6}$</td>
<td>2.5$\times$10$^{-6}$</td>
<td>1.0$\times$10$^{-6}$</td>
<td>0.8$\times$10$^{-6}$</td>
<td>0.5$\times$10$^{-6}$</td>
<td>0.1$\times$10$^{-6}$</td>
<td>0.05$\times$10$^{-6}$</td>
</tr>
</tbody>
</table>

Nordgren et al. 2001[6]

Finally, Lima et al. (see table below) in a study measuring kinetics on synthetic oligonucleotides, reported a range of association constants between $6 \cdot 10^6 - 2 \cdot 10^7 M^{-1}s^{-1}$ (0.36 - 12 M$^{-1}$ min$^{-1}$).

Therefore, the overall range of values for the association rate $k_{on}$ is 0.0027 - 12 M$^{-1}$ min$^{-1}$ and for the equilibrium dissociation constant $K_{d}$ is 0.4 - 89 nM.

$k_{on}$: $0.006 - 6 \text{ min}^{-1}$

<table>
<thead>
<tr>
<th>$k_{off}$</th>
<th>0.01$\text{s}^{-1}$ (0.6 min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bistability range</td>
<td>$0 - 0.2 \text{s}^{-1}$ (0 - 12 min$^{-1}$)</td>
</tr>
</tbody>
</table>

In the study by Nordgren et al., dissociation rate constants in the range $6 \cdot 10^{-4} - 7 \cdot 10^{-3} \text{s}^{-1}$ (0.024 - 0.462 min$^{-1}$) were reported (see figure above). These values are also consistent with the ones published by Lima et al. in a study on antisense oligonucleotide hybridization kinetics, who reported dissociation rate constant values in the range $1 \cdot 10^{-4} - 2 \cdot 10^{-2} \text{s}^{-1}$ (0.006 - 1.2 min$^{-1}$)
Parameters with uncertainty

When deciding how to describe the uncertainty for these parameter we must take into consideration the fact that many of the reported values are derived from in vitro experiments and correspond to mRNA sequences or fragments of different bacteria species. This means that there might be some difference between actual parameter values and the ones reported in literature. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution. Therefore, by assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distributions were designed.

With regards to the $K_{ar}$, the values that correspond to the wild type RNAs are between $1 - 10 nM$, therefore we put the weight of the distribution in this range and we consider as least likely the larger values as they correspond to mutant RNAs. Therefore, the mode of the log-normal distribution is calculated to be $7.8 nM$ and the Spread is 7.27. Thus the range where 68.27% of the values are found is between 1.074 and 56.73.

Similarly, the values reported for $k_{ar}$ are within the range $0.06 - 1.2 min^{-1}$, we put the weight of the distribution within this range but also explore the whole range of values. Therefore, the mode of the log-normal distribution is set to 0.22 and the Spread is 4.8. Thus the range where 68.27% of the values are found is between 0.046 and 1.075 $nM$.

Finally, the probability distribution for $k_{onar}$ is defined accordingly, in order to allow the exploration of the full range of the values retrieved from literature. Therefore, the mode is set to $0.046 nM^{-1} min^{-1}$ and the Spread is 11. In this way the range where
68.27% of the values are found is between 0.004 and 0.51\text{nm}^{-1}\text{min}^{-1}. The initial distribution parameters are $\mu=-1.3114$ and $\sigma=1.3294$.

Since the three parameters are interdependent, thermodynamic consistency also needs to be taken into account. This is achieved by creating a bivariate system as described in Chapter 3. Since $k_{on_{ar}}$ is the parameter with the largest geometric coefficient of variation, this is set as the dependent parameter as per: $k_{on_{ar}} = \frac{k_{ar}^\alpha}{K_{ar}}$, and an updated probability distribution is defined. The location and scale parameters of $k_{on_{ar}}$ ($\mu=-3.9001$ and $\sigma=1.5742$) were calculated from those of $K_{ar}$ and $k_{ar}^\alpha$.

The probability distributions for the three parameters, adjusted accordingly in order to reflect the above values, are the following:
Appendix D. Parameter information for the GBL system model

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines). The correlation matrix which is necessary to define the relationship between the two marginal distributions \( (K_{ar}, k_{ar}^-) \) of the bivariate system is derived by employing random values generated by the two distributions.

The parameter information of the distributions of the multivariate system is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>( \mu )</th>
<th>( \sigma )</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_{ar} )</td>
<td>7.8</td>
<td>7.27</td>
<td>3.4665</td>
<td>1.1882</td>
<td>N/A</td>
</tr>
</tbody>
</table>
| \( k_{on_{ar}} \) | N/A | N/A    | -3.9001  | 1.5742   | \[
\begin{pmatrix}
1 & 0.5214 \\
0.5214 & 1
\end{pmatrix}
\] |
| \( k_{ar}^- \) | 0.223 | 4.8 | -0.43359 | 1.0326   |                   |

The multivariate system of the normal distributions \( \ln(k_{on_{ar}}) \) and \( \ln(k_{ar}^-) \) and the resulting samples of values are presented in the following figure:
In this way, a system of distributions is created where each distribution is described and constrained by the other two. Therefore, the parameters will be sampled by the two marginal distributions in a way consistent with our beliefs and with the relevant thermodynamic constraints. However, since the model’s reaction rate requires the parameters $K_{ar}$ and $k_{ar}$, and not the $k_{onar}$, the value for $K_{ar}$ is calculated by the parameters sampled from the other two distributions in an additional step, as per $K_{ar} = \frac{k_{ar}}{k_{onar}}$.

