

Automatisierung und Optimierung von
Herstellungsprozessen rekombinanter pharmazeutischer Proteine

Von der Naturwissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

M.Sc. Dipl.-Ing. (FH) Andree Ellert

geboren am 19.12.1975 in Celle

Referent: Prof. Dr. Thomas Scheper

Korreferentin: Prof. Dr. Ursula Rinas

Tag der Promotion: 7. April 2015

Veröffentlichungen zur Dissertation

Teile dieser Arbeit wurden in folgenden Beiträgen vorab veröffentlicht:

Schmelig, C.; Wilhelm, B.-U.; Ellert, A.; Peuker, T.; Luttmann, R. (2005): Online Monitoring and Automation of an Integrated Process with *Pichia pastoris*: A contribution to the PAT-Initiative, Beilage in BioProcess International, Trends in Integrated Biomanufacturing 3(9): 22–25.

Ellert, A.; Hahn, B.; Eicke, M.; Treder, S.; Eckl, G.; Luttmann, R. (2006): Monitoring and Control of Integrated Bioprocesses for Production of Active Pharmaceutical Ingredients with *P. pastoris*, Vortrag auf der ACHEMA 2006 M3C in DSP: Measurement, Monitoring, Modelling and Control in Downstream and Integrated Processes, Frankfurt am Main.

Ellert, A.; Tatge, F.; Borchert, S.; Bertelsen, H.P.; Scheffler, U.; Grebe, A.; Luttmann, R. (2008): Setup of a Multi-Fermenter System BIOSTAT[®] Qplus-6 for the Execution of Screening Experiments, Poster auf der European BioPerspectives 2008, Hannover.

Ellert, A.; Tatge, F.; Thiesing, D.; Bertelsen, H.P.; Scheffler, U.; Luttmann, R. (2009): Optimization of Recombinant Protein Production by DoE – Design of Experiments in a Multi-Fermenter System, Poster auf der ACHEMA 2009, Frankfurt am Main.

Ellert, A.; Tatge, F.; Thiesing, D.; Bertelsen, H.P.; Scheffler, U.; Luttmann, R. (2009): Einsatz der Statistischen Versuchsplanung (SVP) zur Steigerung der Raum-Zeit-Ausbeute von nativen Proteinen in *E. coli*, Poster auf der 27. Jahrestagung der Biotechnologen, Mannheim.

Ellert, A.; Thiesing, D.; Tatge, F.; Borchert, S.; Ganzlin, M.; Luttmann, R. (2010): Optimisation of Fluorescent Fusion Protein Expression in *Escherichia coli* via Design of Experiments, Vortrag auf der CAB 2010 – 11th International Symposium on Computer Applications in Biotechnology, Leuven, Belgien.

Vorwort

Die vorliegende Arbeit wurde in der Zeit von Oktober 2004 bis Oktober 2009 im Forschungs- und Transferzentrum Bioprozess- und Analysetechnik der Hochschule für Angewandte Wissenschaften (HAW) Hamburg unter der Leitung von Prof. Dr.-Ing. Reiner Luttmann angefertigt.

Ihm danke ich für sein persönliches Engagement und damit verbunden die Möglichkeit in verschiedenen Projekten mitzuarbeiten, sowie die Bereitstellung eines enormen Fuhrparks von Geräten und Materialien.

Prof. Dr. Gesine Cornelissen vom Forschungs- und Transferzentrum Bioprozess- und Analysetechnik der Hochschule für Angewandte Wissenschaften danke ich für die Übernahme der finalen Betreuung meiner Dissertation.

Herrn Prof. Dr. Thomas Scheper und Frau Prof. Dr. Ursula Rinas vom Institut für Technische Chemie der Leibniz Universität Hannover danke ich für die Übernahme der Referate.

Gunnar Eckl, Frithjof Tatge, David Thiesing und Sven-Oliver Borchert danke ich für den Einsatz im Rahmen ihrer Master-, Diplom- oder Studienarbeiten und für die eingebrachten Ergebnisse.

Des Weiteren danke ich Conny Vikström von der Fa. UMETRICS AB, Schweden, für die Unterstützung über das Projektende hinaus.

Für das tolle Arbeitsklima möchte ich dem gesamten Team des Forschungs- und Transferzentrums und insbesondere meinen Kollegen Christian, David, Fritze, Gunnar, Jenne, Hape, Matze, Sven und Uli danken.

Abschließend möchte ich mich bei meiner Familie bedanken. Ohne die Unterstützung, den Rückhalt und Zuspruch meiner Eltern Otto und Bettina sowie meiner Schwester Juliane und Oma Ruth wäre diese Arbeit nicht entstanden.

*„Viel zu spät begreifen viele die versäumten Lebensziele:
Freuden, Schönheit der Natur, Gesundheit, Reisen und Kultur.
Darum, Mensch, sei zeitig weise! Höchste Zeit ist's! Reise, reise!“*

(Wilhelm Busch)

Kurzfassung

Andree Ellert

Automatisierung und Optimierung von Herstellungsprozessen rekombinanter pharmazeutischer Proteine

Im ersten Teil dieser Arbeit wurde basierend auf vorhandener sowie zusätzlich in das verwendete Bioreaktorsystem integrierter MSR-Technik eine Prozessführungsstrategie erarbeitet, mit der das Chemokin 1-3del I-TAC mit der methylotrophen Hefe *Pichia pastoris* im klassischen dreiphasigen Herstellungsablauf bestehend aus Batch, Fed Batch und Produktion vollautomatisch hergestellt werden konnte.

Da das herzustellende Protein durch *P. pastoris* in das Kulturmedium sekretiert wird, war es möglich, den Upstream-Prozess mit einer Querstrom-Mikrofiltration als primären Downstream-Schritt zu verknüpfen. Der bestehende Ansatz einer Sequentiell Integrierten Bioprozessführung wurde weiter ausgearbeitet und in eine vollautomatische Prozessführungsstrategie übertragen.

Für die Steuerung der sich zyklisch wiederholenden Hauptphasen Produktion, Mikrofiltration, Zellernte und Mediaauffrischung wurden die aus dem Trübungssignal berechnete Zelldichte, die Masse des Reaktionsvolumens und die Masse des produktbehafteten Mediums im Permeattank verwendet. Über S88-Rezepte konnten zuverlässige und reproduzierbare experimentelle Bedingungen geschaffen werden.

Im zweiten Teil dieser Arbeit wurden optimale Betriebsparameter für die Herstellung der humanen Rezeptor-Tyrosinkinase hEphB2 mit zwei Mutanten des Expressionswirts *Escherichia coli* ermittelt. Dafür wurde zunächst ein Multi-Bioreaktorsystem aufgebaut und in Betrieb genommen sowie um verschiedene Analysen- und Automatisierungskomponenten erweitert. Der dreiphasige Produktionsablauf wurde über S88-Rezepte und ein neu entwickeltes MATLAB[®]-Tool automatisiert.

Auf Basis der Statistischen Versuchsplanung wurden Screening-, Optimierungs- und Robustheitsexperimente durchgeführt und ausgewertet. Vorher wurde der reguläre Versuchsbereich durch Spezifikation irregulärer Faktorbereiche experimentell festgelegt sowie eine Zielproteinanalytik etabliert. Zur Bewertung der Versuche wurde die Raum-Zeit-Ausbeute (RZA) als verfahrenstechnisch sinnvolles Gütekriterium definiert.

Im Screening wurde die molare IPTG-Konzentration als statistisch nicht signifikanter Faktor identifiziert. In der Optimierung wurde nur noch eine *E. coli* Mutante aufgrund der höheren RZA an zellinternem löslichem Protein verwendet. Für die zellspezifische Wachstumsrate im Fed Batch und die Flüssigphasentemperatur ab Induktion konnten optimale Einstellungen gefunden und abschließend durch einen Robustheitstest verifiziert werden.

Schlüsselwörter: *P. pastoris*, *E. coli*, I-TAC, hEphB2, ZsGreen, Integrierte Prozessführung, Automatisierung, S88-Rezepte, Statistische Versuchsplanung, Multi-Bioreaktorsystem

Abstract

Andree Ellert

Automation and optimisation of production processes of recombinant pharmaceutical proteins

The first part of this paper describes a process control strategy that was developed based on available and new instrumentation and control technologies that were additionally integrated into the bioreactor system used. Chemokine 1-3del I-TAC was produced using the methylotrophic yeast *Pichia pastoris* in a fully automated, classic manufacturing process consisting of three phases: batch cultivation, fed-batch cultivation and production.

As *P. pastoris* secretes target protein into the culture medium, it was therefore possible to link the upstream process to the primary downstream step entailing crossflow microfiltration. The existing approach of sequentially integrated bioprocessing was further refined and adopted as a fully automatic process control strategy.

The main phases of production, microfiltration, cell harvesting and medium refreshing were repeated cyclically. The transition between the different phases was controlled by the cell density signal calculated from the turbidity measurement; the mass of the liquid reaction volume; and by the mass of the medium containing the product inside the permeate tank. Reliable and robust experimental conditions were achieved with S88 recipes.

The second part of this paper discusses the optimal operating conditions determined for the production of the human receptor tyrosine kinase hEphB2 using two mutants of the expression host *Escherichia coli*. Firstly, a multi-bioreactor system was set up and put into operation, and its functionalities were extended using diverse analytical and automation components. Secondly, the three-phase manufacturing process was fully automated with S88 recipes and a newly developed MATLAB[®] tool.

Based on a Design of Experiments strategy, screening, optimisation and robustness experiments were performed and evaluated. The regular constraints were defined by conducting preliminary experiments based on the specification of irregular factor settings. An analytical method was then established for the target protein. The effect of the factors studied on the space-time yield (STY) was regarded as the most evident quality response for process evaluation.

