Strain improvement with seven-league boots: Application of metabolomics to improve phenylalanine production by Escherichia coli

Improvement of microbial strains to achieve commercially viable production levels is a prerequisite to any bioprocess. In directed strain improvement, potential bottlenecks, feedback inhibition and side routes are removed by overexpression or knock-out of the gene(s) of interest. Up to date, the selection of these targets is based on expert knowledge, but to a large extent also on ‘educated guess’ and ‘gut feeling’. Therefore time, and thus money, is wasted on targets that later on prove to be irrelevant or only result in a very minor improvement.

We have developed a new approach, allowing the open selection and ranking of the most relevant targets. The technology platform used for this comprises three pillars: a validated metabolome analytical platform, data preprocessing and multivariate data analysis tools. This technology platform allows the replacement of current empirical approaches by a scientific approach towards the selection and ranking of targets. This approach was followed to improve a phenylalanine-producing strain of Escherichia coli that had already been optimized by 8 or more steps of rational design.

The step-wise approach for applying this technology platform is as follows:

1. Experimental design and fermentation: generation of samples under (growth) conditions that result in large differences in productivity or yield
   - Controlled batch fermentations
   - A total of 28 samples were generated (Figure 1).

2. Analysis of the metabolome
   - Analyze all metabolites present in the cells by comprehensive LC- and GC-MS methods (Figure 2).
   - Data preprocessing: prepare clean data files suitable for data analysis.

3. Data analysis: ranking of the most relevant biomolecules
   - Rank the relevant metabolites based on their correlation with the phenylalanine yield.
   - For intermediates: Overexpress intermediate converting enzyme (gene).
   - For side-products: Knock-out side route.

4. Biological interpretation
   - What is the function of the identified metabolites in relation to phenylalanine production (Table 1)?
   - Identify the genes that should be deleted or overexpressed (overexpressed) (Figure 5).
   - i. All positive correlations.
   - For intermediates: Overexpress intermediate converting enzyme (gene).
   - For side-products: Knock-out side route.

   Table 1
<table>
<thead>
<tr>
<th>Rank</th>
<th>Intermediate Phenylalanine biosynthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Chorismate</td>
</tr>
<tr>
<td>3</td>
<td>Glycine</td>
</tr>
<tr>
<td>4</td>
<td>Ethylmalate</td>
</tr>
<tr>
<td>5</td>
<td>3,5-Dihydroxybenzoate</td>
</tr>
<tr>
<td>9</td>
<td>Dipeptide with a glycine?</td>
</tr>
<tr>
<td>12</td>
<td>N-Acetylglutamate</td>
</tr>
<tr>
<td>15</td>
<td>Unknown -15.85</td>
</tr>
<tr>
<td>14</td>
<td>Unknown -17.04</td>
</tr>
<tr>
<td>15</td>
<td>Unknown -14.56</td>
</tr>
</tbody>
</table>

5. Validation of the targets
   - Demonstrate that altering the expression of the genes identified in this way results in an increased phenylalanine production (Figure 6).

Conclusions
- The results of this study prove that the metabolomics/multivariate statistics approach can be applied to identify relevant leads for strain improvement.
- 50% improvement of an already optimized strain.
- Multivariate data analysis tools are powerful tools to extract relevant information from functional genomics data sets.
- The combined metabolomics/multivariate data analysis approach works like a navigator.
- Helps you to find the quickest way and make the largest steps.
- Unbiased identification and ranking of the most relevant targets.
- Opens up the black box of cellular metabolism.