*E. coli* (iJO1366) "Arginine and Proline Metabolism" Subsystem

Author: H. Scott Hinton, Utah State University

**INTRODUCTION**

The purpose of this tutorial is to review the basic structure and capabilities of the "Arginine and Proline Metabolism" subsystem of the iJO1366 *E. coli* model.

**MATERIALS**

This tutorial is based on the *Constraint-Based Reconstruction and Analysis* (COBRA) Toolbox [1,2]. To use this tutorial requires the 2016a or newer version of Matlab (https://www.mathworks.com/) and the COBRA toolbox software that can be downloaded from https://opencobra.github.io/cobratoolbox/latest/index.html. The installation instructions and troubleshooting tips are also available on this website.

**EQUIPMENT SETUP**

Initilize the COBRA toolbox.

```
initCobraToolbox
```

> Checking if git is installed ... Done.
> Checking if the repository is tracked using git ... Done.
> Checking if curl is installed ... Done.
> Checking if remote can be reached ... Done.
> Initializing and updating submodules ... Done.
> Adding all the files of The COBRA Toolbox ... Done.
> Define CB map output... set to svg.
> Retrieving models ... Done.
> TranslateSBML is installed and working properly.
> Configuring solver environment variables ...
  - [-----] ILOG_CPLEX_PATH : --> set this path manually after installing the solver ( see instructions )
  - [-----] Gurobi_PATH: C:\gurobi701\win64\matlab
  - [-----] TOMLAB_PATH : --> set this path manually after installing the solver ( see instructions )
  - [-----] MOSEK_PATH : --> set this path manually after installing the solver ( see instructions )
  Done.
> Checking available solvers and solver interfaces ... Done.
> Setting default solvers ... Done.
> Saving the MATLAB path ... Done.
- The MATLAB path was saved in the default location.

> Summary of available solvers and solver interfaces

<table>
<thead>
<tr>
<th>Support</th>
<th>LP</th>
<th>MILP</th>
<th>QP</th>
<th>MIQP</th>
<th>NLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>cplex_direct</td>
<td>full</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dqqMinos</td>
<td>full</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>glpk</td>
<td>full</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>gurobi</td>
<td>full</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ibm_cplex</td>
<td>full</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>matlab</td>
<td>full</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>mosek</td>
<td>full</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>pdcq</td>
<td>full</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>quadMinos</td>
<td>full</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>tomlab_cplex</td>
<td>full</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>qpn</td>
<td>experimental</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>tomlab_snop</td>
<td>experimental</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>gurobi_mex</td>
<td>legacy</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>lindo_old</td>
<td>legacy</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>lindo_legacy</td>
<td>legacy</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>lp_solve</td>
<td>legacy</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>opti</td>
<td>legacy</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Total: 5 2 3 1 2

+ Legend: - = not applicable, 0 = solver not compatible or not installed, 1 = solver installed.

> You can solve LP problems using: 'glpk' - 'gurobi' - 'matlab' - 'pdcq' - 'lp_solve'
> You can solve MILP problems using: 'glpk' - 'gurobi'
> You can solve QP problems using: 'gurobi' - 'pdcq' - 'qpn'
> You can solve MIQP problems using: 'gurobi'
> You can solve NLP problems using: 'matlab' - 'pdcq'

> Checking for available updates ...
> There are 3494 new commit(s) on <master> and 0 new commit(s) on <develop> [5a1c95 @ master]
> You can update The COBRA Toolbox by running updateCobraToolbox() (from within MATLAB).

Select the optimizer solver.

```
% changeCobraSolver('glpk','all');
changeCobraSolver('gurobi17','all');
```

> CBT_LP_SOLVER has been set to gurobi.
> CBT_MILP_SOLVER has been set to gurobi.
> CBT_QP_SOLVER has been set to gurobi.
> CBT_MIQP_SOLVER has been set to gurobi.
> CBT_NLP_SOLVER has been set to gurobi.
% changeCobraSolver('tomlab_cplex','all');
% changeCobraSolver('gurobi6','all');

Load the *E.coli* iJO1366 model.

load('iJO1366.mat');
saved_model = iJO1366;
model = saved_model;

**PROCEDURE**

1. Arginine and Proline Metabolism Subsystem

The purpose of this tutorial is to identify and review the structure and capabilities of the "Arginine and Proline Metabolism" subsystem of the *E.coli* iJO1366 model. This will begin with an overview of the complete subsystem. This overview will be followed by more detailed descriptions of the individual L-Arginine and L-Proline biosynthesis pathways. It will conclude with a simulation that shows the maximum flux that each these amino acids can produce in an oxidative range from anaerobic to aerobic conditions.

The reactions associated with the "Arginine and Proline Metabolism" subsystem can be extracted from the model as shown below.

model = saved_model;
arginineProlineSubSystems = {'Arginine and Proline Metabolism'};
arginineProlineReactions = model.rxns(ismember(model.subSystems,arginineProlineSubSystems));
[tmp,arginineProline_rxnID] = ismember(arginineProlineReactions,model.rxns);
reactionNames = model.rxnNames(arginineProline_rxnID);
reactionFormulas = printRxnFormula(model,arginineProlineReactions,0);
% T = table(reactionNames,reactionFormulas,'RowNames',arginineProlineReactions)
fid = 1;
fprintf(fid,'%12s %7s %50s
','Reaction','Reaction Name','Reaction Formula');

[nrows,ncols] = size(arginineProlineReactions);
for row = 1:nrows
  fprintf(fid,'%12s %7s %50s
',arginineProlineReactions(row,:),reactionNames(row,:),reactionFormulas(row,:));
end