References


Appendix D. Parameter information for the GBL system model

D.10 Translation of R

The \textit{scbR} mRNA (r) is translated into ScbR protein (R).

**Chemical equation**

\[ r \rightarrow r + R \]

**Rate equation**

\[ r = P_R \cdot [r] \]

**Parameters**

The parameter of this reaction is the translation rate of ScbR ($P_R$). The parameter values were derived from calculations of the protein translation rate based on genomic properties of \textit{Streptomyces coelicolor} A3(2) and \textit{E. coli}.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models [1] [2]</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_R$</td>
<td>0.165 – 5.86</td>
<td>(\text{min}^{-1})</td>
<td>0.36s(^{-1})</td>
<td>In a recent publication by R.A. Cox, genomic properties and macromolecular compositions of \textit{Streptomyces coelicolor} A3(2) and \textit{E. coli} were reported, along with equations that connect these properties. For \textit{S. coelicolor}, the polypeptide elongation rate $e_{\text{pol}}$ is reported to be in the range of 0.59 and 3.17 amino acids s(^{-1}). As ScbR has 215 aa, the transcription rate constant can be calculated as per $P_R = \frac{215 \text{aa}/\text{gene}}{0.59 \text{aa}/\text{s}} = 364.48 \text{s/gene} = 6.07 \text{min/gene}$ and $P_R = \frac{215 \text{aa}/\text{gene}}{3.17 \text{aa}/\text{s}} = 67.82 \text{s/gene} = 1.13 \text{min/gene}$, thus resulting in a final value of 0.165s(^{-1}) and 0.89s(^{-1}).</td>
</tr>
</tbody>
</table>

[1] Cox et al., 2004

Appendix D. Parameter information for the GBL system model

Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration the fact that the reported values are either calculated or derived with approximation from experiments and from other macromolecular properties. Additionally, some of the values correspond to protein translation rates of different bacteria species (E. coli). This means that there might be a notable difference between actual parameter values and the ones reported in literature. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution.

Additionally, Cox reports translation rates for E. coli between 12 and 21 aa/s. By the same calculations, the corresponding transcription rate constants are in the range $3.35 - 5.86 \text{min}^{-1}$.

Cox et al. 2004[3]
$P_R$, we wish to explore the full range of reported values. Thus, the Spread is set to 3.2 and the range where 68.27% of the values are found is between 0.23 and 2.38 $\text{min}^{-1}$.

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

![Distribution for $P_R$ values](image.png)

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using a default value of 10% error (orange lines).

The parameter information of the distribution is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_R$</td>
<td>0.744</td>
<td>3.2</td>
<td>0.42952</td>
<td>0.85176</td>
</tr>
</tbody>
</table>

**References**


Appendix D. Parameter information for the GBL system model

D.11 Translation of A

The scbA mRNA (a) is translated into ScbA protein (A).

**Chemical equation**

\[ a \rightarrow a + A \]

**Rate equation**

\[ r = P_A \cdot [a] \]

**Parameters**

The parameter of this reaction is the translation rate of ScbA \((P_A)\). The parameter values were derived from calculations of the protein translation rate based on genomic properties of *Streptomyces coelicolor* A3(2) and *E.coli*.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P_A)</td>
<td>0.11 – 4 [^3]</td>
<td>(min^{-1})</td>
<td>0.066s(^{-1}) ((3.96min^{-1}))[1][2]</td>
<td>In a recent publication by R.A. Cox, genomic properties and macromolecular compositions of <em>Streptomyces coelicolor</em> A3(2) and <em>E.coli</em> were reported, along with equations that connect these properties. For <em>S. coelicolor</em>, the polypeptide elongation rate (e_{pol}) is reported to be in the range between 0.59 and 3.17 amino acids s(^{-1}). As ScbA has 314 aa, the transcription rate constant can be calculated as per: (P_A = \frac{314aa/\text{gene}}{0.59s/aa} = 532.2s/\text{gene} = 8.87min/\text{gene}) and (P_A = \frac{215aa/\text{gene}}{3.17s/aa} = 99.1s/\text{gene} = 1.65min/\text{gene}), thus resulting in a final value of 0.61min(^{-1}) and 0.11min(^{-1}).</td>
</tr>
</tbody>
</table>

---

[1] S. coelicolor A3(2)\(\rightarrow\)E. coli 1996\(\Rightarrow\)1998 (under conditions of high and low aminoglycoside resistance)


---

\[^3\] Cox et al. 2004 [5]
Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration the fact that the reported values are either calculated or derived with approximation from experiments and from other macromolecular properties. Additionally, some of the values correspond to protein translation rates of different bacteria species (E. coli). This means that there might be a notable difference between actual parameter values and the ones reported in literature. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution.

Therefore, although the weight of the distribution is put on the values calculated for S. coelicolor by setting $0.51 \text{ min}^{-1}$ as the mode of the log-normal distribution for the
Appendix D. Parameter information for the GBL system model

$P_A$, we wish to explore the full range of reported values. Thus, the Spread is set to 3.2 and the range where 68.27% of the values are found is between 0.159 and 1.63 min$^{-1}$.

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

![Distribution for $P_A$ values](image)

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using a default value of 10% error (orange lines).

The parameter information of the distribution is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_A$</td>
<td>0.5</td>
<td>3.2</td>
<td>0.053356</td>
<td>0.85327</td>
</tr>
</tbody>
</table>

References


D.12 Formation of homo-dimer $R_2$

Two ScbR (R) proteins bind together to form an ScbR homo-dimer ($R_2$).