During screening, the molar IPTG concentration was revealed to be a statistically insignificant factor. Optimisation experiments were performed using only one *E. coli* strain, which showed a higher STY for soluble protein inside the cells. Optimal parameter settings were found, and subsequently verified by a robustness test, for the cell-specific growth rate during fed-batch cultivation and for the liquid-phase temperature upon induction.

Keywords: *P. pastoris*, *E. coli*, I-TAC, hEphB2, ZsGreen, integrated bioprocessing, automation, S88 recipes, Design of Experiments, multi-bioreactor system

Nomenklatur – Messtechnik und Bioreaktionsmodell

A_F	:= Filterfläche	$[m^2]$
A_{PIM}	:= UV-Absorption des Zielproteins	$[AU]$
c_{IK}	:= Massenkonzentration der Komponente I im Teilsystem K	$[g l^{-1}]$
C_{JtotK}	:= molare Gesamtkonzentration der Komponente J im Teilsystem K	$[mol l^{-1}]$
F_K	:= Volumenstrom in das oder aus dem Teilsystem K	$[l h^{-1}]$
F_{Krel}	:= relative Zufütterrate aus dem Teilsystem K	$[\%]$
F_{nI}	:= Begasungsrate der Komponente I unter Normbedingungen	$[l h^{-1}]$
IER	:= volumetrische Produktionsrate einer Komponente I aller Zellen	$[g l^{-1} h^{-1}]$
ITR	:= volumetrische Transferrate einer Komponente I (von G in L)	$[g l^{-1} h^{-1}]$
IUR	:= volumetrische Aufnahme rate einer Komponente I aller Zellen	$[g l^{-1} h^{-1}]$
J	:= flächenbezogener Flux	$[l h^{-1} m^{-2}]$
K_1	:= Parameter zur Anpassung der Trübungsmessung	$[g l^{-1}]$
K_2	:= Parameter zur Anpassung der Trübungsmessung	$[AU^{-1}]$
k_{La}	:= volumetrischer O_2 -Transferkoeffizient	$[h^{-1}]$
$K_{P1/A}$:= Verhältnis Offline- zu Atline-Proteinmessung	$[g l^{-1} AU^{-1}]$
K_{S2}	:= Parameter zur Anpassung der Methanolmessung	$[g l^{-1} ppm^{-1}]$
M_I	:= Molmasse der Komponente I	$[g mol^{-1}]$
m_{IK}	:= Masse der Komponente I im Teilsystem K	$[g]$
N_{St}	:= Rührerdrehzahl	$[min^{-1}]$
pH	:= pH-Wert	$[-]$
p_K	:= Gesamtdruck im Teilsystem K	$[bar]$
p_{O_2}	:= relativer Sauerstoffpartialdruck in der Flüssigphase	$[\%]$
PRD	:= Produktivität	$[g h^{-1}]$
p_{TM}	:= Transmembrandruck	$[bar]$
Q_I	:= volumetrische Ver- oder Entsorgungsrate der Komponente I	$[g l^{-1} h^{-1}]$
$q_{I/X}$:= zellspezifische Reaktionsrate der Komponente I	$[g g^{-1} h^{-1}]$
R	:= allgemeine Gaskonstante	$[J mol^{-1} K^{-1}]$
RQ	:= molarer Respirationsquotient	$[mol mol^{-1}]$
RZA	:= volumetrische Raum-Zeit-Ausbeute	$[RFU h^{-1}]$
$S_{ex/em}$:= Fluoreszenzmessung einer Wellenlängenkombination	$[RFU]$
S_{turb}	:= optisches Signal der Trübungsmessung in der Flüssigphase	$[AU]$
t	:= Zeit	$[h]$
T_K	:= Temperatur des Teilsystems K	$[K]$
V_K	:= Volumen des Teilsystems K	$[l]$

V_{nM}	:= Molvolumen unter Gasnormbedingungen	[mol^{-1}]
x_{iG}	:= Stoffmengenanteil der Komponente I in der Gasphase	[-]
y_K	:= Stellgröße eines Reglers im Teilsystem K	[%]
$y_{X/I}$:= Ausbeutekoeffizient Zellmasse/Komponente I	[gg^{-1}]

Griechische Formelzeichen

$\alpha_{Z/X}$:= Biofeucht-/Biotrockenmasseverhältnis	[-]
ϑ_L	:= Temperatur in der Flüssigphase	[°C]
λ	:= Wellenlänge	[nm]
μ	:= zellspezifische Wachstumsrate	[h^{-1}]
ρ_K	:= Dichte im Teilsystem K	[gl^{-1}]

Indizes

0	:= Anfangsbedingung (Zeitpunkt null)
A	:= Analysensystem
Ac	:= Säure (<i>acid</i>)
AF	:= Antischaum (<i>antifoam</i>)
AIR	:= Luft
Al	:= Lauge (<i>alkali</i>)
at	:= atline
B	:= Puffer (<i>buffer</i>)
blk	:= Leerwert einer Messung (<i>blank</i>)
C, CO ₂	:= Kohlenstoffdioxid
CDW	:= Biotrockenmasse (<i>cell dry weight</i>)
crit	:= kritischer Wert
DJ	:= Doppelmantel (<i>double jacket</i>)
E	:= Eppendorf Tube
em	:= Emission
est	:= abgeschätzt (<i>estimated</i>)
ex	:= Exzitation
feed	:= Cross flow-Eingang
FID	:= Flammenionisationsdetektor
FMC	:= Fermentation Micro Computer
G	:= Gasphase
gr	:= Wachstumsanteil (<i>growth</i>)

H	:= Ernte (<i>harvest</i>)
high	:= oberer Grenzwert, Höchstwert
i, j, k	:= Laufvariable für ein Ereignis o. ä.
IB	:= Inclusion Body
in	:= Eingang (<i>input</i>)
ind	:= Induktion einer Proteinexpression
L	:= Reaktionsraum Flüssigphase (<i>liquid</i>)
LC	:= Flüssigchromatographie
low	:= unterer Grenzwert, Mindestwert
lys	:= lysiert
m	:= Maintenanceanteil
M	:= Medienphase
max	:= maximaler Grenzwert, Maximalwert
min	:= minimaler Grenzwert, Minimalwert
n	:= Gasnormbedingungen
O, O ₂	:= Sauerstoff
OD	:= Optische Dichte
off	:= offline
on	:= online
opt	:= optimal
out	:= Ausgang (<i>output</i>)
P1	:= Produkt 1 (Zielprotein, <i>P. pastoris</i> : I-TAC, <i>E. coli</i> : hEphB2)
P3	:= Produkt 3 (<i>P. pastoris</i> : Pyruvat, <i>E. coli</i> : Acetat)
per	:= Permeat
raw	:= Rohwert einer Messung
rel	:= relative (auf Maximalwert bezogene) Zufütterung
ret	:= Retentat
R1	:= Reservoir 1 (<i>P. pastoris</i> : Glycerol, <i>E. coli</i> : Glukose)
R2	:= Reservoir 2 (<i>P. pastoris</i> : Methanol, <i>E. coli</i> : IPTG)
R3	:= Reservoir 3 (<i>P. pastoris</i> : Refreshmedium)
S1	:= Substrat 1 (<i>P. pastoris</i> : Glycerol, <i>E. coli</i> : Glukose)
S2	:= Substrat 2 (<i>P. pastoris</i> : Methanol)
S3	:= Substrat 3 (<i>P. pastoris</i> : Pyruvat, <i>E. coli</i> : Acetat)
S	:= Probenahme (<i>sampling</i>)
sol	:= löslich (<i>soluble</i>)

T1	:= Titration 1 (Säure)
T2	:= Titration 2 (Lauge)
T	:= Transfer
tot	:= Gesamtmenge dissoziierfähiger Größen
turb	:= Trübung (<i>turbidity</i>)
w	:= Sollwert
X	:= Biotrockenmasse
Z	:= Biofeuchtmasse

Nomenklatur – Statistische Versuchsplanung

a_i	:= tabellierter Koeffizient für den W-Test
c_{ii}	:= Element der Hauptdiagonalen einer Matrix
c	:= Anzahl der Mehrfachrealisierungen
D	:= Gesamterwünschtheit (<i>overall desirability</i>)
d	:= Wunschfunktion
DF	:= Freiheitsgrad (<i>degree of freedom</i>)
\underline{e}	:= Vektor der geschätzten Residuen
F	:= Quantil der F-Verteilung
f	:= Anzahl der Versuche mit einfacher Zählung von c Mehrfachrealisierungen
G	:= Testgröße für den Grubbs-Test
g	:= Normierungsfunktion
\underline{H}	:= Dach- oder Hut-Matrix (<i>hat matrix</i>)
H_0	:= Nullhypothese
H_A	:= Alternativhypothese
h_{ii}	:= Element der Hauptdiagonalen der Dach- oder Hut-Matrix
Log(D)	:= Gütemaß für die Entfernung zum Gesamtziel
MS	:= mittlere quadratische Abweichung (<i>mean sum of squares</i>)
MV	:= Gütemaß für die Modellvalidität
n	:= Gesamtanzahl der Versuche
p	:= Überschreitungswahrscheinlichkeit, Signifikanzwert
PRESS	:= Quadratsumme der prognostizierten Residuen (<i>prediction error sum of squares</i>)
Q^2	:= Vorhersagegüte, Prognosegüte (<i>goodness of prediction</i>)
r	:= Anzahl der zu schätzenden Regressionskoeffizienten
R^2	:= Bestimmtheitsmaß, Korrelationskoeffizient (<i>goodness of fit</i>)
R_{adj}^2	:= korrigiertes Bestimmtheitsmaß

