Application of chemometrics for advanced bioprocess monitoring and simulation in view of the FDA’s PAT initiative

Markus Luchner

Department of Biotechnology
University of Natural Resources and Applied Life Sciences, Vienna
Outline

• Introduction to the kinetics of microbial recombinant protein expression
• Process monitoring: an overview
• Case studies: Prediction of complex process variables by chemometric modelling
• Process Analytical Technology (PAT)
• Conclusions
Outline

- Introduction to the kinetics of microbial recombinant protein expression
- Process monitoring: an overview
- Case studies: Prediction of complex process variables by chemometric modelling
- Process Analytical Technology (PAT)
- Conclusions
Principle configuration of a bioprocess

**BIOPROCESS**

**UPSTREAM PROCESSING**

- Substrate tank
- Fermenter

**DOWNSTREAM PROCESSING**

- Storage tank
- Isolation & Purification

Product

Biotechno 2008, Bucharest
Principles of recombinant protein production strategy

Objective: adaptation of recombinant protein production to host cell metabolic capacity

Priority TASK: control → need of specific monitoring
Identification of key variables

Transcription
- Genomic DNA
- Plasmidic DNA

Translation
- mRNA

Metabolic building blocks
- Monomers and energy from host cell metabolism

Process kinetics
- Environmental conditions

DNA microarrays

2-D differential gel electrophoresis (DIGE)

Biotechno 2008, Bucharest
Limits for process optimisation

Inadequate understanding of biological system and observability in real-time

- Complexity
- Lack of on- and in-line sensors
- Unpredictable interaction of recombinant protein with host metabolism
Outline

• Introduction to the kinetics of microbial recombinant protein expression
• Process monitoring: an overview
• Case studies: Prediction of complex process variables by chemometric modelling
• Process Analytical Technology (PAT)
• Conclusions
Information content of the cell

optical properties

(bio-)chemical properties

“recombinant protein factory”

electrical properties

physiological properties

Only few sensors for direct measurement of key process variables available!

Biotechno 2008, Bucharest
State of the art of bioprocess monitoring – availability of off-, at-, on- and in-line measurements

### At-, on- and in-line measurements:
- Classical signals (exhaust gas, base/acid consumption)
- Spectroscopic methods (Optical-, Infrared-, Dielectric spectroscopy, Mass-spectrometry)
- Biosensors
- Electrochemical sensors
- Flow injection analysis

### Off-line analysis:
- Lab-on-a-chip (DNA / RNA / protein quantification)
- Proteomics (DIGE)
- DNA µ-arrays (transcription profiling)
- Surface plasmon resonance (biomolecular interaction)
- Chromatographic methods (GC, HPLC)

Real time data + highly significant off-line data sets

Solution: generate correlations

“on-line” monitoring of complex variables by simulation

Biotechno 2008, Bucharest
Overview of our currently used in- and on-line sensor systems

- Compact proton transfer reaction mass spectrometry
- 2-D multi-wavelength fluorescence spectroscopy
- Base consumption
- Near infrared spectroscopy
- Dielectric spectroscopy
- O₂ und CO₂ off-gas analyzer

Biotechno 2008, Bucharest
Dielectric spectroscopy:

- Intact cells build up charge in electrical field (0.2 - 10 MHz) due to non-conducting nature of the cell
- Plasma membrane act as capacitors
- Resulting capacitance (pF) is proportional to number and cell size

Optical fluorescence spectroscopy:

- Two-dimensional, multi-wavelength fluorescence spectroscopy
- Excitation 270nm – 550nm / emission 310nm - 590nm → resulting in 150 excitation/emission wavelength combinations

Near Infra Red spectroscopy:

- NIR 850 nm

Biotechno 2008, Bucharest
Dielectric spectroscopy:

Physical principle

- Application of a radio-frequency electrical field (0.2 - 10 MHz) to fermentation broth

- Intact cells build up charge due to non-conducting nature of the cell plasma membrane and therefore act as capacitors

- Measurement of resulting capacitance (pF), which is proportional to number and cell size (= measurement of membrane enclosed volume)
  - Pro’s:
    - Good correlation to biomass
  - Con’s:
    - No direct calibration possible due to changes in cell size
    - Additional measurement of conductivity (mS/cm) required

Biomass Monitor ABER Instruments BM214M®
Optical fluorescence spectroscopy:

- Two-dimensional, multi-wavelength fluorescence spectroscopy
- Fluorescent properties of biogenic substances are measured
- Wavelength range: excitation 270nm – 550nm / emission 310nm - 590nm in steps of 20 nm → resulting in 150 excitation/emission wavelength combinations

- Pro’s:
  - Measurement of biogenic fluorophores which are directly involved in metabolic pathways and components
  - Multivariate data set
  - No fouling
  - Rapid measurement (interval for a full scan 90 sec.)