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Reaction Name</th>
<th>Reaction Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABTA</td>
<td>4-aminobutyrate transaminase</td>
<td>4abut_c + akg_c -&gt; glu__l_c + succal_c</td>
</tr>
<tr>
<td>ABUTD</td>
<td>Aminobutyraldehyde dehydrogenase</td>
<td>4abutn_c + h2o_c + nad_c -&gt; 4abut_c + 2 h_c + nadh_c</td>
</tr>
<tr>
<td>ACBK</td>
<td>Acetylglutamate kinase</td>
<td>acglu_c + atp_c -&gt; acg5p_c + adp_c</td>
</tr>
<tr>
<td>ACBS</td>
<td>N-acetylglutamate synthase</td>
<td>accoa_c + glu___l_c -&gt; acglu_c + coa_c + h_c</td>
</tr>
<tr>
<td>ACODA</td>
<td>Acetylornithine deacetylase</td>
<td>acorn_c + h2o_c -&gt; ac_c + orn_c</td>
</tr>
<tr>
<td>ACOTA</td>
<td>Acetylornithine transaminase</td>
<td>acorn_c + akg_c &lt;= acg5sa_c + glu___l_c</td>
</tr>
<tr>
<td>AMDMC</td>
<td>Adenosylmethionine decarboxylase</td>
<td>amet_c + h_c -&gt; ametam_c + co2_c</td>
</tr>
<tr>
<td>AGMT</td>
<td>Agmatinase</td>
<td>agm_c + h2o_c -&gt; ptrc_c + urea_c</td>
</tr>
<tr>
<td>AGPR</td>
<td>N-acetyl-g-glutamyl-phosphate reductase</td>
<td>acg5sa_c + nadp_c + pi_c &lt;= acg5p_c + h_c + nadph_c</td>
</tr>
<tr>
<td>AMPTASEPG</td>
<td>Aminopeptidase (pro-gly)</td>
<td>h2o_c + progly_c -&gt; gly_c + pro___l_c</td>
</tr>
<tr>
<td>APCS</td>
<td>Aminopropylcadaverine synthase</td>
<td>15dap_c + ametam_c -&gt; 5mta_c + h_c + na15dap_c</td>
</tr>
<tr>
<td>ARGDC</td>
<td>Arginine decarboxylase</td>
<td>arg__l_c + h_c -&gt; agm_c + co2_c</td>
</tr>
</tbody>
</table>
Table 1. Reaction names and formulas for the reactions associated with the "Arginine and Proline Metabolism" subsystem.

The connectivity between these reactions can be visualized through an Escher [3] map of this "Arginine and Proline Metabolism" subsystem as shown below.
Figure 1. "Arginine and Proline Metabolism" subsystem which includes Putrescine and Spermidine biosynthesis pathways (Proline_Arginine_Subsystem.json, Proline_Arginine_Subsystem.png, Proline_Arginine_Subsystem.svg).
This subsystem not only includes the biosynthesis pathways for L-Arginine and L-Proline but also includes the biosynthesis pathways for Putrescine ('ptrc_c') and Spermidine ('spmd_c'). Putrescine and Spermidine are precursors to polyamines and will not be discussed in more detail in this tutorial. This subsystem also includes the pathways that can be used to convert L-Arginine ('arg__L_c') to L-Glutamate ('glu__L_c') and Succinate ('succ_c'). The details of the biosynthesis pathways of L-Arginine and L-Proline will be discussed later.

Let's look at the active reactions in the "Arginine and Proline Metabolism" subsystem under aerobic conditions.

```python
model = saved_model;
model = changeRxnBounds(model, 'EX_glc__D_e\', -10,'l'); % Set maximum glucose uptake
model = changeRxnBounds(model, 'EX_o2_e\',-30,'l'); % Set oxygen uptake
model = changeRxnBounds(model, 'BIOMASS_Ec_iJO1366WT_53p95M', -0,'b'); % Disable WT biomass reaction
model = changeObjective(model, 'BIOMASS_Ec_iJO1366_core_53p95M'); % Set the objective function
FBA_solution = optimizeCbModel(model,'max',0,0); % Perform FBA
printLabeledData(arginineProlineReactions, round(FBA_solution.x(arginineProline_rxnID),3))
```
Table 2. Flux values for the reactions in the "Arginine and Proline Metabolism" subsystem under aerobic conditions.

From this it can be seen that the primary L-Arginine pathway begins with L-Glutamate and is composed of the following reactions; 'ACGS', 'ACGK', 'AGPR', 'ACOTA', 'ACODA', 'OCBT', 'ARGSS', and 'ARGSL'. These are the reactions that have the flux value of +/- 0.291 mmol \cdot gDW^{-1} \cdot hr^{-1}. On the other hand, the primary L-Proline pathway starts with L-Glutamate and is composed of the following reactions; 'GLU5K', 'G5SD', 'G5SADs', and 'P5CR'. These are the reactions that have the flux value of +/- 0.217 mmol \cdot gDW^{-1} \cdot hr^{-1}.

Below is an Escher [3] figure of the "Arginine and Proline Metabolism" subsystem in an aerobic condition using "Aerobic_Reaction_Flux.csv" as the data file.
Figure 2. The flux flowing through the "Arginine and Proline Metabolism" subsystem under aerobic conditions. The thickness of a line is proportional to the amount of flux flowing through a reaction, the "red" lines correspond to reactions operating in the forward direction while the "blue" lines correspond to reactions operating in the reverse direction (Arginine_Proline_Subsystem_Aerobic.png and Arginine_Proline_Subsystem_Aerobic.svg).
This visualization shows that the primary L-Arginine pathway begins with L-Glutamate and is composed of the following reactions; 'ACGS', 'ACGK', 'AGPR', 'ACOTA', 'ACODA', 'OCBT', 'ARGSS', and 'ARGSL'. On the other hand, the primary L-Proline pathway starts with L-Glutamate and is composed of the following reactions; 'GLU5K', 'G5SD', 'G5SADs', and 'P5CR'.

Now let’s look at the different biosynthesis pathways in more detail.

3. L-Proline Biosynthesis

The chemical structure for L-Proline ($C_3H_7NO_2$) is shown below.