RMSE	:= Wurzel des mittleren quadratischen Fehlers (<i>root mean square error</i>)
RP	:= Gütemaß für die Reproduzierbarkeit, Wiederholbarkeit (<i>reproducibility</i>)
SD	:= Standardabweichung (<i>standard deviation</i>)
SE	:= Standardfehler (<i>standard error</i>)
SS	:= Summe der Fehlerquadrate (<i>sum of squares</i>)
t	:= Quantil der t-Verteilung
W	:= Testgröße für den W-Test
w	:= transformierte Einflussgröße der kanonischen Analyse
\underline{X}	:= Designmatrix mit den normierten Stufen einer Einflussgröße
X	:= Messwert einer Einflussgröße
x	:= Einflussgröße, Faktor, Regressor, unabhängige Variable
\underline{x}	:= Vektor der Einflussgröße
Y	:= Messwert einer Zielgröße
y	:= Zielgröße, Antwort, Regressand, abhängige Variable
\underline{y}	:= Vektor der Zielgröße
\hat{y}	:= Schätzwert für die Zielgröße
$\hat{\underline{y}}$:= Schätzvektor der Zielgröße
\bar{y}	:= Mittelwert über alle Elemente im Vektor der Zielgröße

Griechische Formelzeichen

α	:= Irrtumswahrscheinlichkeit, Signifikanzniveau
β	:= Regressionskoeffizient
$\underline{\beta}$:= Vektor der Regressionskoeffizienten
$\hat{\beta}$:= Schätzwert eines Regressionskoeffizienten
$\hat{\underline{\beta}}$:= Schätzvektor der Regressionskoeffizienten
Δ	:= Schrittweite
ε	:= Residuum, Modellfehler
$\underline{\varepsilon}$:= Vektor der Residuen
γ	:= Sternpunktabstand
κ	:= Konditionszahl
λ	:= Eigenwert
μ	:= Mittelwert
v	:= Wichtungparameter

Σ	:=	Summe
σ	:=	Standardabweichung
σ^2	:=	Varianz

Indizes

cal	:=	berechnet (<i>calculated</i>)
can	:=	kanonisch (<i>canonical</i>)
COE	:=	Regressionskoeffizient (<i>coefficient</i>)
cp	:=	Zentralpunkt (<i>center point</i>)
high	:=	oberer Grenzwert, Höchstwert
i, j, k, m	:=	Laufvariable für ein Ereignis o. ä.
LOF	:=	Anpassungsfehler, Modellfehler (<i>lack of fit</i>)
low	:=	unterer Grenzwert, Mindestwert
max	:=	maximaler Grenzwert, Maximalwert
MEA	:=	Mittelwert (<i>mean</i>)
min	:=	minimaler Grenzwert, Minimalwert
PE	:=	Wiederholungsfehler (<i>pure error</i>)
REG	:=	Regression
RES	:=	Residuum
SP	:=	stationärer Punkt
spx	:=	Simplex
sta	:=	standardisiert
stu	:=	studentisiert
T	:=	transponierter Vektor bzw. Matrix
tab	:=	tabelliert
tar	:=	Ziel (<i>target</i>)
TC	:=	Gesamtabweichung vom Mittelwert (<i>total corrected for the mean</i>)
wgt	:=	Wichtung (<i>weighting</i>)

Inhaltsverzeichnis

1 Einleitung	1
1.1 Aspekte der pharmazeutischen Wirkstoffherstellung	1
1.2 Zielsetzung der Arbeit	2
2 Eingesetzte mikrobielle Expressionssysteme und hergestellte Produkte	3
2.1 Extrazelluläre Produktion von chemotaktischen Zytokinen mit <i>P. pastoris</i>	3
2.1.1 Die methylo tro phe Hefe <i>Pichia pastoris</i> SMD1168	3
2.1.2 Eigenschaften des Zielprodukts 1-3del I-TAC	4
2.1.3 Herstellung und Transformation des Expressionsvektors pPIC9K	7
2.2 Intrazelluläre Produktion von Rezeptor-Tyrosinkinasefragmenten mit <i>E. coli</i>	9
2.2.1 Die humane Rezeptor-Tyrosinkinase hEphB2	9
2.2.2 Das Reef Coral Fluorescent Protein ZsGreen als Expressionsreporter	11
2.2.3 Der Stamm <i>Escherichia coli</i> BL21 (DE3)	13
2.2.4 Der Expressionsvektor pZT7 und seine funktionellen Bestandteile	14
2.2.5 Expression des Fusionsproteins in Form von Inclusion Bodies	16
3 Monitoring der Prozesse	18
3.1 Einbindung einer Bioprozessanalytik	18
3.2 Prozessüberwachung mit Inline-, Online- und Atline-Messtechnik	19
3.2.1 Bestimmung der Zelldichte durch Trübungsmessung	19
3.2.2 Abgasmessung für O ₂ und CO ₂	21
3.2.3 Verfahren zur Methanolbestimmung	22
3.2.4 HPLC-Analytik der Proteinkonzentration in der Flüssigphase	29
3.2.5 Produktmessung mittels 2D-Fluoreszenzspektroskopie	32
3.3 Erweitertes Prozessmonitoring	34
3.3.1 Einsatz von Software-Sensorik an biotechnologischen Prozessen	34
3.3.2 Modellbasiertes Monitoring mit MATLAB [®] Applikationen	34
4 Herstellung von pharmazeutischen Proteinen im dreiphasigen Produktionsablauf	39
4.1 Der klassische dreiphasige Produktionsablauf	39
4.2 Der Aufbau der Bioreaktoranlage BIOSTAT [®] C	40
4.3 Mess-, Steuer- und Regelungsaufgaben für eine Kultivierung mit <i>P. pastoris</i>	42
4.4 Erforderliche Infrastruktur zur automatischen Prozessführung	44
4.4.1 Vor-Ort-Automatisierung mit dem Prozesscontroller DCU	44
4.4.2 Datenerfassung mit dem SCADA-System MFCS/win	45
4.4.3 Prozessautomatisierung mit der Visual Basic [®] Applikation MFCS-Tools	47
4.4.4 Prozessführung mit Hilfe von Rezepten nach ANSI/ISA-88.01	49
4.5 Vollautomatischer Prozessablauf	54
4.5.1 S88-gesteuerte Inokulation	54

4.5.2	Automatische Batchende-Erkennung.....	55
4.5.3	Steuerung der zellspezifischen Wachstumsrate	56
4.5.4	Induktion der Produktbildung	57
4.5.5	Sequentielle pO ₂ -Regelung	58
5	Integrierte Prozessführung – die Verbindung von Up- und Downstream	59
5.1	Das Konzept der Integrierten Prozessführung	59
5.1.1	Gewählter Prozessaufbau	59
5.1.2	Die Sequentiell Integrierte Prozessführung.....	62
5.1.3	MSR-Aufgaben während der Integrierten Prozessführung.....	64
5.1.4	<i>In-situ</i> product removal (ISPR) mit der Crossflow-Filtrationseinheit	65
5.2	Automatisierung der Sequentiell Integrierten Prozessführung	66
5.2.1	Umsetzung mit S88-Steuerrezepten und DCU-Sequenzen.....	66
5.2.2	Vorbereitung der Filtration durch Aufbau eines Filterkuchens	67
5.2.3	Durchführung der Mikrofiltration	69
5.2.4	Vorbereitung neuer Sequenzen durch Zellernte und Medienrefresh	72
5.2.5	Fouling an der Crossflow-Membran	73
6	Grundlagen der Prozessoptimierung mit Statistischer Versuchsplanung	76
6.1	Der dreiphasige Produktionsablauf und dessen Optimierungspotential	76
6.1.1	Klassischer Prozessablauf einer HCDC mit <i>E. coli</i>	76
6.1.2	Der modulare Aufbau des Multi-Bioreaktorsystems BIOSTAT® Qplus.....	77
6.1.3	Anbindung der Analysen- und Automatisierungskomponenten an MFCS/win	79
6.1.4	Einfluss- und Zielgrößen für eine optimierte Proteinproduktion.....	80
6.2	Prozessoptimierung mit Statistischer Versuchsplanung	83
6.2.1	Genereller Ablauf und Modelle der Statistischen Versuchsplanung	83
6.2.2	Die Versuchsplanungs- und Auswertungssoftware MODDE.....	86
6.3	Festlegung des Versuchsbereichs durch Spezifikation irregulärer Faktorbereiche	87
6.3.1	Zusammenhang zwischen Kultivierungstemperatur und Wachstumsrate	87
6.3.2	Zusammenhang zwischen Acetatbildung und Wachstumsrate	91
6.3.3	Einfluss der IPTG-Konzentration auf die Proteinexpression.....	93
6.3.4	Grenzen des regulären Versuchsbereichs.....	94
6.3.5	Koordinatentransformation der Faktorstufen.....	95
6.4	Ermittlung der Raum-Zeit-Ausbeute als relevante Zielgröße	96
6.4.1	Fluoreszenzmessung mit dem GENios™ Reader	96
6.4.2	Berechnung der Raum-Zeit-Ausbeute.....	96
7	Ermittlung optimaler Betriebsparameter mit Statistischer Versuchsplanung	100
7.1	Die Durchführung der Screeningversuche	101
7.1.1	Festlegung der Faktorstufen für das vollfaktorielle Screening-Design.....	101
7.1.2	Aufstellung der Designmatrix und Bestimmung der Konditionszahl	103