**Chemical equation**

$$2R \rightleftharpoons R_2$$

**Rate equation**

$$r = \frac{k^-}{K_{d6}} \cdot [R]^2 - k_6^- \cdot [R_2]$$

**Parameters**

The parameters of this reaction are the dissociation constant for binding of one ScbR to another ($K_{d6}$) and the dissociation rate for binding of one ScbR to another ($k_6^-$).

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{d6}$</td>
<td>$2.58 - 42.9$</td>
<td>nM</td>
<td>Majka et al. published a study on dimerization of the initiator Protein DNA A of Streptomycetes and on its mutants, where they report dissociation constants in the range $35.3 - 114\text{ nM}$.</td>
</tr>
</tbody>
</table>

Majka et al. 2004[3]
Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the values reported in literature correspond to *in vitro* testing of different protein-protein interaction and dimerization reactions than ScbR, although they refer to another *Streptomyces* protein (DnaA). This means that there might be a difference between actual parameter values and the ones reported in literature. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distributions. By assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distributions were designed.
Appendix D. Parameter information for the GBL system model

More specifically, for $K_{d6}$ the mode of the log-normal distribution is $3.89\text{nM}$ (value that is close to the wild type protein homodimerization). However, in order to explore a larger part of the nanomolar scale when sampling for parameter values, the Spread is calculated to be 1.9. In this way, the range where 68.27% of the values are found is between 2 and $7.4\text{nM}$.

With regards to the parameter $k_{on6}$, in order to explore the full range of plausible values, the mode of the log-normal distribution is set to $0.3\text{nM}^{-1}\text{min}^{-1}$ and the Spread is 15.1. Thus the range where 68.27% of the values are found is between 0.02 and $4.66\text{nM}^{-1}\text{min}^{-1}$.

Since the two parameters are interdependent, thermodynamic consistency also needs to be taken into account. This is achieved by creating a bivariate system as described in Chapter 3. Since no information was retrieved for $k_{6}^{-}$ and therefore is the parameter with the largest geometric coefficient of variation, this is set as the dependent parameter as per: $k_{6}^{-} = k_{on6} \cdot K_{d6}$. The location and scale parameters of $k_{6}^{-}$ ($\mu=2.5303$ and $\sigma=1.5324$) were calculated from those of $K_{d6}$ and $k_{on6}$.

The probability distributions for the two parameters, adjusted accordingly in order to reflect the above values, are the following:

![Distribution for $K_{d6}$ values](image)
The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines). The correlation matrix which is necessary to define the relationship between the two marginal distributions \((k_{d6}, k_6)\) of the bivariate system is derived by employing random values generated by the two distributions.

The parameter information of the distributions of the multivariate system is:
### Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>Correlation matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{on6}$</td>
<td>0.307</td>
<td>15.1</td>
<td>0.85865</td>
<td>1.4273</td>
<td>N/A</td>
</tr>
<tr>
<td>$K_{d6}$</td>
<td>3.89</td>
<td>1.9</td>
<td>1.6716</td>
<td>0.55779</td>
<td>$\begin{pmatrix} 1 &amp; 0.2176 \ 0.2176 &amp; 1 \end{pmatrix}$</td>
</tr>
<tr>
<td>$k_{6}^{-}$</td>
<td>N/A</td>
<td>N/A</td>
<td>2.5303</td>
<td>1.5324</td>
<td></td>
</tr>
</tbody>
</table>

The multivariate system of the normal distributions ($ln(k_{d6})$ and $ln(k_{6}^{-})$) and the resulting samples of values are presented in the following figure:

![Bivariate system of parameters](image)

In this way, a system of distributions is created where each distribution is described and constrained by the other two. Therefore, the parameters will be sampled by the two marginal distributions in a way consistent with our beliefs and with the relevant thermodynamic constraints.
Appendix D. Parameter information for the GBL system model

References


Appendix D. Parameter information for the GBL system model

D.13 Degradation of $r$

The mRNA of $scbR$ ($r$) degrades.

**Chemical equation**

$$r \rightarrow \emptyset$$

**Rate equation**

$$r = d_{mR} \cdot [r]$$

**Parameters**

The parameter of this reaction is the degradation rate of $r$ ($d_{mR}$). The parameter values were derived from calculations of the mRNA half life in gram positive bacteria with low GC content and in *Actinomycetes*.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{mR}$</td>
<td>0.03 – 0.365</td>
<td>min$^{-1}$</td>
<td>$0.0017s^{-1}$ ($0.102min^{-1}$)$^{[1]}$</td>
<td>In a study on global analysis of mRNA decay in gram positive bacteria by RNA-Seq, Kristoffersen et al. reported a range of values for the half-life ($t_{1/2}$) of mRNA in <em>B. cereus</em> between $1.9 – 5.6min$, with a median of $2.4min$. From these values we calculated the mRNA degradation rate of $scbR$ as per $d_{mR} = \frac{\ln(2)}{t_{1/2}}$, which resulted in a degradation rate constant value range $0.124 – 0.365min^{-1}$ and a median $0.29min^{-1}$, $0.007s^{-1}$ ($0.42min^{-1}$)$^{[2]}$</td>
</tr>
</tbody>
</table>
Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are either calculated or derived with approximation from experiments. Additionally they correspond to mRNA degradation rates of different bacteria species (B. cereus) and to unstable mRNA in S. antibioticus spores. This means that there might be a notable difference between the actual parameter values and the ones reported in literature. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution. By assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distributions were designed.
Therefore, although the weight of the distribution is put on the values calculated for the *Streptomyces* species by setting the mode of the log-normal distribution of the $d_{mR}$ to 0.14 min$^{-1}$, we wish to explore the full range of reported values and maybe even sample a small percentage of values outside the reported range. Thus, the Spread was set to 2.06, which means that the range where 68.27% of the values are found is between 0.068 and 0.29 min$^{-1}$.