![Chemical structure of L-Proline](image)

**Figure 3.** The chemical structure of L-Proline ($C_3H_7NO_2$) - Wikipedia

3.1 L-Proline Biosynthesis Pathways

Let’s begin by identifying the reactions that directly produce L-Proline.

```bash
surfNet(model, 'pro__L_c')
```

Met #834  pro__L_c, L-Proline, C5H9NO2
Consuming reactions:

- #7  BIOMASS_Ec_iJO1366_growth_with_53_carbon_sources  E. coli biomass objective function (iJO1366) - WT - with 53.95 GAM estimate
  
  ```bash
  0.000223 18fthf_c + 0.000223 2dmq18_c + 2.5e-05 2fe2s_c +
  0.000284 4fe4s_c + 0.000223 5athc_c + 0.000279 accoa_c + 0.000223 adocl2_c
  + 0.49915 ala__L_c + 0.000223 amet_c + 0.28742 arg__L_c + 0.23423 asp__L_c
  + 0.23423 asp__L_c + 54.12 atp_c + 0.00016 bmocgdpp_c + 2e-06 btm_c
  + 0.004952 ca2_c + 0.000223 chor_c + 0.004952 ci_c + 0.000168 coa_c
  + 2.4e-05 cobalt2_c + 0.1298 ctp_c + 0.000674 cu2_c + 0.08988 cys__L_c
  + 0.024805 datp_c + 0.000261 dctp_c + 0.000261 dgtp_c + 0.034805 dtpp_c
  + 0.000223 dhtp_c + 0.000223 fad__C_c + 0.006388 fe2_c + 0.000742 fe3_c
  + 0.25571 gln__L_c + 0.25571 glu__L_c + 0.5953 gly_c + 0.15419 glycogen_c
  + 0.000223 gthrd_c + 0.20912 gtp_c + 48.7529 h2o_c + 0.000223 heme0_c
  + 0.02056 his__L_c + 0.28231 1le__L_c + 0.18569 k_c + 0.43778 leu__L_c
  + 3e-06 lipo_pb_c + 0.33345 lys__L_c + 3.1e-05 malcoa_c + 0.14934 met__L_c
  + 0.008253 mg2_c + 0.000223 mlthf_c + 0.000858 mn2_c + 7e-06 mobd_c
  + 7e-06 mocogdp_c + 7e-06 mocogdp_c + 0.000223 mql8_c + 0.001787 nad_c
  + 4.5e-05 nadh_c + 0.000112 naph_c + 0.000835 nadph_c + 0.012379 nh4_c
  + 0.000307 ni2_c + 0.012366 pe160_c + 0.09618 pe161_c + 0.04957 pe181_c
  + 0.005797 pg160_c + 0.004439 pg161_c + 0.002288 pg181_c + 0.18002 phe__L_c
  + 0.000223 phecoa_c + 0.2148 pro__L_c + 0.03327 ptrc_c + 0.000223 pydx5p_c
  + 0.000223 q8h2_c + 0.000223 ribf1v_c + 0.20968 ser__L_c + 0.000223 sheme_c
  + 0.004126 so4_c + 0.006744 spmd_c + 9.8e-05 succoa_c + 0.000223 thf_c
  + 0.000223 thmpc_c + 0.24651 thr__L_c + 0.055234 trp__L_c + 0.13399 tyr__L_c
  ```
Table 3. L-Proline consuming and producing reactions.

The reactions that directly produce L-Proline include 'AMPTASEPG', 'P5CR', 'PROabcpp', 'PROt2rpp', and 'PRO4pp', and 'PROGLYabcpp'. The first reaction ('AMPTASEPG') converts L-Prolylglutamine to L-Proline. The second reaction ('P5CR') is the final reaction in the primary pathway from L-Glutamate to L-Proline. The
transporter reactions (‘PROabcpp’, ‘PROt2rpp’, and ‘PROt4pp’) allow the transport of L-Proline between the periplasm and the cytoplasm. The first of these transporters is an ABC transporter that requires ATP. The second transporter requires energy from the proton-motive force, while the third reaction is a sodium/L-Proline symporter. L-Proline is allowed to diffuse from the extracellular space to the periplasm through the reaction ‘PROt Knox’. The final reaction is an ABC transporter that moves diffused L-Prolinylglycine from the periplasm to the cytoplasm. L-Prolinylglycine is diffused from the extracellular space into the periplasm through the reaction ‘PROGLYt Knox’ (‘EX_progly e’).

In addition to the primary L-Proline biosynthesis pathways there is an alternate pathway that also begins with L-glutamate and includes the reactions ‘ACGS’, ‘ACGK’, ‘AGPR’, and ‘NACODA’. This pathway requires the precursor acetyl-CoA.

Now by combining the L-Proline pathways seen in Figure 1, with the reactions associated with alternate methods of producing L-Proline described above, we can define all the reactions that can lead to the biosynthesis and/or production of L-Proline. These reactions include; ‘GLU5K’, ‘G5SD’, ‘G5SADs’, ‘P5CR’, ‘ACGS’, ‘ACGK’, ‘AGPR’, ‘NACODA’, ‘AMPTASEPG’, ‘PROGLYabcpp’, ‘PROGLYt Knox’, ‘EX_progly e’, ‘PROabcpp’, ‘PROt2rpp’, and ‘PROt4pp’, ‘PROt Knox’ and ‘EX_pro l_e’. These pathways for the production of L-Proline are shown in the following Escher-based figure.

![Figure 4. Biosynthesis pathways for L-Proline (Proline_Biosynthesis.json, Proline_Biosynthesis.png, Proline_Biosynthesis.svg).](image)

In the figure, we can see the primary biosynthesis pathway (‘GLU5K’, ‘G5SD’, ‘G5SADs’, ‘P5CR’) and alternate pathway (‘ACGS’, ‘ACGK’, ‘AGPR’, ‘NACODA’) that begins with L-Glutamate. There are two pathways that begin with the extracellular metabolites, L-Proline and L-Prolinylglycine. Extracellular L-Proline can be brought into cytoplasm with ‘PROabcpp’, ‘PROt2rpp’, and ‘PROt4pp’, ‘PROt Knox’ and ‘EX_pro l_e’ while the Extracellular L-Prolinylglycine pathways includes ‘AMPTASEPG’,
'PROGLYabcpp', 'PROGLYtexp', 'EX__pro__L_e'

Note that this figure also includes the key exchange reactions ('EX__pro__L_e', and 'EX__pro__L_e') associated with the metabolites that diffuse into the periplasm.