7.1.3	Aufstellung der Planmatrix und Auswertung der Antwortmatrix.....	104
7.1.4	Prüfung der Antwortgrößen auf Normalverteilung mit dem W-Test.....	105
7.1.5	Prüfung der Antwortgrößen auf Ausreißer mit dem Grubbs-Test	106
7.1.6	Statistische Modellbildung über Multiple Lineare Regression (MLR).....	108
7.1.7	Prüfung der Residuen auf Normalverteilung und Ausreißer.....	110
7.1.8	Identifikation statistisch signifikanter Faktoren.....	111
7.1.9	Mehrfaktorielle Varianzanalyse und Signifikanztest der Regressionsmodelle...	116
7.1.10	Gütemaße der Multiplen Linearen Regression.....	121
7.1.11	Auswahl eines <i>E. coli</i> Stamms für Optimierungsversuche	125
7.2	Die Durchführung der Optimierungsversuche	127
7.2.1	Festlegung der Faktorstufen und Auswahl des Optimierungs-Designs	127
7.2.2	Aufstellung der Planmatrix und Auswertung der Antwortmatrix	128
7.2.3	Ergebnisse der Regressionsanalyse.....	130
7.2.4	ANOVA für die Optimierungsversuche.....	133
7.2.5	Nichtlineare Optimierung zur Identifikation von optimalen Betriebsbedingungen	135
7.2.6	Wunschfunktionen und das Simplexverfahren zur Optimumsuche.....	138
7.3	Verifikation der optimalen Betriebsparameter durch den Robustheitstest.....	142
7.3.1	Festlegung der Faktorabweichungen und Erstellung des Designs	142
7.3.2	Regressionsanalyse und ANOVA	144
7.3.3	Beurteilung der Variabilität der Antworten	146
8	Zusammenfassung	147
9	Anhang	149
9.1	Eingesetzte Kultivierungsmedien.....	149
9.1.1	Verwendete Medien für die Kultivierung von <i>P. pastoris</i>	149
9.1.2	Verwendete Medien für die Kultivierung von <i>E. coli</i>	150
9.2	Bestimmung der optischen Dichte, der Biotrockenmasse und der Biofeuchtmasse ...	151
9.3	Acetatmessung über enzymatisches Testkit.....	152
9.4	Vorgehensweise für die Zielproteinanalytik bei <i>E. coli</i> Kultivierungen.....	152
9.5	Automatisierungsaufgaben.....	159
10	Literaturverzeichnis	161

1 Einleitung

1.1 Aspekte der pharmazeutischen Wirkstoffherstellung

In den vergangenen Jahren hat die Biotechnologie verstärkt Einzug in die pharmazeutische Wirkstoffentwicklung gehalten. Biotechnologisch hergestellte Protein-Pharmazeutika – sog. Biologicals – sind aufgrund ihrer Wirksamkeit und Spezifität eines der zentralen Innovationsfelder in der Pharmakotherapie. Bereits heute werden in Deutschland mehr als 16 % des Gesamtpharmamarktes mit Biotech-Arzneien erzielt, im Jahr 2009 entsprach dies einer erwirtschafteten Summe von knapp 4.7 Mrd. € (Michl and Heinemann, 2010).

2009 wurden in Deutschland insgesamt 44 neue Wirkstoffe (inkl. neuer Kombinationen bekannter Wirkstoffe) zugelassen, darunter 12 Biopharmazeutika. Damit waren Ende des Jahres 188 Biologicals auf dem deutschen Markt vertreten, 468 befanden sich in der klinischen Entwicklung. Die zugelassenen Präparate teilen sich in die drei Bereiche monoklonale Antikörper, andere rekombinante Proteine wie Insuline, Epoetine, Hormone, Enzyme oder Interferone sowie Impfstoffe auf (Michl and Heinemann, 2010).

Die Betrachtung der Entwicklungspipeline lässt erwarten, dass auch in Zukunft ein beachtlicher Teil an Neuzulassungen erfolgt, deren Herstellung ein hohes Wertschöpfungspotential birgt. Dennoch kann in dieser Branche kein schneller Return-on-Invest erwartet werden. Der Hauptgrund dafür ist, dass die Produkt- und Prozessentwicklungszeiten von acht bis zwölf Jahren sehr lang sind und die Entwicklung von NBEs¹ risikobehaftet ist, da im Schnitt nur ca. 30 % der Kandidaten für neue Produkte die klinischen Studien bestehen (Werner, 2004; Dickson and Gagnon, 2004; Grabowski, 2008).

Die Entwicklung innovativer Arzneimittel ist daher mit hohen Kosten verbunden, und wird in der Literatur abhängig von Medikament, herstellendem Unternehmen und Berechnungsgrundlage mit 500 Mio. bis 2000 Mio. US-Dollar geschätzt (DiMasi *et al.*, 2003; Adams and Brantner, 2006; Adams and Brantner, 2010; Vernon *et al.*, 2010).

Bisher sichert eine Kombination aus definierten, unflexiblen Herstellschritten und umfangreichen Tests die Qualität von Arzneimitteln. Diese Methodik wird häufig als „Quality by Testing“ (QbT) bezeichnet (Yu, 2008). Zukünftig sollen jedoch die PAT²-Initiative und der QbD³-Ansatz der FDA⁴ bzw. ICH⁵ zur Entwicklung von robusten Herstellprozessen und damit zur Qualitätssicherung im Arzneimittelsektor beitragen. Durch diese Ansätze wird der bisherige Fokus, der auf die Endbewertung produzierter Arzneimittel gerichtet war, weiter nach vorn auf die kontinuierliche Verbesserung des Entwicklungs- und Herstellungsprozesses verschoben (FDA, 2004; ICH, 2009; Schmidt-Bader, 2010).

¹ NBE - New Biological Entity

² PAT - Process Analytical Technology

³ QbD - Quality by Design

⁴ FDA - U.S. Food and Drug Administration

⁵ ICH - International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use

Zur Erlangung eines besseren Prozessverständnisses empfiehlt sich der Einsatz prozessanalytischer Technologien bereits in den frühen Phasen der Prozessentwicklung. Durch die Kombination von analytischen und wissensbasierten Methoden können aus Daten Informationen gewonnen werden, die den Zusammenhang von Produktqualität/-quantität und Produktionsprozess verdeutlichen. Das aufgebaute Wissen kann direkt in das validierte Umfeld übertragen sowie einer kontinuierlichen Weiterentwicklung zugeführt werden und so in optimierten Produktionsprozessen resultieren (Dünnebier and Tups, 2007; Herwig, 2010; Maiwald, 2010).

1.2 Zielsetzung der Arbeit

Das Ziel dieser Arbeit war es, durch Kombination von Methoden und Werkzeugen der klassischen Bioverfahrenstechnik sowie der Chemometrie und Informatik, Bioreaktionsprozesse zur Herstellung pharmazeutischer Proteine mit *Pichia pastoris* bzw. *Escherichia coli* zu automatisieren und optimale Betriebsparameter zu identifizieren.

Damit die Einbindung bioverfahrenstechnischer Operationen in einen vollautomatisierten Prozess einwandfrei erfolgt, muss die Gesamtanlage Möglichkeiten zur Messung, Steuerung und Regelung prozessrelevanter Größen gewährleisten. Die vorhandene Automatisierungsinfrastruktur sollte daher zunächst über die Integration von kommerziell erhältlicher Inline-, Online- und Atline-Messtechnik erweitert werden.

Auf der Basis von verlässlichen Prozessdaten sowie der im Prozessleitsystem MFCS/win implementierten Rezeptursteuerung sollte eine robuste Grundautomatisierung geschaffen werden, die es ermöglicht, die durchzuführenden biotechnologischen Prozesse zuverlässig, reproduzierbar und zielgerichtet unter optimalen Bedingungen ablaufen zu lassen. Diese sollte unter Berücksichtigung der am Bioprozess gegebenen Möglichkeiten durch ein zu entwickelndes prozessspezifisches Software-Tool auf der Basis von MATLAB[®] ergänzt werden.

Im ersten Teil dieser Arbeit sollte die experimentelle Umsetzung der ausgearbeiteten Konzepte durch Kultivierung der Hefe *P. pastoris* erfolgen, die das herzustellende Protein in das umgebende Kulturmedium sekretiert. Dadurch ergab sich die Möglichkeit, den klassischen dreiphasigen Produktionsablauf mit einer unlimitierten Batch-Phase sowie einer substratlimitierten Fed Batch- und Produktionsphase um primäre Downstream-Schritte zu einem vollautomatisierten Integrierten Bioprozess zu komplettieren.

Im zweiten Teil dieser Arbeit sollte die Ermittlung optimaler Betriebsparameter über den Einsatz Statistischer Versuchsplanung durch Kultivierung von Mutanten des weitgehend erforschten prokaryontischen Expressionssystems *E. coli* in einem Multi-Bioreaktorsystem erfolgen. In Screeningversuchen sollten Variablen identifiziert werden, die einen signifikanten Einfluss auf die Erhöhung der Gesamtausbeute aktiver Proteine in der löslichen Fraktion haben. In darauf folgenden Optimierungsversuchen sollte dann der Wertebereich der signifikanten Faktoren festgelegt und schließlich in einem Robustheitstest überprüft werden.