- Con’s:
  - No direct correlation with variables of process operation
  - Interference of sample matrix

DELTA BioView®

Biotechno 2008, Bucharest
Physiological relevant wavelength combinations

- Riboflavin, FAD, FMN 460/520, 380/520
- NAD(P)H 340/460
- Pyridoxine, Pyridoxamine, Pyridoxal-5-P 330/400, 400/500
- Tryptophane 290/350
- Tyrosine 280/310
- Phenylalanine 270/290

(Marose et al., 1998)

Spectra of a bioprocess

Biotechno 2008, Bucharest
Near Infra Red spectroscopy:

- Principle: NIR 850 nm
- range: 0 – 4 AU (Absorbance Units)
- until OD\textsubscript{600} >350

- Pro’s:
  - Good correlation to biomass

- Con’s:
  - No direct calibration possible

TruCell™ www.finesse.com
In- and on-line signals

<table>
<thead>
<tr>
<th>Sensor device</th>
<th>Number of signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>O₂ off gas</td>
<td>1</td>
</tr>
<tr>
<td>CO₂ off gas</td>
<td>1</td>
</tr>
<tr>
<td>Base consumption</td>
<td>1</td>
</tr>
<tr>
<td>Dielectric spectroscopy (capacity, conductivity)</td>
<td>2</td>
</tr>
<tr>
<td>Multi-wavelength fluorescence</td>
<td>150</td>
</tr>
<tr>
<td>NIR</td>
<td>1</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td><strong>156</strong></td>
</tr>
</tbody>
</table>

Large data sets

→ Data mining – screening of relevant variables
Application of chemometric methods for data analysis

- No direct measurement of physiological meaningful variables possible
- Variety of on- and in-line signals available
- Highly developed off-line analytics

Needs:
- Mathematical (chemometric) methods to extract meaningful, yet hidden information and find correlations to off-line variables

Goal:
- Real-time estimation of complex biological variables utilising available on-line sensor signals
Outline

• Introduction to the kinetics of microbial recombinant protein expression
• Process monitoring: an overview
• Case studies: Prediction of complex process variables by chemometric modelling
• Process Analytical Technology (PAT)
• Conclusions
Data flow and tools for pre-processing and modelling of data

**Pre-processing of data**
Filtering and interpolation (Matlab)

**Data mining - screening of relevant on-line data**
Kohonen Self Organizing Maps based on Ward distance Cluster analysis (Viscovery® Profiler (Eudaptics GmbH Vienna))

**Modelling of data - selection of model type**
Partial Least Squares (PLS)
Artificial Neuronal Network (ANN)
Data mining - screening of relevant on-line data

SOM’s (Self organising maps – Kohonen* algorithm)

\[ E = \int \sum h_{ci} | \vec{w}_i - \vec{x}_i |^2 g(\vec{x}) d^n x \]

<table>
<thead>
<tr>
<th>in-line signal</th>
<th>NIR</th>
<th>ex510/em550</th>
<th>ex270/em550</th>
<th>ex550/em590</th>
</tr>
</thead>
<tbody>
<tr>
<td>correlation-coefficient</td>
<td>0.9845</td>
<td>0.9982</td>
<td>0.8894</td>
<td>0.1403</td>
</tr>
</tbody>
</table>

approx. 60 % of fluorescence signals: correlation coeffitient > 0.75

Visualisation by “false colour presentation” and calculation of correlation coefficient

Model types

• Non-linear model:
  
  – Partial least squares (PLS)
    - reduction of multidimensional data sets to lower dimensions for analysis
  
  – Radial Basis Function Neural Network (RBF): Neural networks are better suited for non-linear data
    - supervised learning method
    - non-linear transfer function
    - training by vector weighting

Quality of estimation:
  - Root Mean Square Error of Prediction (RMSEP): RMSEP represents the overall error of the modelled data
Case study: prediction of complex variables

Used data sets

• On-line
  – Exhaust gas composition: $O_2$, $CO_2$
  – Base consumption rate
  – Fluorescence signals
  – Capacity, conductivity

• Off-line:
  – Bacterial Dry Matter (BDM) (gravimetric)
  – Total Cell Number (TCN) / Dead Cell Number (DC) (flow cytometry)
  – Product (mg/g BDM) (electrophoretic)
  – Plasmid Copy Number (PCN) (electrophoretic)
Prediction of key variables in fed-batch cultivation applying RBF-Network

Input: Classical signals (base consumption, exhaust-gas analysis)

Arrows indicate induction of recombinant protein expression

F. Clementschitsch et al. (2005), Journal of Biotechnology, 2, 120, 183-196.

F. Clementschitsch et al. (2006), Microbial Cell Factories, 5, 19-30
Prediction of key variables in fed-batch cultivation applying RBF-Network

Input: Dielectric spectroscopy signals, classical signals (capacity, conductivity, exhaust-gas analysis, base consumption)

Arrows indicate induction of recombinant protein expression

F. Clementschitsch et al. (2005), Journal of Biotechnology, 2, 120, 183-196.
F. Clementschitsch et al. (2006), Microbial Cell Factories, 5, 19-30.
Prediction of key variables in fed-batch cultivation applying RBF-Network