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

![Distribution for $d_{mR}$ values](image)

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using a default value of 10% error (orange lines).

The parameter information of the distribution is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{mR}$</td>
<td>0.14</td>
<td>2.06</td>
<td>-1.5916</td>
<td>0.60824</td>
</tr>
</tbody>
</table>

References


D.14 Degradation of R

The ScbR monomer (R) degrades.

Chemical equation

\[ R \rightarrow \emptyset \]

Rate equation

\[ r = d_R \cdot [R] \]

Parameters

The parameter of this reaction is the degradation rate of R \((d_R)\). The parameter values were derived from proteomics studies on different proteins of \(S. coelicolor\).

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_R)</td>
<td>(5.02 \cdot 10^{-4} )</td>
<td>(min^{-1})</td>
<td>(0.00578min^{-1})</td>
<td>In a quantitative proteomics study on protein turnover rates in dynamic systems, Jayapel et al. reported the degradation rates of 115 proteins in (S. coelicolor) cultures undergoing transition from exponential growth to stationary phase. The values were in the range (0.05 \pm 0.014h^{-1} - 0.223 \pm 0.031h^{-1}) ((8.3 \pm 2.3 \cdot 10^{-4} - 0.0037 \pm 0.0005min^{-1})) with a median of (0.097h^{-1}(0.00162min^{-1})).</td>
</tr>
</tbody>
</table>
Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are estimated or calculated from proteomics experiments, using methods which were prone to measurement errors (i.e. optical density...
of fragmented mycelia to estimate the growth rate ($\mu$)) and extrapolations from cell density to total cellular protein without taking into account the dilution due to cell division. Additionally, the study did not include measurements on ScbR protein specifically, although it was conducted on a wide range of S. coelicolor proteins which enables a rough estimation of the range of values of protein degradation in this species. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution. By assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distribution was designed.

The mode of the log-normal distribution of the $d_R$ to $0.00144 \text{min}^{-1}$. In order to explore the full range of published values and even sample a percentage of values outside the reported range, the Spread is set to 1.78, so that the range where 68.27% of the values are found is between $8 \cdot 10^{-4}$ and $0.0026 \text{min}^{-1}$.

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines).

The parameter information of the distribution is:
Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>( \mu )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_R )</td>
<td>0.00144</td>
<td>1.78</td>
<td>6.2837</td>
<td>0.50981</td>
</tr>
</tbody>
</table>

References


D.15 Degradation of $R_2$

The ScbR homo-dimer ($R_2$) degrades.

**Chemical equation**

$$R_2 \rightarrow \emptyset$$

**Rate equation**

$$r = d_{R_2} \cdot [R_2]$$

**Parameters**

The parameter of this reaction is the degradation rate of $R_2$ ($d_{R_2}$). The parameter values were derived from proteomics studies on different proteins of *S. coelicolor*.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models [1] [2]</th>
<th>Remarks-Reference</th>
</tr>
</thead>
</table>
| $d_{R_2}$ | $5.02 \cdot 10^{-4}$ | $min^{-1}$ | 0.004$s^{-1}$  
               $0.24min^{-1}$[1]  
               $0.36s^{-1}$  
               $21.6min^{-1}$[2]  
               Range tested:  
               $10^{-7} - 10^{-1}s^{-1}$  
               $(6 \cdot 10^{-6}$  
               $- 6min^{-1})$  
               Bistability range:  
               $0.0022$  
               $0.0067$s$^{-1}$[1]  
               $0.132$  
               $0.402min^{-1}$  
               and $0.09$  
               $0.38s^{-1}[2]$  
               $(5.4$  
               $- 22.8min^{-1})$ | In a quantitative proteomics study on protein turnover rates in dynamic systems, Jayapal et al. reported the degradation rates of 115 proteins in *S. coelicolor* cultures undergoing transition from exponential growth to stationary phase. The values were in the range $0.05 \pm 0.014h^{-1} - 0.223 \pm 0.031h^{-1}$  
               $(8.3 \pm 2.3 \cdot 10^{-4} - 0.0037 \pm 0.00052min^{-1})$  
               with a median of $0.097h^{-1}(0.00162min^{-1})$. |
Appendix D. Parameter information for the GBL system model

Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are estimated or calculated from proteomics experiments, using methods which were prone to measurement errors (i.e. optical density
of fragmented mycelia to estimate the growth rate ($\mu$)) and extrapolations from cell density to total cellular protein without taking into account the dilution due to cell division. Additionally, the study did not include measurements on ScbR protein specifically, although it was conducted on a wide range of *S. coelicolor* proteins which enables a rough estimation of the range of values of protein degradation in this species. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution. By assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distribution was designed.

The mode of the log-normal distribution of the $d_{R_2}$ to $0.00144 \text{min}^{-1}$. In order to explore the full range of published values and even sample a percentage of values outside the reported range, the Spread is set to 1.78, so that the range where 68.27% of the values are found is between $8 \cdot 10^{-4}$ and $0.0026 \text{min}^{-1}$.

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

![Distribution for $d_{R_2}$ values](image)

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines).