2 Eingesetzte mikrobielle Expressionssysteme und hergestellte Produkte

2.1 Extrazelluläre Produktion von chemotaktischen Zytokinen mit *P. pastoris*

2.1.1 Die methylotrophe Hefe *Pichia pastoris* SMD1168

Die fakultativ methylotrophe Hefe *P. pastoris* gehört zur Pilzgruppe der Ascomyceten und ist für die Expression heterologer Proteine neben *E. coli* eines der am besten charakterisierten Systeme. Seit der Markteinführung wurden zahlreiche Proteine aus Bakterien, Viren, Pilzen, Protisten, Pflanzen, Wirbellosen und Wirbeltieren, den Menschen inbegriffen, mit Hilfe dieses Systems exprimiert (Cregg *et al.*, 1993; Cregg *et al.*, 2000; Daly and Hearn, 2005).

Hefen bieten im Gegensatz zu *E. coli* eine Anzahl von Vorteilen in Bezug auf die Expression komplexer Proteine. Als einzellige Organismen ähneln sie trotz ihrer geringen Größe der Komplexität von Pflanzen- und Tierzellen. Sie besitzen daher einerseits die Vorteile bakterieller Systeme hinsichtlich Manipulation und Wachstum sowie andererseits die Vorteile eukaryotischer Systeme hinsichtlich der Fähigkeit, eine posttranslationale Prozessierung und die Modifikation von Proteinen durchzuführen (Buckholz and Gleeson, 1991; Li *et al.*, 2007).

Die in dieser Arbeit verwendete Hefe *P. pastoris* SMD1168 mit dem Genotyp *his4 pep4Δ* wurde von der Fa. Invitrogen bezogen. Der Mut^S-Stamm weist neben der Auxotrophie für Histidin, die für die Isolierung erfolgreich transformierter Klone genutzt wird, eine Protease-defizienz auf. Durch die Deletion des für die Protease A kodierenden Gens *pep4* kann die Ausbeute an funktionellem, rekombinanten Protein erhöht werden (Weiss *et al.*, 1995; Gleeson *et al.*, 1998).

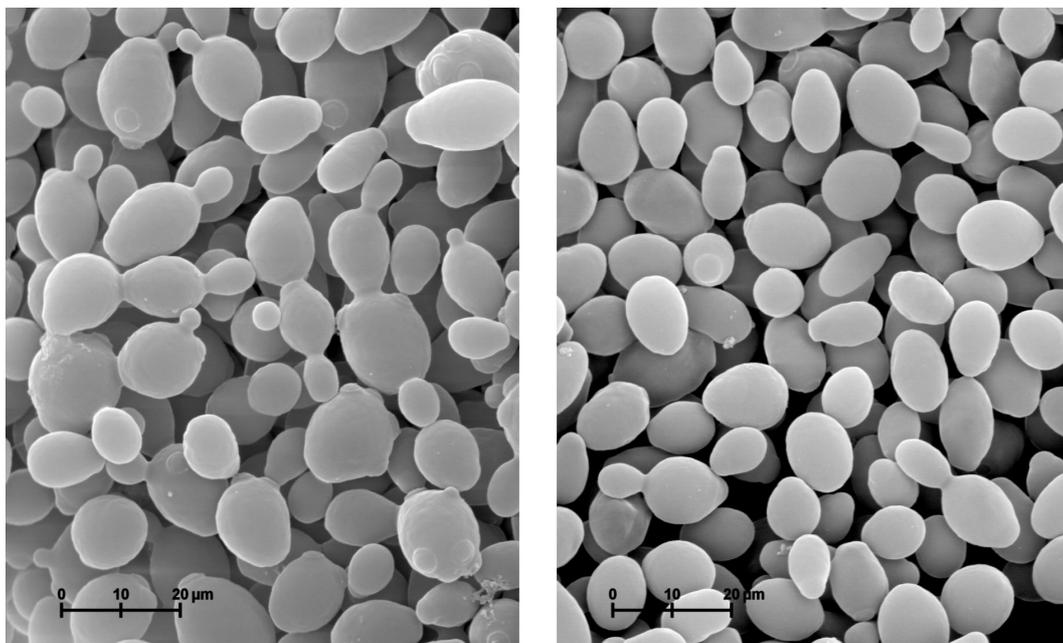


Bild 2.1: Elektronenmikroskopische Aufnahme von *Pichia pastoris* bei Wachstum auf Glycerol (links) und Methanol (rechts), Scanning Electron Microscopy (SEM) Aufnahme mit 1600-facher Vergrößerung (Dennis Kunkel Microscopy, Inc.)

P. pastoris besitzt zwei Gene *AOX1* und *AOX2* zur Bildung von Alkoholoxidase (AOX), wobei *AOX1* für den überwiegenden Teil der Enzymaktivität in den Zellen verantwortlich ist (Couderc and Baratti, 1980; Roggenkamp *et al.*, 1984; Tschopp *et al.*, 1987). Damit ist der Stamm in der Lage, neben Glycerol (oder Glukose) auch Methanol als einzige Energie- und Kohlenstoffquelle zu nutzen. Bild 2.1 zeigt eine SEM-Aufnahme der Zellen bei Wachstum auf den verschiedenen Substraten.

Bei Wachstum auf Glycerol ist eine stärkere Sprossung klar erkennbar und es kommt zu einer vollständigen Repression der *AOX*-Gene (Tschopp *et al.*, 1987). Da der *AOX1*-Promotor einer sehr effizienten Regulation unterliegt, wird er für die Expression von Fremdgenen genutzt. So kann die Hefe auf Glycerol zu einer vordefinierten Zelldichte anwachsen und durch anschließenden Wechsel zum Substrat Methanol der *AOX1*-Promotor induziert werden. AOX wird in Zellen, die auf Glycerol wachsen, vollständig abgebaut. In Zellen dagegen, die auf Methanol als einziger C-Quelle wachsen, kann das Enzym bis zu 30 % des Zellproteins ausmachen (Couderc and Baratti, 1980).

Hinsichtlich der Methanolverwertung werden drei Phänotypen von *P. pastoris* Stämmen unterschieden. Zusätzlich zum methanolverwertenden Stamm Mut⁺ (Methanol utilization positive) gibt es weitere Klone mit AOX-Mutationen. Dies sind einerseits Mut^S-Stämme (Methanol utilization slow), bei denen *AOX1* durch produktcodierende Gene ersetzt wurde, so dass dieser Stamm auf das schwächere *AOX2* zurückgreifen muss und auf Methanol wesentlich langsamer wächst. Bei Mut⁻-Stämmen (Methanol utilization negative) hingegen wurden beide *AOX*-Gene deletiert, so dass sie unfähig sind, auf Methanol als Kohlenstoffquelle zu wachsen. Die Induktion des Zielproteins bleibt dennoch erhalten.

2.1.2 Eigenschaften des Zielprodukts 1-3del I-TAC

Die Entzündung ist eine wichtige unspezifische Immunantwort des menschlichen Körpers als Reaktion auf Schädigungen und wird z. B. durch Infektionen oder Fremdstoffen hervorgerufen. Die Entzündungsreaktion wird mit der Freisetzung von Mediatoren eingeleitet, die einen kontrollierten Einstrom von Leukozyten (z. B. Makrophagen, Granulozyten, B- oder T-Lymphozyten) und Plasma aus dem Blut in den traumatisierten Bereich ermöglichen. Die Rekrutierung immunologischer Zellen in den Entzündungsherd wird als Chemotaxis bezeichnet und insbesondere durch transient exprimierte Chemokine (Chemotaktische Zytokine) koordiniert. Bild 2.2 zeigt die agonistische Wirkung von Chemokinen bei inflammatorischen Prozessen.

Chemokine sind sekretierte Polypeptide mit einem geringen Molekulargewicht von ca. 8 - 16 kDa und einer Länge von ca. 70 - 130 Aminosäuren, wovon ca. 20 - 25 Aminosäuren für eine Signalsequenz kodieren, die während der Sekretion durch Signalpeptidasen abgespalten wird. Sie werden aufgrund der Zahl und der Anordnung konservierter Cysteinreste am aminoterminalen Ende in ihrer Sequenz in die vier Subfamilien CC-, CXC-, XC- und CX3C-

Chemokine unterteilt, wobei X eine beliebige Aminosäure kodiert. Zusätzlich werden die Chemokine durchnummeriert und mit dem Buchstaben „L“ für „Ligand“ ergänzt.

Die Nomenklatur der Chemokin-Rezeptoren, die auf den Oberflächen der immunologisch aktiven Zellen gebunden sind, orientiert sich an den bindenden Liganden sowie der Kennzeichnung mit dem Buchstaben „R“. Die Chemokinfamilie besteht inzwischen aus über 50 Chemokinen und mindestens 19 Chemokin-Rezeptoren, wobei das Chemokinsystem in seiner Wirkung redundant und überlappend ist, da Chemokine an mehrere unterschiedliche Rezeptoren binden können (Baggiolini *et al.*, 1997; Murphy *et al.*, 2000; Zlotnik and Yoshie, 2000).