Input: Selected signals (capacity, conductivity, selected fluorescence wavelength combinations, exhaust-gas analysis, base consumption)

Arrows indicate induction of recombinant protein expression

F. Clementschitsch et al. (2005), Journal of Biotechnology, 2, 120, 183-196.
F. Clementschitsch et al. (2006), Microbial Cell Factories, 5, 19-30

Biotechno 2008, Bucharest
Prediction of key variables in fed-batch cultivation applying PLS

Input: Selected signals (capacity, conductivity, selected fluorescence wavelength combinations, exhaust-gas analysis, base consumption)

Arrows indicate induction of recombinant protein expression

F. Clementschitsch et al. (2005), Journal of Biotechnology, 2, 120, 183-196.
F. Clementschitsch et al. (2006), Microbial Cell Factories, 5, 19-30
Example: application of NIR for monitoring

\[ f(x) = p1 \cdot x^4 + p2 \cdot x^3 + p3 \cdot x^2 + p4 \cdot x + p5 \]

\[ p1 = -81.13, \ p2 = 650.3, \ p3 = -1825, \ p4 = 2201, \ p5 = -968.9 \]

max. deviation $< 2g$ (6 %)
• Achievements:
  – On-line prediction of key variables
  – Set up of control loops enabled

• Limitations
  – Monitoring of deviations on molecular level (e.g. stress response)
  – Validability of prediction by chemometric methods not fully accepted by regulatory authorities

Extension of on-line data base
Proton Transfer Reaction Mass Spectrometry:

- Reaction: VOCs charged by
  \[ \text{H}_3\text{O}^+ + R \rightarrow \text{RH}^+ + \text{H}_2\text{O} \]
- Detection limit: 500 pptv
- Mass range: 1 – 300 am

- Pro’s:
  - Non invasive measurement
  - Measurement of metabolites
  - Rapid measurement (approximately 3 minutes per cycle)
  - Soft ionization – no fragmentation

- Con’s:
  - Mass information but no structure information

www.ptrms.com
www.ionimed.com

Biotechno 2008, Bucharest
## In- and on-line signals

<table>
<thead>
<tr>
<th>Sensor device</th>
<th>Number of signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>O$_2$ off gas</td>
<td>1</td>
</tr>
<tr>
<td>CO$_2$ off gas</td>
<td>1</td>
</tr>
<tr>
<td>Base consumption</td>
<td>1</td>
</tr>
<tr>
<td>Dielectric spectroscopy (capacity, conductivity)</td>
<td>2</td>
</tr>
<tr>
<td>Multi-wavelength fluorescence</td>
<td>150</td>
</tr>
<tr>
<td>NIR</td>
<td>1</td>
</tr>
<tr>
<td>PTR-MS</td>
<td>Up to 60</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td><strong>Up to 216</strong></td>
</tr>
</tbody>
</table>
Application of PTR-MS for process monitoring

PTR-MS enables the transition from pattern recognition to quantitative analysis of volatile metabolites

Biotechno 2008, Bucharest
• Non invasive sampling device

• Sensor for early detecting of different physiological states
  – e.g. growth and non growth associated recombinant protein production and overburden of the cell

• Real time availability of complex variables for process control
Outline

• Introduction to the kinetics of microbial recombinant protein expression
• Process monitoring: an overview
• Case studies: Prediction of complex process variables by chemometric modelling
• Process Analytical Technology (PAT)
• Conclusions
Process Analytical Technology (PAT) and Quality by Design (QbD)

- **Process Analytical Technology initiative:**
  - a system for designing, analyzing, and controlling manufacturing through timely measurements (i.e. during processing) of critical quality and performance attributes of raw- and in-process materials and processes with the goal of ensuring final product quality. (http://www.fda.gov/Cder/OPS/pat.htm)

- **Required tools for the implementation of PAT:**
  - Multivariate data acquisition and data analysis tools
  - Modern process analyzers or process analytical chemistry tools
  - Extension of process monitoring and control tools

**GOAL:** definition of the design space to gain more flexibility in operation
Outline

• Introduction to the kinetics of microbial recombinant protein expression
• Process monitoring: an overview
• Case studies: Prediction of complex process variables by chemometric modelling
• Process Analytical Technology (PAT)
• Conclusions
Conclusions I

• Chemometric modelling and prediction contribute to the improvement of process monitoring and control

• Contribution of individual sensor signals:
  – Classical signals do not contain enough information to allow the estimation of complex process variables
  – Monitoring of key variables achieved through signal combination
  – Selection of input signal improves quality of prediction

• PTR-MS technology enables
  – early detection of deviations and different physiological states
  – real-time quantification of specific process relevant compounds
Conclusions II

Complex diagnostics platform comprising in-, on- and off-line data delivers a broad spectrum of information

→ basis for PAT and QbD compliance
→ enables the definition of the designs pace (ICH Q8)
Acknowledgements

Department of Biotechnology

Group of Microbial Fermentation

Funding by the Austrian Center of Biopharmaceutical Technology (www.acbt.at)