The parameter information of the distribution is:
Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{R_2}$</td>
<td>0.00144</td>
<td>1.78</td>
<td>$-6.2837$</td>
<td>0.50981</td>
</tr>
</tbody>
</table>

References


D.16 Degradation of a

The mRNA of scbA (a) degrades.

Chemical equation

\[ a \rightarrow \emptyset \]

Rate equation

\[ r = d_{mA} \cdot [a] \]

Parameters

The parameter of this reaction is the degradation rate of a \((d_{mA})\). The parameter values were derived from calculations of the mRNA half life in gram positive bacteria with low GC content and in \textit{Actinomycetes}.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_{mA})</td>
<td>0.03 - 0.365</td>
<td>(min^{-1})</td>
<td>0.0018(s^{-1}) (0.108(min^{-1})) [1][2]</td>
<td>In a study on global analysis of mRNA decay in gram positive bacteria by RNA-Seq, Kristoffersen et al. reported a range of values for the half-life ((t_{1/2})) of mRNA in \textit{B. cereus} between 1.9 - 5.6(min), with a median of 2.4(min). From these values we calculated the mRNA degradation rate of scbA as per (d_{mA} = \frac{ln(2)}{t_{1/2}}), which resulted in a degradation rate constant value range 0.124 - 0.365(min^{-1}) and a median 0.29(min^{-1}).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8.1 (\cdot 10^{-4})(s^{-1}) (0.0486(min^{-1})) [2]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.6 (\cdot 10^{-4}) - 1.2 (\cdot 10^{-2})(s^{-1}) (0.0396 - 0.72(min^{-1}))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bistability range: 0.0016 - 0.019(s^{-1}) [1]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.096 - 1.14(min^{-1}))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and 4.1 (\cdot 10^{-4}) - 1.6 (\cdot 10^{-2})(s^{-1}) [2]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0246 - 0.96(min^{-1}))</td>
<td></td>
</tr>
</tbody>
</table>
Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are either calculated or derived with approximation from experiments. Additionally they correspond to mRNA degradation rates of different bacteria species (B. cereus) and to unstable mRNA in S. antibioticus spores. This means that there might be a notable difference between the actual parameter values and the ones reported in literature. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution. By assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distributions were designed.
Therefore, although the weight of the distribution is put on the values calculated for the Streptomyces species by setting the mode of the log-normal distribution of the \( d_{mA} \) to 0.14 min\(^{-1}\), we wish to explore the full range of reported values and maybe even sample a small percentage of values outside the reported range. Thus, the Spread was set to 2.06, which means that the range where 68.27\% of the values are found is between 0.068 and 0.29 min\(^{-1}\).

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using a default value of 10\% error (orange lines).

The parameter information of the distribution is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>( \mu )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_{mA} )</td>
<td>0.14</td>
<td>2.06</td>
<td>-1.5916</td>
<td>0.60824</td>
</tr>
</tbody>
</table>

References


D.17 Degradation of A

The ScbA protein (A) degrades.

**Chemical equation**

\[ A \rightarrow \emptyset \]

**Rate equation**

\[ r = d_A \cdot [A] \]

**Parameters**

The parameter of this reaction is the degradation rate of A \((d_A)\). The parameter values were derived from proteomics studies on different proteins of *S. coelicolor*.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models [1] [2]</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_A)</td>
<td>(5.02 \cdot 10^{-4}) (-0.00578min^{-1}) ([3] [4] [5])</td>
<td>(min^{-1})</td>
<td>(0.0018s^{-1}) ((0.108min^{-1})[1]) (0.066s^{-1}) ((3.96min^{-1})[2])</td>
<td>In a quantitative proteomics study on protein turnover rates in dynamic systems, Jayapal et al. reported the degradation rates of 115 proteins in <em>S. coelicolor</em> cultures undergoing transition from exponential growth to stationary phase. The values were in the range (0.05 \pm 0.014h^{-1} - 0.223 \pm 0.031h^{-1}) ((8.3 \pm 2.3 \cdot 10^{-4} - 0.0037 \pm 0.00052min^{-1})) with a median of (0.097h^{-1}(0.00162min^{-1})).</td>
</tr>
</tbody>
</table>
Appendix D. Parameter information for the GBL system model

Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are estimated or calculated from proteomics experiments, using methods which were prone to measurement errors (i.e. optical density...
of fragmented mycelia to estimate the growth rate ($\mu$) and extrapolations from cell density to total cellular protein without taking into account the dilution due to cell division. Additionally, the study did not include measurements on ScbR protein specifically, although it was conducted on a wide range of *S. coelicolor* proteins which enables a rough estimation of the range of values of protein degradation in this species. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution. By assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distribution was designed.

The mode of the log-normal distribution of the $d_A$ to $0.00144 min^{-1}$. In order to explore the full range of published values and even sample a percentage of values outside the reported range, the Spread is set to 1.78, so that the range where 68.27% of the values are found is between $8 \cdot 10^{-4}$ and $0.0026 min^{-1}$.

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

![Distribution for $d_A$ values](image)

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines).

The parameter information of the distribution is:
Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_A$</td>
<td>0.00144</td>
<td>1.78</td>
<td>$-6.2837$</td>
<td>0.50981</td>
</tr>
</tbody>
</table>

References


Appendix D. Parameter information for the GBL system model

D.18 Degradation of C

The SCB protein (C) degrades.