The formulas for these L-Proline producing and consuming reactions can be found as shown below.

```
model = saved_model;
ProlineBiosynthesisReactions = transpose({'GLUSK', 'GSSD', 'GSSADs', 'PSCR', 'AGCS', 'ACGK', 'AGPR', 'NACODA', ...
    'AMPTASEPG', 'PROGLYabcpp', 'PROGLYtexp', 'EX__pro__L_e', 'PROabcpp', 'PROt2rpp', 'PROt4pp', 'PROtext', 'EX__pro__L_e'});
[tmp,ProlineBiosynthesis_rxnID] = ismember(ProlineBiosynthesisReactions, model.rxns);
reactionNames = model.rxnNames(ProlineBiosynthesis_rxnID);
reactionFormulas = printRxnFormula(model, ProlineBiosynthesisReactions,0);
% T = table(reactionNames, reactionFormulas, 'RowNames', ProlineBiosynthesisReactions)
 fid = 1;
fprintf(fid, '%-18s %6s %5s
', 'Reaction', 'Reaction Name', 'Reaction Formula');

[nrows,ncols] = size(ProlineBiosynthesisReactions);
for row = 1:nrows
    fprintf(fid, '%-18s %6s %5s
', ProlineBiosynthesisReactions(row,:), reactionNames{row,:}, reactionFormulas{row,:});
end
```

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Reaction Name</th>
<th>Reaction Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLUSK</td>
<td>Glutamate 5-kinase</td>
<td>ATP_c + Glu_L_c -&gt; Adp_c + Glu5p_c</td>
</tr>
<tr>
<td>GSSD</td>
<td>Glutamate-5-semialdehyde dehydrogenase</td>
<td>Glu5p_c + H_c + NADPH_c -&gt; Glu5sa_c + NADP_c + Pi_c</td>
</tr>
<tr>
<td>GSSADs</td>
<td>L-glutamate 5-semialdehyde dehydratase (spontaneous)</td>
<td>Glu5sa_c -&gt; 1Pyr5c_c + H_c + H2O_c</td>
</tr>
<tr>
<td>PSCR</td>
<td>Pyrroline-5-carboxylate reductase</td>
<td>1Pyr5c_c + 2 H_c + NADH_c -&gt; NADP_c + Pro__L_c</td>
</tr>
<tr>
<td>AGCS</td>
<td>N-acetylglutamate synthase</td>
<td>AcCoa_c + Glu_L_c -&gt; AcGlu_c + Coa_c + H_c</td>
</tr>
<tr>
<td>ACGK</td>
<td>Acetylglutamate kinase</td>
<td>AcGlu_c + ATP_c -&gt; AcG5p_c + ADP_c</td>
</tr>
<tr>
<td>AGPR</td>
<td>N-acetyl-glutamyl-phosphate reductase</td>
<td>AcG5sa_c + NADP_c + Pi_c &lt;=&gt; AcG5p_c + H_c + NADPH_c</td>
</tr>
<tr>
<td>NACODA</td>
<td>N-acetylorrhizine deacetylase</td>
<td>AcG5sa_c + H2O_c -&gt; Ac_c + Glu5sa_c</td>
</tr>
<tr>
<td>AMPTASEPG</td>
<td>Aminopeptidase (Pro-poly)</td>
<td>H2O_c + Prolyg_c -&gt; Gly_c + Pro__L_c</td>
</tr>
<tr>
<td>PROGLYabcpp</td>
<td>L-Prolineglycine (Pro-Gly) transport via ABC system (periplasm)</td>
<td>ATP_c + H2O_c + Progly_p -&gt; Adp_c + H_c + Pi_c + Progly_p</td>
</tr>
<tr>
<td>PROGLYtexp</td>
<td>L-Prolineglycine transport via diffusion (extracellular to periplasm)</td>
<td>Progly_e &lt;=&gt; Progly_p</td>
</tr>
<tr>
<td>EX__pro__L_e</td>
<td>L-Prolineglycine exchange</td>
<td>Progly_e -&gt;</td>
</tr>
<tr>
<td>PROabcpp</td>
<td>L-proline transport via ABC system (periplasm)</td>
<td>ATP_c + H2O_c + Pro__L_p -&gt; Adp_c + H_c + Pi_c + Pro__L_c</td>
</tr>
<tr>
<td>PROt2rpp</td>
<td>L-proline reversible transport via proton symporter (periplasm)</td>
<td>H_p + Pro__L_p &lt;=&gt; H_c + Pro__L_c</td>
</tr>
<tr>
<td>PROt4pp</td>
<td>Na+/Proline-L symporter (periplasm)</td>
<td>Na1_p + Pro__L_p -&gt; Na1_c + Pro__L_c</td>
</tr>
<tr>
<td>PROtext</td>
<td>L-proline transport via diffusion (extracellular to periplasm)</td>
<td>Pro__L_e &lt;=&gt; Pro__L_p</td>
</tr>
<tr>
<td>EX__pro__L_e</td>
<td>L-Proline exchange</td>
<td>Pro__L_e -&gt;</td>
</tr>
</tbody>
</table>

**Table 4. Reactions names and formulas for L-Proline biosynthesis reactions.**

We can also find the subsystems that are associated with each of these reactions.

```
reactionSubsystems = model.subSystems(ProlineBiosynthesis_rxnID);
% T = table(reactionNames, reactionSubsystems, 'RowNames', MethionineBiosynthesisReactions)  
 fid = 1;
```
Table 5. Reactions names and subsystems for L-Proline biosynthesis reactions.