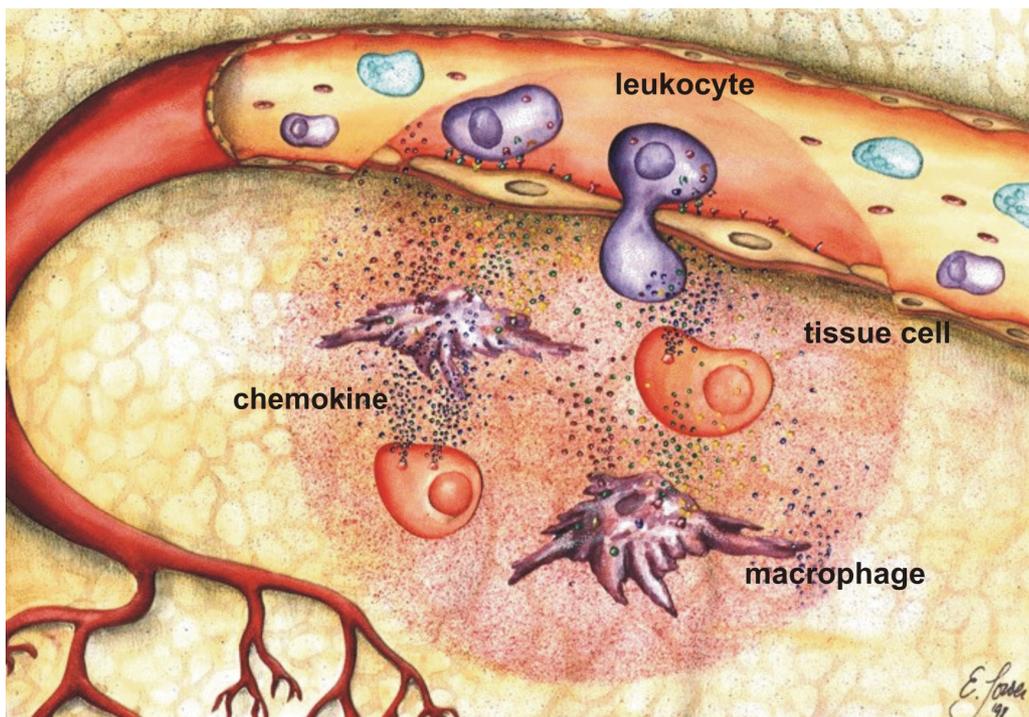


Bild 2.2: Agonistische Wirkung von Chemokinen bei inflammatorischen Prozessen (Sowa, 1998)

Funktioniert die Regulation des Chemokin-Netzwerkes, führt die Entzündungsreaktion im optimalen Fall zur Heilung. Bei chronisch entzündlichen Erkrankungen wie z. B. Asthma, Multipler Sklerose oder Rheumatoider Arthritis ist die Regulation des Chemokin-Netzwerkes jedoch gestört und führt zur lokalen Akkumulation von infiltrierenden Leukozyten (Stellato *et al.*, 2001; van der Voorn *et al.*, 1999; Matsui *et al.*, 2001).

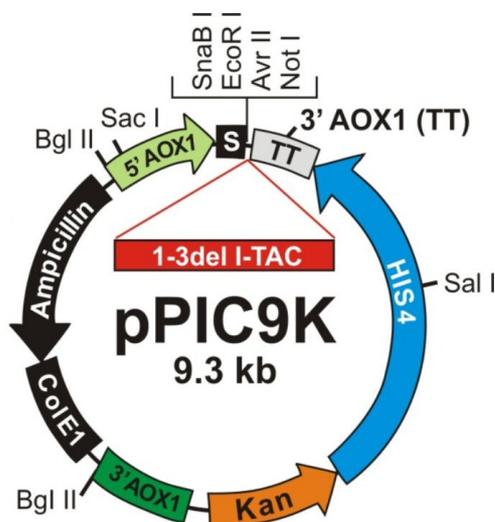
Therapeutische Strategien zielen daher darauf ab, durch Herstellung spezifischer Chemokin-Rezeptor-Antagonisten die Aktivierung und Rekrutierung von Immunzellen zum Entzündungsort und die darauf folgende Verstärkung der Immunantwort selektiv zu unterbinden. Insbesondere scheint eine Modifizierung der N-terminalen Aminosäuresequenz zu potenten Antagonisten zu führen (Gong *et al.*, 1997; Elsner *et al.*, 1997; Panzer *et al.*, 1999).

Die unterstrichenen Aminosäuren F (Phenylalanin), L (Leucin) und M (Methionin) stellen den Deletionsbereich von 1-3del I-TAC dar, die Aminosäuresequenz des Antagonisten beginnt also erst bei F4. Die grau hervorgehobenen Cysteine C9 und C36 sowie C11 und C53 bilden Disulfidbrücken aus, welche zur Stabilisierung der Proteinstruktur beitragen (Baggiolini, 2001). Die ersten beiden Cysteine C9 und C11 sind durch eine Aminosäure voneinander getrennt, wodurch die Zugehörigkeit von I-TAC zu den CXC-Chemokinen begründet wird. Bei der Nummerierung wird von der fragmentierten Proteinsequenz ausgegangen.

2.1.3 Herstellung und Transformation des Expressionsvektors pPIC9K

Pichia pastoris besitzt keine eigenen stabilen, episomalen Plasmide, so dass zur Produktion rekombinanter Proteine ins Genom integrierende Vektoren verwendet werden. Grundsätzlich sind *Pichia*-Vektoren sog. „shuttle“-Vektoren, die also auch für die Propagierung in *E. coli* geeignet sind (Weiss and Reiländer, 1999).

Um die im *E. coli*-Hintergrund ablaufende Klonierung zu erlauben, besitzt der in Bild 2.5 dargestellte und in dieser Arbeit verwendete pPIC9K-Vektor der Fa. Invitrogen ein ColE1-Replikon mit dem oriV als Replikationsursprung und eine in *E. coli* wirksame Antibiotikaresistenz als Selektionsmarker.



Zwischen der 5'- und 3'-AOX1-Region ist eine multiple Klonierungsstelle lokalisiert, die die Insertion des Zielgens ermöglicht. Nach PCR-Amplifizierung der 1-3del I-TAC Ziel-DNA wird diese mit den Restriktionsenzymen EcoRI und NotI in den geschnittenen Vektor pPIC9K kloniert. Durch die Transformation erhält *E. coli* zwei Resistenzen, Ampicillin und Kanamycin. Eine anschließende Anzucht auf LB-Amp-Platten ermöglicht aus positiven Klonen Plasmid-DNA zu isolieren (Rubant, 2005).

Bild 2.5: Der Expressionsvektor pPIC9K

Vor der Transformation in *P. pastoris* SMD1168 wird der aus *E. coli* isolierte Vektor über Restriktion linearisiert. Durch die Wahl der Linearisierungsschnittstelle kann der Integrationsort im *Pichia*-Genom bestimmt und damit der „methanol-utilization“-Phänotyp der Wirtszellen verändert werden. In dieser Arbeit wird die Mut^S-Rekombinante favorisiert, mit dem Ziel, eine reine Wachstumsphase der Hefe auf Glycerol sowie eine wachstumsreduzierte Produktionsphase des Zielproteins auf Methanol sicherzustellen.

Eine Restriktion mit Sall und SacI erzeugt freie DNA-Enden, die eine homologe Rekombination in den *HIS4*- bzw. den *AOX1*-Locus stimulieren. Bei dieser Art der Linearisierung bleibt der *Mut*⁺-Phänotyp des histidinauxotrophen SMD1168-Stamms erhalten. Daher wurde eine Restriktion mit BglII gewählt, bei der zwei gleichzeitige cross over-Reaktionen zum Verlust des *AOX1*-Wildtypgens (Bild 2.6) führen und ca. 10 - 20 % der auf einem histidindefizienten Medium selektierten Transformanten den favorisierten *Mut*^S-Phänotyp aufweisen (Weiss and Reiländer, 1999; Rubant, 2005).

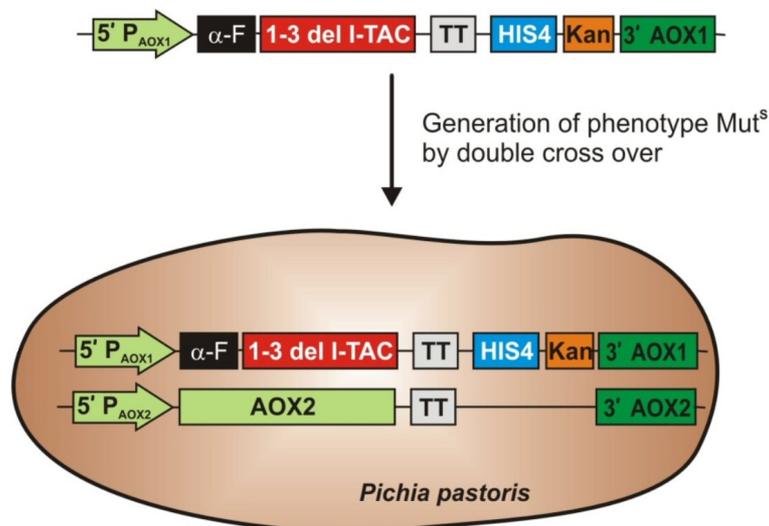


Bild 2.6: Integration des linearisierten Vektorfragments in das *AOX1*-Gen von *P. pastoris*

Das integrierte Vektorfragment enthält, flankiert von der 5'- und 3'-*AOX1*-Region, neben der kodierenden Sequenz für das Zielprotein ein Histidinol-Dehydrogenase-Gen *HIS4* als weiteren Selektionsmarker. Das Kanamycingen *Kan* verleiht Resistenz gegen Geneticin[®] in *Pichia* und wird für die Selektion von Klonen mit multiplen Insertionen benutzt (Cregg *et al.*, 1985; Scorer *et al.*, 1994).

Da sich die Strukturgene des Chemokins 1-3del I-TAC hinter der AOX-Promotorregion befinden, kann die methanolinduzierte Expression unter die Kontrolle des *AOX1*-Promotors gestellt werden. Das *TT-AOX1*-Gen sorgt für eine effiziente Transkriptionstermination und Polyadenylierung der mRNA. Generell sekretiert *Pichia* kaum native Proteine ins Medium (Cregg *et al.*, 1993). Umso effektiver lassen sich rekombinante Proteine exportieren, denen die α -Faktor-Propeptid-Signalsequenz aus *Saccharomyces cerevisiae* vorgeschaltet ist.