Chemical equation

\[ C \rightarrow \emptyset \]

Rate equation

\[ r = d_C \cdot |C| \]

Parameters

The parameter of this reaction is the degradation rate of C \((d_C)\). The parameter values were derived from measurements and estimations of autoinducer degradation rates from other bacteria (AHL, PAI2).

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models ([1][2])</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_C)</td>
<td>(7.68 \cdot 10^{-5} - 2 \cdot 10^{-2}) ([3][4][5][6])</td>
<td>(min^{-1})</td>
<td>(6.7 \cdot 10^{-5} s^{-1}) ((0.00402 min^{-1}))</td>
<td>Chen et al. empirically calculated the degradation rate of the autoinducer PAI2 in <em>Pseudomonas aeruginosa</em> cultures and reported a best-fit degradation constant of (0.195 h^{-1}(0.00325 min^{-1})). Additionally, the degradation rates of the quorum sensing autoinducer AHL have been measured \textit{in vitro} by Kaufmann et al. and the reported rates for different AHLs are between (1.28 \cdot 10^{-6} - 3.07 \cdot 10^{-5} s^{-1}) ((7.68 \cdot 10^{-5} - 1.84 \cdot 10^{-2} min^{-1})). Weber et al. also reported that the degradation rate of AHL \textit{in vivo} has been estimated, and is within the range (5 \cdot 10^{-3} - 2 \cdot 10^{-2} min^{-1}).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Range tested: (0 - 2 \cdot 10^{-4} s^{-1}) ((0 - 0.012 min^{-1}))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bistability range: (0 - 0.0085 s^{-1}) ((0 - 0.51 min^{-1}))</td>
<td></td>
</tr>
</tbody>
</table>
Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are either calculated or estimated based on \textit{in vitro} and \textit{in vivo} experiments in different bacteria species. Additionally, the values correspond to various autoinducers but not to SCBs. This means that there might be a notable difference between actual parameter values and the ones reported in literature. However, from the published information we can safely assume that SCBs are very stable molecules and their degradation rate will not exceed the maximum rate published for the other autoinducers. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution. By assigning the
Appendix D. Parameter information for the GBL system model

appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distribution was designed.

Thus, the weight of the distribution is put to $0.0032 \text{min}^{-1}$ which is set as the mode of the log-normal distribution for the $d_C$ and the Spread is 5.1, so the range where 68.27% of the values are found is between $6 \cdot 10^{-4}$ and $0.016 \text{min}^{-1}$. Thus, it will be possible to also explore potentially slower degradation rates.

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using a default value of 10% error (orange lines).

The parameter information of the distribution is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_C$</td>
<td>0.0032</td>
<td>5.1</td>
<td>-4.6234</td>
<td>1.0539</td>
</tr>
</tbody>
</table>

References


Appendix D. Parameter information for the GBL system model

D.19 Degradation of $C_2 \cdot R_2$

The SCB-ScbR protein complex ($C_2 \cdot R_2$) degrades.

**Chemical equation**

$$C_2 \cdot R_2 \rightarrow \emptyset$$

**Rate equation**

$$r = d_{CR} \cdot [C_2 \cdot R_2]$$

**Parameters**

The parameter of this reaction is the degradation rate of $C_2 \cdot R_2$ ($d_{CR}$). The parameter values were derived from proteomics studies on different proteins of *S. coelicolor*.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models [1] [2]</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{CR}$</td>
<td>$5.02 \cdot 10^{-4}$</td>
<td>$min^{-1}$</td>
<td>$0.062 s^{-1}$ (3.72 min$^{-1}$)[1]</td>
<td>In a quantitative proteomics study on protein turnover rates in dynamic systems, Jayapal et al. reported the degradation rates of 115 proteins in <em>S. coelicolor</em> cultures undergoing transition from exponential growth to stationary phase. The values were in the range $0.05 \pm 0.014 h^{-1} - 0.223 \pm 0.031 h^{-1}$ ($8.3 \pm 2.3 \cdot 10^{-4} - 0.0037 \pm 0.00052 min^{-1}$) with a median of $0.097 h^{-1} (0.00162 min^{-1})$.</td>
</tr>
</tbody>
</table>
Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are estimated or calculated from proteomics experiments, using methods which were prone to measurement errors (i.e. optical density...
Appendix D. Parameter information for the GBL system model

of fragmented mycelia to estimate the growth rate ($\mu$)) and extrapolations from cell density to total cellular protein without taking into account the dilution due to cell division. Additionally, the study did not include measurements on ScbR protein specifically, although it was conducted on a wide range of *S. coelicolor* proteins which enables a rough estimation of the range of values of protein degradation in this species. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution. By assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distribution was designed.

The mode of the log-normal distribution of the $d_{CR}$ to $0.00144\text{min}^{-1}$. In order to explore the full range of published values and even sample a percentage of values outside the reported range, the Spread is set to 1.78, so that the range where 68.27% of the values are found is between $8 \cdot 10^{-4}$ and $0.0026\text{min}^{-1}$.

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

![Distribution for $d_{CR}$ values](image)

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines).

The parameter information of the distribution is:
Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{CR}$</td>
<td>0.00144</td>
<td>1.78</td>
<td>$-6.2837$</td>
<td>0.50981</td>
</tr>
</tbody>
</table>

References


D.20  Degradation of AR$_2$

The ScbR-ScbA protein complex (AR$_2$) degrades.