3.2. L-Proline Aerobic Operation

Let's look at the aerobic response of all these reactions.

```python
model = saved_model;
model = changeRxnBounds(model,'EX_glc__D_e','10','1'); % Set maximum glucose uptake
model = changeRxnBounds(model,'EX_o2_e',-30,'1'); % Set oxygen uptake
model = changeRxnBounds(model,'BIOMASS_Ec_iJO1366_WT_53p95M',-30,'1'); % Disable WT biomass reaction
model = changeObjective(model,'BIOMASS_Ec_iJO1366_core_53p95M'); % Set the objective function
FBAsolution = optimizeCbModel(model,'max',0,0); % Perform FBA
prolineReactions = transpose(['AMPTASEPG','GLUSK','GSSD','GSSADs','PSCR','PROabcpp','PROt2rpp','PROt4pp','PROGLYabcpp',...
   'PROtex','PROGLYtex','ACGS','ACGK','AGPR','NACODA']);
[tmp,prolineReactions_rxnID] = ismember(prolineReactions,model.rxns);
printLabeledData(prolineReactions_rxnID,round(FBAsolution.x(prolineReactions_rxnID),3))
```
Table 6. Flux values for L-Proline biosynthesis reactions.

In this case, it can be seen that only the primary pathway is be used, while the secondary pathway is blocked since ‘NACODA’ has zero flux. The flux through ‘ACGS’, ‘ACGK’, and ‘AGPR’ is also a path of the L-Proline pathway.

These fluxes can be seen in the following Escher-based visualization [3].

Figure 5. Aerobic fluxes flowing through the L-Proline biosynthesis pathways. The thickness of a line is proportional to the amount of flux flowing through a reaction, the
"red" lines correspond to reactions operating in the forward direction while the "blue" lines correspond to reactions operating in the reverse direction (Proline_Biosynthesis_Aerobic.svg or Proline_Biosynthesis_Aerobic.png).

In this figure we can see that for aerobic conditions the primary pathways are used to produce L-Proline. The flux through the lower pathway ('ACGS','ACGK','AGPR') is precursor flux to L-Arginine production.

3.3. Excess L-Proline Production

When a cell is producing a recombinant protein, it might be required to produce additional L-Proline for the desired bioproduct. What is the maximum amount of L-Proline that can be produced for a given growth-rate?

```python
model = saved_model;
model = changeRxnBounds(model,'EX_glc__D_e',-10,1); % Set maximum glucose uptake
model = changeRxnBounds(model,'EX_o2_e',-30,1); % Set oxygen uptake
model = changeRxnBounds(model,'BIOMASS_Ec_iJO1366_WT_53p95M',0,1); % Disable WT biomass reaction
model = changeObjective(model,'BIOMASS_Ec_iJO1366_core_53p95M'); % Set the objective function
FBAsolution = optimizeCbModel(model,'max'); % Perform FBA to find optimal growth-rate
model = changeRxnBounds(model,'BIOMASS_Ec_iJO1366_WT_53p95M',FBAsolution.f,1); % Set fixed growth-rate
model = addDemandReaction(model,'pro__l_c');

DM_pro__l_c  pro__l_c  ->
```

```python
[tmp,phenylalanine_MAX_rxnID] = ismember({'DM_pro__l_c'},model.rxns);
model = changeObjective(model,'DM_pro__l_c'); % Set the objective function
FBAsolution_phenylalanine = optimizeCbModel(model,'max'); % Perform FBA to find optimal growth-rate
% printLabeledData({'DM_pro__l_c'}, round(FBAsolution_tyrosine.x(phenylalanine_MAX_rxnID),3))

xMin = 0.5;
xMax = FBAsolution.f;
xInc = (xMax - xMin)/20;
x = xMin;
excessProline = [];
growthRate = [];
for i = 1:21
    model = changeRxnBounds(model,'BIOMASS_Ec_iJO1366_WT_53p95M',x,1); % Set fixed growth-rate
    FBAsolution_proline = optimizeCbModel(model,'max'); % Perform FBA
    excessProline(i) = FBAsolution_proline.f;
growthRate(i) = x;
    x = x + xInc;
end
plot(growthRate,excessProline)
title('Excess L-Proline');
xlabel('Growth-rate (h-1)'); ylabel('Excess L-Proline (mmol.gDW-1.h-1)');
```
Figure 6. A plot showing the maximum amount of L-Proline that can be produced for a given growth-rate.

This figure illustrates that as the need for excess L-Proline increases the growth-rate will need to decrease. In this figure we can see that the excess L-Proline flux can increase from 0.04324 mmol · gDW⁻¹ · hr⁻¹ when the cell is at maximum growth-rate to 5.046 mmol · gDW⁻¹ · hr⁻¹ when it is at 50% of that optimal growth-rate. Finally, to increase the L-Proline flux beyond these levels will require using the pathways that allow for the transport of L-Proline from the extracellular media.

4. L-Arginine Biosynthesis

The chemical structure for L-Arginine (C₆H₁₄N₄O₂) is shown below.

![Chemical structure of L-Arginine](image)

Figure 7. The chemical structure of L-Arginine (C₆H₁₄N₄O₂) - Wikipedia

4.1 L-Arginine Biosynthesis Pathways

Let's begin by identifying the reactions that directly produce L-Arginine.
#7 BIOMASS_Ec_iJO1366_w6_53p95M E. coli biomass objective function (iJO1366) - WT - with 53.95 GAM estimate