Die Sekretion des rekombinanten Fusionsproteins α -F-(1-3)del I-TAC in das Kulturmedium ermöglicht die Primäraufreinigung über eine direkt an den Bioreaktor angeschlossene Crossflow-Filtrationseinheit. Die molekularbiologischen Arbeiten zur Herstellung des Expressionsstamms wurden vom Pharmazentrum Frankfurt der Johann-Wolfgang-Goethe Universität, Frankfurt a. M. durchgeführt, während die Kultivierung des Klons unter anderem im Rahmen dieser Arbeit erfolgte.

2.2 Intrazelluläre Produktion von Rezeptor-Tyrosinkinasefragmenten mit *E. coli*

2.2.1 Die humane Rezeptor-Tyrosinkinase hEphB2

Die zelluläre Signaltransduktion ist eine essentielle Grundvoraussetzung für das Leben einzelner oder mehrzelliger komplexer Organismen, um angemessen auf Veränderungen in ihrer Umwelt zu reagieren. Ein Signalweg besteht aus Faktoren, Rezeptoren, Adapterproteinen, Enzymen und Second-Messenger-Molekülen, die kaskadenartig ein Signal zur Aktivierung oder Hemmung weitergeben.

Rezeptor-Tyrosinkinasen (RTKs) gehören zu einer Gruppe von membranständigen Rezeptoren mit einer intrinsischen Tyrosinkinaseaktivität (Offermanns and Rosenthal, 2004). Die Tyrosinkinase katalysiert die Übertragung des endständigen γ -Phosphats von ATP auf den Tyrosinrest eines Proteins. Nach Bindung eines Liganden werden intrazelluläre Rezeptor-Domänen autophosphoryliert und dadurch unterschiedliche Signalwege in der Zelle initiiert (van der Geer *et al.*, 1994; Hubbard and Till, 2000).

Eph-Rezeptoren sind mit zurzeit 14 im humanen Genom identifizierten Mitgliedern die größte Familie der Rezeptor-Tyrosinkinasen (Pasquale, 2008). Die Bezeichnung Eph¹ bezieht sich auf die Erstbeschreibung eines Mitglieds in einem erythropoietinproduzierenden hepatozellulären Karzinom des Menschen (Hirai *et al.*, 1987). Eph-Rezeptoren interagieren mit membran gebundenen Liganden, den Ephrinen², und werden wie die Liganden aufgrund ihrer Struktur- und Bindungseigenschaften der A- bzw. B-Klasse zugeordnet (Eph Nomenclature Committee, 1997).

Bild 2.7 zeigt die Struktur eines EphB2/Ephrin-B2-Komplexes. Die extrazelluläre Domäne des EphB2-Rezeptors besteht aus zwei Fibronectin-Typ-III-Repeats, einem cysteinreichen EGF³-ähnlichen Motiv und einer aminoterminalen globulären Ephrin-Bindungsdomäne. Der zytoplasmatische Anteil des Rezeptors besteht aus vier funktionellen Einheiten; der Juxtamembranregion mit zwei konservierten Tyrosinresten, der ATP-bindenden Tyrosinkinasedomäne sowie zwei C-terminalen Protein-Interaktionsdomänen. Der transmembrane Ephrin-B2-Ligand enthält eine extrazelluläre Eph-Bindungsdomäne sowie eine zytoplasmatische Region mit einer PDZ-Domäne (Himanen *et al.*, 1998; Binns *et al.*, 2000; Himanen *et al.*, 2001).

Da sowohl der EphB2-Rezeptor als auch der Ephrin-B2-Ligand membrangebunden vorliegen, ist eine Zell-Zell-Interaktion zur Aktivierung der Tyrosinkinase und damit der intrazellulären Vorwärts- bzw. Rückwärtssignalübertragung notwendig (Kullander and Klein, 2002; Himanen *et al.*, 2007).

¹ Eph - Erythropoietin producing human hepatocellular carcinoma

² Ephrin - Eph family receptor interacting protein

³ EGF - epidermal growth factor

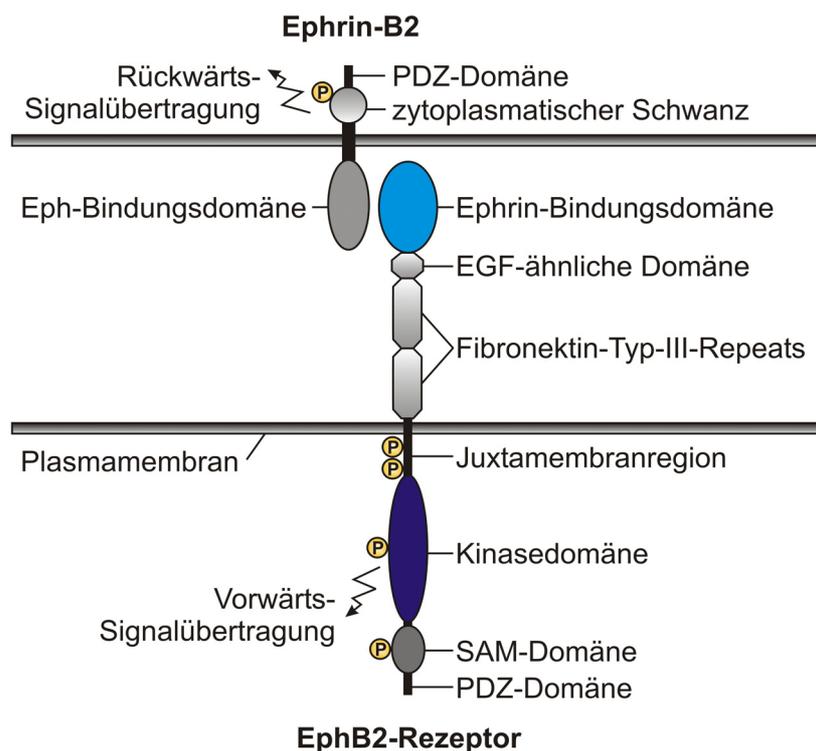


Bild 2.7: Struktur des EphB2-Rezeptor/Ephrin-B2-Komplexes (verändert nach Kullander and Klein, 2002)

Neben einer Reihe anderer biologischer Funktionen spielen Eph-Rezeptoren und Ephrine, deren Aktivierung oder Überexpression mit zahlreichen Arten von Krebs assoziiert ist (Übersicht in Surawska *et al.*, 2004), eine wichtige Rolle bei der arteriovenösen Differenzierung des Blutgefäßsystems (Angiogenese). Klinische und histopathologische Befunde an menschlichen Tumoren weisen darauf hin, dass der Anschluss solider Tumoren an das Blutgefäßsystem (Tumorangiogenese) dem ungehemmten Wachstum und der weiteren Progression vorausgeht (Hanahan and Folkman, 1996; Folkman, 2006; Baeriswyl and Christofori, 2009; Moserle *et al.*, 2009).

Therapeutische Ansätze zur Hemmung der Blutzufuhr zu Tumoren zielen daher darauf ab, Tumorgefäße selektiv zu zerstören bzw. zu okkludieren oder die Angiogenese zu hemmen und damit das Tumorwachstum anzuhalten (Folkman, 2007; Abdelrahim *et al.*, 2010).

Für die Entwicklung von anti-angiogenischen und somit anti-proliferativen Wirkstoffen wie z. B. monoklonalen Antikörpern oder Tyrosinkinaseinhibitoren sind biochemische und röntgenkristallografische Untersuchungen bzw. NMR-Strukturanalysen der interagierenden Signalmoleküle notwendige Voraussetzung.

In der vorliegenden Arbeit dienen zwei zytoplasmatische Fragmente des humanen EphB2-Rezeptors als beispielhafte Zielproteine zur biotechnologischen Herstellung mit dem Expressionwirt *E. coli*. Die Fragmente hEphB2 (D604-S898) und hEphB2 [L767P] (D604-S898) bestehen aus 295 Aminosäuren, weisen beide ein rechnerisches Molekulargewicht von ca. 33.3 kDa auf und umfassen die gesamte katalytische Kinasedomäne.

Bild 2.8 zeigt die Aminosäuresequenz des wt-hEphB2-Fragments, in grau hervorgehoben das 23 Aminosäuren umfassende Kinase-Aktivierungssegment (G766 bis I788) mit der Autophosphorylierungsstelle Tyrosin (Y780). Die hydrophoben Aminosäuren Alanin (A), Valin (V), Leucin (L), Isoleucin (I), Phenylalanin (F) und Methionin (M) haben einen relativen Anteil von 36.6 % der gesamt vorhandenen Aminosäuren. Die Interaktion zwischen exponierten hydrophoben Proteindomänen ist förderlich für die Bildung von Proteinaggregaten (Kiefhaber *et al.*, 1991; Baneyx and Mujacic, 2004).