Chemical equation

$$AR_2 \rightarrow \emptyset$$

Rate equation

$$r = d_{AR} \cdot [AR_2]$$

Parameters

The parameter of this reaction is the degradation rate of AR$_2$ ($d_{AR}$). The parameter values were derived from proteomics studies on different proteins of *S. coelicolor*.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models [1] [2]</th>
<th>Remarks-Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{AR}$</td>
<td>$5.02 \cdot 10^{-4}$</td>
<td>$min^{-1}$</td>
<td>$0.062s^{-1}$</td>
<td>In a quantitative proteomics study on protein turnover rates in dynamic systems, Jayapal et al. reported the degradation rates of 115 proteins in <em>S. coelicolor</em> cultures undergoing transition from exponential growth to stationary phase. The values were in the range $0.05 \pm 0.014h^{-1} - 0.223 \pm 0.031h^{-1}$ ($8.3 \pm 2.3 \cdot 10^{-4} - 0.0037 \pm 0.00052min^{-1}$) with a median of $0.097h^{-1}(0.00162min^{-1})$.</td>
</tr>
</tbody>
</table>

$$-0.00578min^{-1}$$

Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are estimated or calculated from proteomics experiments, using methods which were prone to measurement errors (i.e. optical density
Appendix D. Parameter information for the GBL system model

of fragmented mycelia to estimate the growth rate ($\mu$)) and extrapolations from cell density to total cellular protein without taking into account the dilution due to cell division. Additionally, the study did not include measurements on ScbR protein specifically, although it was conducted on a wide range of *S. coelicolor* proteins which enables a rough estimation of the range of values of protein degradation in this species. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution. By assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distribution was designed.

The mode of the log-normal distribution of the $d_{AR}$ to $0.00144\, \text{min}^{-1}$. In order to explore the full range of published values and even sample a percentage of values outside the reported range, the Spread is set to 1.78, so that the range where 68.27% of the values are found is between $8 \cdot 10^{-4}$ and $0.0026\, \text{min}^{-1}$.

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using the reported experimental error (green lines) or a default value of 10% error (orange lines).

The parameter information of the distribution is:
Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{AR}$</td>
<td>0.00144</td>
<td>1.78</td>
<td>−6.2837</td>
<td>0.50981</td>
</tr>
</tbody>
</table>

References


Appendix D. Parameter information for the GBL system model

D.21 Degradation of \( r \cdot a \)

The mRNA complex of \( scbR-scba \) degrades very soon after its formation.

**Chemical equation**

\[
 r \cdot a \rightarrow \emptyset
\]

**Rate equation**

\[
 r = d_{mRA} \cdot [r \cdot a]
\]

**Parameters**

The parameter of this reaction is the degradation rate of the \( r \cdot a \) complex (\( d_{mRA} \)).

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL model ([1])</th>
<th>Remarks-Reference</th>
</tr>
</thead>
</table>
| \( d_{mRA} \) | 0.35 – 69.3 [2] [3] | \( min^{-1} \) | 0.01s\(^{-1}\) (0.6\(min^{-1} \)) [1] | In a study on the properties of antisense RNA and on the effect it has on gene regulation in bacteria, Georg et al. report that the antisense RNAs resulting from overlapping or divergent promoters (i.e. isiA/IsrR sense/antisense pair in *Synechocystis* PCC6803) that form a complex, co-degrade almost immediately.  
  Georg et al. 2012 [3]  
This opinion is further supported by A.J. Carposis who, in a review on gene regulation by antisense RNA and quantifies a short mRNA half-life as approximately 2\(min\). |
Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are either calculated with approximation from experiments corresponding to different antisense RNA degradation rates of various bacteria species and are mostly a quantification of our understanding of immediate degradation. This means that there might be a difference between the actual parameter values and the ones reported in literature, although the reported values in literature suggest that the upper bound for the half-life of such a complex would not exceed 2 minutes. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution. By assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distribution was designed.

Therefore, in order to keep the parameter values as strictly as possible within the range of the values reported in literature, the prediction interval in this case for the definition of the distribution is set to 95.45% rather than 68.27%. Therefore, the mode
of the log-normal distribution of the $d_{mRA}$ is set to $4.9\text{min}^{-1}$ and the Spread is set to 14.1, which means that the range where 95.45\% of the values are found is between 0.35 and 69.3\text{min}^{-1}.

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using a default value of 10\% error (orange lines).

The parameter information of the distribution is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{mRA}$</td>
<td>4.9</td>
<td>14.1</td>
<td>2.5669</td>
<td>0.98835</td>
</tr>
</tbody>
</table>

References


D.22 Diffusion of C and $C_e$

The SCB1 protein (C) diffuses from each cell to the environment and back.

**Chemical equation**

$$C \xleftarrow{D} \xrightarrow{r \cdot D} C_e$$

**Rate equation**

The rate equation is different for the two reactions in order to express the dynamics of the cell density in the environment. Therefore, an additional term $\rho = \frac{V_{cell}}{V_{ext}}$ is added in the rate of diffusion from the external environment to the cell. This term represents the additional propensity given to the diffusion of extracellular SCBs by the accumulation of cells in the environment and the consequent increase of cell density. The two rate laws are therefore formed as follows:

$$r_{out} = D \cdot ([C_e] - [C])$$
$$r_{in} = \rho \cdot D \cdot ([C] - [C_e])$$