0.000223 10fthf_c + 0.000223 2dmmq_8c + 2.5e-05 2fe25s_c + 0.000248 4fe4s_c + 0.000223 5mthf_c + 0.000279 accoa_c + 0.000223 adocbl_c
+ 0.49915 ala_l_c + 0.000223 amet_c + 0.28742 arg_l_c + 0.23423 asn_l_c + 0.23423 asp_l_c + 54.12 atp_c + 0.000116 bmocgdpg_c + 2e-06 bnn_c
+ 0.004952 ca2_c + 0.000223 chor_c + 0.000452 cl_c + 0.000168 coa_c + 2.4e-05 cobalt2_c + 0.1298 ctp_c + 0.000674 cu2_c + 0.000898 cys_l_c
+ 0.024805 datp_c + 0.025612 dctp_c + 0.025612 dgtp_c + 0.024805 dttp_c + 0.000223 enter_c + 0.000223 fad_c + 0.000638 fe2_c + 0.007428 fe3_c
+ 0.25571 gln_l_c + 0.25571 glu_l_c + 0.5953 gly_c + 0.15419 glycogen_c + 0.000223 gthrd_c + 0.20912 gtp_c + 48.7529 h2o_c + 0.000223 hemeo_c
+ 0.092056 his_l_c + 0.28231 ile_l_c + 0.18569 k_c + 0.47787 leu_l_c + 3e-06 lipobp_c + 0.33345 lys_l_c + 3.1e-05 malcoa_c + 0.14934 met_l_c
+ 0.008253 mg2_c + 0.000223 mlthf_c + 0.000658 mn2_c + 7e-06 mobd_c + 7e-06 moccodpg_c + 7e-06 moccodpg_c + 0.000223 ngl8_c + 0.001787 nad_c
+ 4.5e-05 nadh_c + 0.000112 napd_c + 0.000335 napdh_c + 0.012379 nh4_c + 0.000307 ni2_c + 0.012366 pe160_c + 0.000618 pe161_c + 0.000957 pe181_c
+ 0.000757 pg100_c + 0.000439 pg161_c + 0.002288 pg181_c + 0.18002 phe_l_c + 0.000223 pheme_c + 0.2148 pro_l_c + 0.03327 ptrc_c + 0.000223 pydx5p_c
+ 0.000223 qh2c + 0.000223 ribflv_c + 0.20968 ser_l_c + 0.000223 sheme_c + 0.004126 so4_c + 0.006744 spmd_c + 9.8e-05 succoa_c + 0.000223 thf_c
+ 0.000223 thmph_c + 0.24651 thr_l_c + 0.055234 trp_l_c + 0.13399 tyr_l_c + 5.5e-05 udcpd_c + 0.1401 utp_c + 0.41118 val_l_c + 0.000324 zn2_c
+ 0.000151 collpa_e + 0.000244 clpn160_p + 0.00229 clpn161_p + 0.00118 clpn181_p + 0.001345 murein3p3p_p + 0.000605 murein3p4p_p +
+ 0.000066 murein4p4p_p + 0.000944 murein4p4p4p_p + 0.000673 murein4p4p4p_p + 0.031798 pe160_p + 0.024732 pe161_p + 0.012747 pe181_p + 0.004892 pg160_p
+ 0.003085 pg161_p + 0.001961 pg181_p -> 53.95 adp_c + 53.95 h_c + 53.9459 pi_c + 0.74983 ppi_c

#8 BIOMASS_Ec_iJO1366_w6_53p95M E. coli biomass objective function (iJO1366) - core - with 53.95 GAM estimate

0.000223 10fthf_c + 2.6e-05 2fe25s_c + 0.000223 2ohph_c + 0.00026 4fe4s_c
+ 0.51369 ala_l_c + 0.000223 amet_c + 0.29579 arg_l_c + 0.24105 asn_l_c + 0.24105 asp_l_c + 54.12 atp_c + 0.000122 bmocgdpg_c + 2e-06 bnn_c
+ 0.00528 ca2_c + 0.00528 cl_c + 0.000576 coa_c + 2.5e-05 cobalt2_c
+ 0.13351 ctp_c + 0.000709 cu2_c + 0.01588 cys_l_c + 0.026166 datp_c + 0.027017 dctp_c + 0.027017 dgtp_c + 0.026166 dttp_c + 0.000223 fad_c
+ 0.00715 fe2_c + 0.00715 fe3_c + 0.26316 gln_l_c + 0.26316 glu_l_c + 0.61264 gly_c + 0.2151 gtp_c + 48.6015 h2o_c + 0.004738 his_l_c
+ 0.29053 ile_l_c + 0.19519 k_c + 0.45053 leu_l_c + 0.34316 lys_l_c + 0.15369 met_l_c + 0.008675 mg2_c + 0.000223 mlthf_c + 0.000691 mn2_c
+ 7e-06 mobd_c + 0.001831 nad_c + 0.000447 napd_c + 0.013813 nh4_c + 0.000323 ni2_c + 0.017868 pe160_c + 0.05454 pe161_c + 0.18527 phe_l_c
+ 0.000223 pheme_c + 0.22104 pro_l_c + 0.000223 pydx5p_c + 0.000223 ribflv_c + 0.21579 ser_l_c + 0.000223 sheme_c + 0.004338 so4_c +
+ 0.000223 thf_c + 0.000223 thmph_c + 0.25369 thr_l_c + 0.056843 trp_l_c + 0.1379 tyr_l_c + 5.5e-05 udcpd_c + 0.1441 utp_c + 0.42316 val_l_c +
0.000341 zn2_c + 0.019456 kdo2lipid_d_e + 0.013894 murein5px4_p_p + 0.045946 pe160_p
+ 0.02106 pe161_p -> 53.95 adp_c + 53.95 h_c + 53.9457 pl_c +
0.7739 ppi_c

#05 ARGDC Arginine decarboxylase
    arg__l_c + h_c -> agm_c + co2_c

#06 ArgTRS Arginyl-tRNA synthetase
    arg__l_c + atp_c + trnaarg_c -> amp_c + argtrna_c + ppi_c

#07 ArgTpp L-arginine transport out via proton antiport (cytoplasm to periplasm)
    arg__l_c + h_p -> h_c + arg__l_p

#08 AST Arginine succinytransferase
    arg__l_c + succoa_c -> coa_c + h_c + sucarg_c

Producing reactions:

#09 ARGAGM7pp Arginine/agmatine antiport (periplasm)
    agm_c + arg__l_l_p <-> arg__l_c + agm_p

#10 ARGORN7pp Arginine/ornithine antiporter (periplasm)
    orn_c + arg__l_l_p <-> arg__l_c + orn_p

#11 ARGSL Argininosuccinate lyase
    argsucc_c <-> arg__l_c + fum_c

#12 Argabcpp L-arginine transport via ABC system (periplasm)
    atp_c + h2o_c + arg__l_l_p -> adp_c + arg__l_c + h_c + pl_c

Show previous steps...

**Table 7.** L-Arginine consuming and producing reactions.

From these results, combined with Figure 1, we can see that there are four reaction that directly produce L-Arginine, ‘ARGAGM7pp’, ‘ARGORN7pp’, ‘ARGabcpp’, and ‘ARGSL’. The first three of these reactions are transporters that move periplasmic L-Arginine to the cytoplasm at the energy cost of either ‘atp_c’ or proton-motive force. The final reaction, ‘ARGSL’ is the final step in the primary biosynthesis pathway. Using the Escher tool we can explore all the paths that feed into the L-Arginine primary pathway of ‘ACGS’, ‘ACGK’, ‘AGPR’, ‘ACOTA’, ‘ACODA’, ‘OCBT’, ‘CBPS’, ‘ARGSS’, and ‘ARGSL’. These include 3 reactions (‘CBPS’, ‘CBMKr’, and ‘OXAMTC’) that provide three separate pathways for the production of Carbamoyl phosphate (‘cbp_c’) which is required by ‘OCBT’. The is also a pathway to provide additional Ornithine (‘orn_c’) from the extracellular media. All these reaction are shown in the following Escher-based figure.
In this figure, we can also note that the precursor Acetyl-CoA and the amino acids L-Glutamate and L-Aspartate are required for the biosynthesis of L-Arginine. Now let's look at the names of these reactions and their stochiometric formulas.

```r
model = saved_model;
arginineReactions = transpose(['ACGS', 'ACGK', 'AGPR', 'ACOTA', 'ACODA', 'OCBT', 'ARGSS', 'ARGSL', 'ARGabcpp', 'ARGAGmt7pp', 'ARGNrt7pp', ' ARGtex', 'EX_arg__L_e', 'CBPS', 'CBMKr', 'OXAMTC', 'ORNabcpp', 'ORNex', 'EX_orn_e']);
[tmp,arginine_rxnID] = ismember(arginineReactions,model.rxns);
reactionNames = model.rxnNames(arginine_rxnID);
```
Table 8. Reaction names and their stoichiometric formulas for reactions associated with L-Arginine biosynthesis.

We can also find the subsystems that are associated with each of these reactions.
Table 9. Reactions names and subsystems for L-Arginine biosynthesis reactions.

4.2. L-Arginine Aerobic Operation

Let’s look at the aerobic response of these reactions.

```plaintext
model = saved_model;
model = changeRxnBounds(model,'EX_glc_D_e',-10,'l'); % Set maximum glucose uptake
model = changeRxnBounds(model,'EX_o2_e',-30,'l'); % Set oxygen uptake
model = changeRxnBounds(model,'BIO MASS_Ec_iJO1366 WT_53p95M',-30,'l'); % Disable WT biomass reaction
model = changeObjective(model,'BIO MASS_Ec_iJO1366 core_53p95M'); % Set the objective function
FBA solution = optimizeCbModel(model,'max',0,0); % Perform FBA
printLabelledData(arginineReactions, round(FBA solution.x(arginine_rxnID),3))
```

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Subsystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGCS</td>
<td>N-acetylglutamate synthase</td>
</tr>
<tr>
<td>AGX</td>
<td>Acetylglutamate kinase</td>
</tr>
<tr>
<td>AGPR</td>
<td>N-acetyl-g-glutamyl-phosphate reductase</td>
</tr>
<tr>
<td>ACOTA</td>
<td>Acetylornithine transaminase</td>
</tr>
<tr>
<td>ACODA</td>
<td>Acetylornithine deacetylase</td>
</tr>
<tr>
<td>OCBT</td>
<td>Ornithine carbamoyltransferase</td>
</tr>
<tr>
<td>ARGSS</td>
<td>Argininosuccinate synthase</td>
</tr>
<tr>
<td>ARGL</td>
<td>Argininosuccinate lyase</td>
</tr>
<tr>
<td>ARGabcpp</td>
<td>L-arginine transport via ABC system (periplasm)</td>
</tr>
<tr>
<td>ARGAGMT7pp</td>
<td>Arginine/agmatine antiport (periplasm)</td>
</tr>
<tr>
<td>ARGORN7pp</td>
<td>Arginine/ornithine antiporter (periplasm)</td>
</tr>
<tr>
<td>ARGtext</td>
<td>L-arginine transport via diffusion (extracellular to periplasm)</td>
</tr>
<tr>
<td>EX_argL_e</td>
<td>L-arginine exchange</td>
</tr>
<tr>
<td>CBPS</td>
<td>Carbamoyl-phosphate synthase (glutamine-hydrolysing)</td>
</tr>
<tr>
<td>CBMk</td>
<td>Carbamate kinase</td>
</tr>
<tr>
<td>OXAMTC</td>
<td>Oxamate transcarbamoylase</td>
</tr>
<tr>
<td>ORNabcpp</td>
<td>Ornithine transport via ABC system (periplasm)</td>
</tr>
<tr>
<td>ORNtex</td>
<td>Ornithine transport via diffusion (extracellular to periplasm)</td>
</tr>
<tr>
<td>EX_orn_e</td>
<td>Ornithine exchange</td>
</tr>
</tbody>
</table>

ACGS 0.291
AGX 0.291
AGPR -0.291
ACOTA 0.291
ACODA 0.291
OCBT 0.291
ARGSS 0.291
ARGL 0.291
ARGabcpp 0
ARGAGMT7pp 0
ARGORN7pp 0
ARGtext 0
EX_argL_e 0
CBPS 0
Table 10. Flux values for L-Proline biosynthesis reactions.