**Parameters**

The parameter of this reaction is the diffusion rate of C ($D$). In bacteria communication, the autoinducer can either passively diffuse in and out of the cell or, in cases of larger signalling molecules, can be actively transported through the cell membrane. Generally, the diffusion rates when driven by active processes are four orders of magnitude smaller than the passive diffusion rates of small molecules [1]. In this model, active transport is not explored, as SCBs are small molecules that can easily diffuse though the cell membrane. However, transport driven by SCB concentration differences due to increase of the cell density in the environment is taken into account. This means that as cell density increases, more autoinducer molecules enter the cell from the outside. The parameter values were derived from measurements and estimations of different autoinducers’ passive diffusion and active transport rates in bacteria.
Appendix D. Parameter information for the GBL system model

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Units</th>
<th>Value in previous GBL models [2] [3]</th>
<th>Remarks/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>0.1 – 24</td>
<td>min$^{-1}$</td>
<td>$8.3 \cdot 10^{-2} s^{-1}$ (4.98 min$^{-1}$)</td>
<td>Kaplan et al. conducted experiments on how AHL exits and enters the cell in <em>Vibrio fischeri</em> and reported that the equimolar internal and external concentrations were established within 20s (estimated rate ~3 min$^{-1}$), suggesting a very fast diffusion rate. If the signal concentration in the medium is sufficiently large, the time for intracellular concentration to reach equilibrium with the extracellular, is $\frac{D}{20} s^{-1} = 3$ min$^{-1}$.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0 – 24 min$^{-1}$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bistability range: 0.02 – 0.097 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.2 – 5.82 min$^{-1}$)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>and $8.3 \cdot 10^{-6}$ – 4.2 s$^{-1}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.000498 – 252 min$^{-1}$)</td>
<td></td>
</tr>
</tbody>
</table>

Goryachev et al. have published a stochastic model on quorum sensing in *Agrobacterium tumefaciens*, where they investigated the role of active transport and passive diffusion of the autoinducer AAI in and out of the cell. The values used for their simulations were 0.1 – 0.4 s$^{-1}$ (6 – 24 min$^{-1}$) for the passive diffusion and 0.02 – 0.6 s$^{-1}$ (1.2 – 36 min$^{-1}$) for the active transport. An estimation within the same range was made by Groisman et al. in a study using a microfluidic chemostat to analyze the cell response to an exogenously added autoinducer, where the reported time span for diffusive exchange of small molecules between different compartments was 40 s. Therefore, the resulting diffusion rate constant is $\frac{1}{40} s^{-1} = 0.025$ min$^{-1}$.

### Kinetic Constants

<table>
<thead>
<tr>
<th>Kinetic Constant</th>
<th>Legend</th>
<th>Value/Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$</td>
<td>production of AAI by TiO$_2$</td>
<td>0.02–0.4 s$^{-1}$</td>
</tr>
<tr>
<td>$k_2$</td>
<td>passive diffusion of AAI through the cell wall</td>
<td>0.1–0.4 s$^{-1}$</td>
</tr>
<tr>
<td>$k_3$</td>
<td>active transport of AAI across membrane</td>
<td>0.02–0.6 s$^{-1}$</td>
</tr>
</tbody>
</table>

Supplementary material

(http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0010037)

-Goryachev et al. 2005

Groisman et al. 2005

Finally, Weber et al. reported some cases of active transport where the diffusion rates are much slower, between 0.01 and 0.1 min$^{-1}$.

actively transported in and out of the cell. For example, in the bacterial species *Pseudomonas aeruginosa*, C4-HSL can freely diffuse but C12-HSL, a larger signaling molecule, is subjected to active influx and efflux at rates of $–10^{-6}$ min$^{-1}$ and $–10^{-1}$ min$^{-1}$ respectively [55].

Weber et al. 2011
Parameters with uncertainty

When deciding how to describe the uncertainty for this parameter we must take into consideration that the reported values are either calculated or estimated based on *in vitro* and *in vivo* experiments in different bacteria species. Additionally, the values correspond to various autoinducers but not to SCBs. This means that there might be a notable difference between actual parameter values and the ones reported in literature. However, the experiments conducted set the general context of diffusion rates for smaller and larger molecules. Therefore, it is safe to assume that GBLs will exhibit a similar behaviour to other small autoinducer molecules. These facts influence the quantification of the parameter uncertainty and therefore the shape of the corresponding distribution. By assigning the appropriate weights to the parameter values and using the method described in Chapter 3, the appropriate probability distributions were designed.

For this reason, the weight of the distribution is put to $2.8\, \text{min}^{-1}$ which is set as the mode of the log-normal distribution for the $D$ and the Spread is $4.27$. This means that the range where $68.27\%$ of the values are found is between $0.656$ and $11.95\, \text{min}^{-1}$, which also explores a part outside the reported values but also noting the risk that the marginal values would better fit active transport or larger molecules.

The probability distribution for the parameter, adjusted accordingly in order to reflect the above values, is the following:

![Distribution for D values](image-url)
Appendix D. Parameter information for the GBL system model

The values retrieved from literature and their weights are indicated by the blue dashed lines, and the uncertainty for each value is indicated using a default value of 10% error (orange lines).

The parameter information of the distribution is:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mode</th>
<th>Spread</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>2.8</td>
<td>4.27</td>
<td>1.9947</td>
<td>0.98233</td>
</tr>
</tbody>
</table>

References


D.23 Degradation of $C_e$

The extracellular SCB1 protein ($C_e$) degrades.

**Chemical equation**

$$C_e \rightarrow \varnothing$$

**Rate equation**

$$r = d_C \cdot [C_e]$$

**Parameters**

The parameter of this reaction is the degradation rate of $C_e$ ($d_C$), which is assumed to be the same as the degradation of the cellular SCB1 ($C$) in this model. Detailed information on this parameter can be found in Section D.18.
Bibliography