These fluxes can be seen in the following Escher-based visualization [3].
Figure 9. Aerobic fluxes flowing through the L-Arginine biosynthesis pathways. The thickness of a line is proportional to the amount of flux flowing through a reaction, the "red" lines correspond to reactions operating in the forward direction while the "blue" lines correspond to reactions operating in the reverse direction (Arginine_Biosynthesis_Aerobic.svg or Arginine_Biosynthesis_Aerobic.png).

In this figure, we can see that for aerobic conditions the primary pathways are used to produce L-Arginine. This figure illustrate the primary pathway ('ACGS', 'ACGK', 'AGPR', 'ACOTA', 'ACODA', 'OCBT', 'ARGSS', and 'ARGSL') with 'CBMKr' providing the amino group for L-Arginine.

4.3. Excess L-Arginine Production

When a cell is producing a recombinant protein, it might be required to produce additional L-Arginine for the desired bioproduct. What is the maximum amount of L-Arginine that can be produced for a given growth-rate?

```matlab
model = saved_model;
model = changeRxnBounds(model,'EX_glc_D_e',-10,'l'); % Set maximum glucose uptake
model = changeRxnBounds(model,'EX_o2_e',-30,'l'); % Set oxygen uptake
model = changeRxnBounds(model,'BIOMASS_Ec_iJO1366_WT_53p95M',-10,'b'); % Disable WT biomass reaction
model = changeObjective(model,'BIOMASS_Ec_iJO1366_core_53p95M'); % Set the objective function
FBSolution = optimizeCbModel(model,'max'); % Perform FBA to find optimal growth-rate
model = changeRxnBounds(model,'BIOMASS_Ec_iJO1366_WT_53p95M',FBSolution.f,'b'); % Set fixed growth-rate
model = addDemandReaction(model, 'arg__L_c');
```

```matlab
[tmp,phenylalanine_MAX_rxnID] = ismember({'DM_arg__L_c'},model.rxns);
model = changeObjective(model,'DM_arg__L_c'); % Set the objective function
FBSolution_phenylalanine = optimizeCbModel(model,'max'); % Perform FBA to find optimal growth-rate
xMin = 0.5;
xMax = FBSolution.f;
xInc = (xMax - xMin)/20;
x = xMin;
excessArginine = [];
growthRate = [];
for i = 1:21
    model = changeRxnBounds(model,'BIOMASS_Ec_iJO1366_WT_53p95M',x,'b'); % Set fixed growth-rate
    FBSolution_arginine = optimizeCbModel(model,'max'); % Perform FBA
end
```
excessArginine(i) = FBAsolution_arginine.f;
growthRate(i) = x;
x = x + xInc;
end
plot(growthRate,excessArginine)
title('Excess L-Arginine');
xlabel('Growth-rate (h^{-1})'); ylabel('Excess L-Arginine (mmol.gDW^{-1}.h^{-1})');

![Graph showing excess L-Arginine]

**Figure 10.** A plot showing the maximum amount of L-Arginine that can be produced for a given growth-rate.

This figure illustrates that as the need for excess L-Arginine increases the growth-rate will need to decrease. In this figure we can see that the excess L-Arginine flux can increase from 0.03718 mmol · gDW^{-1} · hr^{-1} when the cell is at maximum growth-rate to 4.353 mmol · gDW^{-1} · hr^{-1} when it is at 50% of that optimal growth-rate. Finally, to increase the L-Arginine flux beyond these levels will require using the pathways that allow for the transport of L-Arginine from the extracellular media.

**5. Aerobic vs Anaerobic Amino Acid Production**

Now let's look at the total amount of flux that is created for each of these amino acids as the oxygen content varies from anaerobic to aerobic.

model = saved_model;
model = changeRxnBounds(model,'EX_glc__D_e','10','1'); % Set maximum glucose uptake
model = changeRxnBounds(model,'BIOMASS_Ec_iJO1366_WT_53p95M','-b','b'); % Disable WT biomass reaction
model = changeObjective(model,'BIOMASS_Ec_iJO1366_core_53p95M'); % Set the objective function
arg_flux = [];
pro_flux = [];
for k = 1:31
    model = changeRxnBounds(model,'EX_o2_e',-(k-1),'b'); % Set oxygen uptake
    FBAsolution = optimizeCbModel(model,'max'); % Perform FBA
    [P, C, vP, vC] = computeFluxSplits(model, {'arg__L_c'}, FBAsolution.x);
    arg_flux(k) = sum(vP);
    [P, C, vP, vC] = computeFluxSplits(model, {'pro__L_c'}, FBAsolution.x);
    pro_flux(k) = sum(vP);
    growthRate(k) = FBAsolution.f;
end

figure(1)
oFlux = 0:30;
ax1 = subplot(3,1,1); % top subplot
ax2 = subplot(3,1,2); % middle subplot
ax3 = subplot(3,1,3); % bottom subplot

plot(ax1,oFlux,arg_flux)
title(ax1,'Total L-Arginine Flux');
xlabel(ax1,'Oxygen uptake flux (mmol.gDW-1.h-1)'); ylabel(ax1,'L-Arg Flux');

plot(ax2,oFlux,pro_flux);
title(ax2,'Total L-Proline Flux');
xlabel(ax2,'Oxygen uptake flux (mmol.gDW-1.h-1)'); ylabel(ax2,'L-Pro Flux');

plot(ax3,oFlux,growthRate);
title(ax3,'Growth-rate');
xlabel(ax3,'Oxygen uptake flux (mmol.gDW-1.h-1)'); ylabel(ax3,'Growth-rate');
5. Conclusion

The purpose of this tutorial was to identify and review the structure and capabilities of the "Arginine and Proline Metabolism Subsystem" subsystem of the *E.coli* iJO1366 model. It began with an overview of the complete subsystem. This was followed by more detailed descriptions of the individual L-Arginine and L-Proline biosynthesis pathways. It concluded with a simulation showing the maximum flux that each of these amino acids can produce in a range from anaerobic to aerobic conditions.

References