Durch das Einfügen der Punktmutation L767P wurde in der nativen Sequenz die häufig an der Ausbildung einer α -helikalen Struktur beteiligte Aminosäure Leucin gegen Prolin ausgetauscht. Prolin hat statt der üblichen primären Aminogruppe (NH_2) eine sekundäre Aminogruppe (NH), deren H-Atom bei der Peptid-Bindung verlorenght und deswegen für Wasserstoffbrückenbindungen zur Ausbildung der α -Helix nicht mehr zur Verfügung steht und so erheblich die Faltung des hEphB2-Fragments beeinflusst.

```

        604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624
          D P N E A V R E F A K E I D I S C V K I E
625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649
  Q V I G A G E F G E V C S G H L K L P G K R E I F
650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674
  V A I K T L K S G Y T E K Q R R D F L S E A S I M
675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699
  G Q F D H P N V I H L E G V V T K S T P V M I I T
700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724
  E F M E N G S L D S F L R Q N D G Q F T V I Q L V
725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749
  G M L R G I A A G M K Y L A D M N Y V H R D L A A
750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774
  R N I L V N S N L V C K V S D F G L S R F L E D D
775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799
  T S D P T Y T S A L G G K I P I R W T A P E A I Q
800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824
  Y R K F T S A S D V W S Y G I V M W E V M S Y G E
825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849
  R P Y W D M T N Q D V I N A I E Q D Y R L P P P M
850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874
  D C P A A L H Q L M L D C W Q K D R N H R P K F G
875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898
  Q I V N T L D K M I R N P N S L K A M A P L S S

```

Bild 2.8: Aminosäuresequenz des zytoplasmatischen wt-hEphB2-Fragments (Wybenga-Groot *et al.*, 2001)

2.2.2 Das Reef Coral Fluorescent Protein ZsGreen als Expressionsreporter

Nach der Entdeckung des Grün Fluoreszierenden Proteins GFP aus der Qualle *Aequorea victoria* (Shimomura *et al.*, 1962) und dessen Weiterentwicklung (Chalfie *et al.*, 1994; Tsien, 1998) gibt es nunmehr eine Reihe weiterer fluoreszierender Proteine, die für wissenschaftliche Forschungszwecke eingesetzt werden (Übersicht in Veith and Veith, 2005).

Eine neue Gruppe von fluoreszierenden Proteinen wurde aus nicht-biolumineszierenden Spezies von Riffkorallen isoliert (Matz *et al.*, 1999) und kommerziell durch Clontech Laboratories Inc., USA, zur Verfügung gestellt. Dazu zählt auch das Reef Coral Fluorescent Protein (RCFP) ZsGreen (syn. zFP506) aus der zu den Krustenanemonen gehörenden Gattung *Zoanthus sp.*, die in Bild 2.9 gezeigt ist.



Bild 2.9: Kolonie- und Einzelaufnahme von *Zoanthus sp.* (mit freundlicher Genehmigung von Massimo Boyer und Gudmundur Geir)

Durch den Einsatz von ZsGreen als Fusionspartner kann die Expression des Zielproteins in *E. coli* im Bioreaktor mittels nicht-invasiver 2D-Fluoreszenzspektroskopie während des Prozesses verfolgt bzw. eine einfache Offline-Proteinanalytik über einen Fluoreszenzreader etabliert werden.

Im Gegensatz zum monomeren GFP neigt ZsGreen zur Ausbildung einer tetrameren Struktur und somit zur Aggregation. Um das Protein dennoch als Expressionsreporter nutzen zu können, wurden eine Reihe von Mutationen in die ursprüngliche Aminosäuresequenz eingefügt, mit dem Ziel, ZsGreen-Mutanten zu erzeugen, die nicht aggregieren, heller leuchten und eine schnellere Reifung des Chromophors aufweisen (Yanushevich *et al.*, 2002).

ZsGreen setzt sich aus 231 Aminosäuren zusammen, hat ein Molekulargewicht von ca. 26.1 kDa und eine Sequenzhomologie zum *Aequorea* GFP von nur ca. 26 % (Matz *et al.*, 1999). Trotz der geringen Übereinstimmung in der Primärstruktur, weist ZsGreen die für alle Proteine aus der GFP-Familie typische Sekundärstruktur auf, die sog. β -can Struktur. Demnach bildet sich aus 11 β -Faltblättern ein Zylinder, in dessen Zentrum sich eine α -Helix mit dem für die Fluoreszenz verantwortlichen Chromophor befindet (Yang *et al.*, 1996).

Eine schematische Darstellung der Sekundärstrukturen von GFP (Ormö *et al.*, 1996) und ZsGreen (Pletneva *et al.*, 2007) ist in Bild 2.10 gezeigt. Das ZsGreen-Chromophor entsteht durch Zyklisierung und Oxidation des Tripeptids Asn-Tyr-Gly an den Positionen 66-68, wodurch sich eine photochemisch wirksame Struktur ausbildet (Shimomura, 1979; Cody *et al.*, 1993). Die Reaktion findet autokatalytisch statt und benötigt bis auf molekularen Sauerstoff kein Substrat oder externen Kofaktor (Barondeau *et al.*, 2003). Die Fluoreszenz von ZsGreen

kann bei einer Wellenlänge λ_{ex} von 496 nm angeregt werden, bei einer Wellenlänge λ_{em} von 506 nm wird das Licht wieder emittiert.

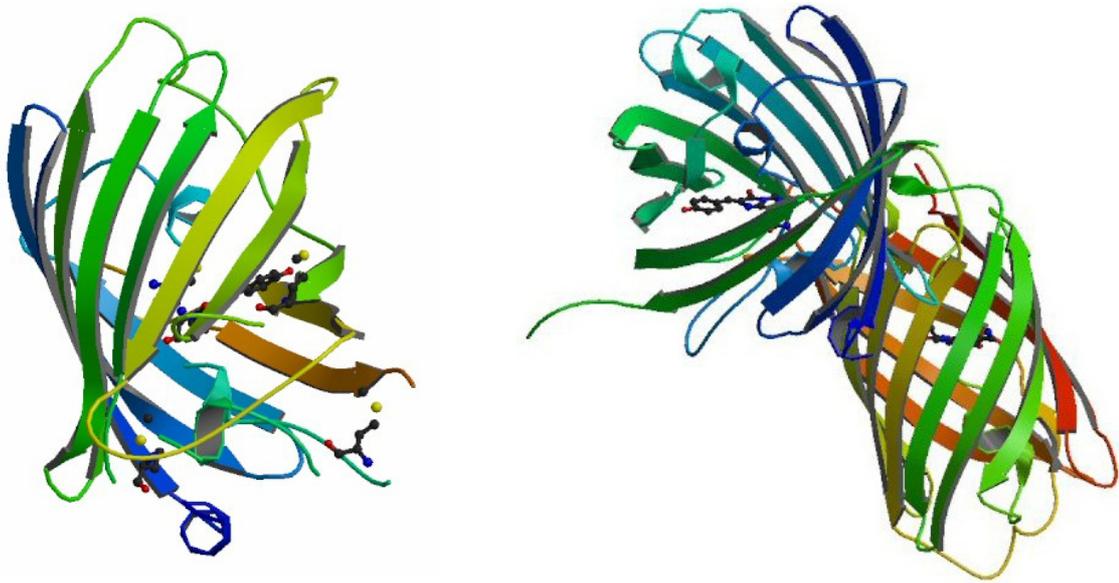


Bild 2.10: 3D-Struktur eines GFP-Monomers aus *Aequorea victoria* (links) sowie eines ZsGreen-Dimers aus *Zoanthus sp.* (rechts) auf Basis der Datenbankeinträge 1EMA und 2OJK in der PDB-Datenbank (www.pdb.org)

2.2.3 Der Stamm *Escherichia coli* BL21 (DE3)

Für die Produktion eines rekombinanten Proteins stehen zahlreiche Expressionssysteme zur Verfügung. Mit seiner Vielzahl an unterschiedlichen Stämmen und Expressionsvektoren stellt jedoch die Spezies *E. coli* das zurzeit populärste prokaryotische Wirtssystem dar (Baneyx, 1999; Schmidt, 2004).

In dieser Arbeit wurde der Laborstamm *Escherichia coli* BL21 (DE3) mit dem Genotyp $F^- \text{ompT hsdS}_B(\gamma_B^- \text{m}_B^-) \text{gal dcm}$ verwendet. Da dieser Stamm über keine spezifischen Pathogenitätsmerkmale verfügt, wird er nach der Gentechnik-Sicherheitsverordnung (GenTSV) in die Risikogruppe 1 eingeordnet. Der Sicherheitsaspekt wird durch das Fehlen des extrachromosomalen, autonom-replizierenden F-Plasmids, das Gene für den Vorgang der Konjugation enthält, unterstützt.

Wichtig für die Produktion und Aufreinigung rekombinanter Proteine ist die *ompT*-Defizienz, sowie die für alle *E. coli* B-Stämme natürlich vorhandene *lon*-Defizienz. Durch diese defekten Gene, die für die membranassoziierte Protease VII (outer membrane protein T) und die cytoplasmatische Lon-Protease kodieren, kann die Proteolyse exprimierter Proteine *in vitro* und *in vivo* reduziert und damit die Ausbeute erhöht werden (Grodberg and Dunn, 1988; Baneyx and Georgiou, 1990; Goldberg *et al.*, 1994).

E. coli BL21 (DE3) ist prädestiniert für die Verwendung von T7-Promotor getriebenen Expressionssystemen (Rosenberg *et al.*, 1987; Studier *et al.*, 1990). Bei diesem Stamm wurde über den Prophagen DE3, dessen Immunitätsregion durch die des Phagen 21 ersetzt wurde, eine Kopie des T7-RNA-Polymerasegens als exzisionsdefizientes λ -Lysogen stabil in das Chromosom integriert und unter die Kontrolle des IPTG-induzierbaren *lacUV5*-Promotors gestellt (Studier und Moffat, 1986; Davanloo *et al.*, 1984), bei dem Katabolitregulation vermittelnde Promotorbereiche mittels Mutagenese ausgeschaltet wurden.

Die schwache aber konstitutive Expression des am Lac-Operator bindenden Lac-Repressors, der vom ebenfalls chromosomal vorhandenen *lacI*-Gen kodiert wird, verhindert theoretisch bis zur Induktion mit dem Galaktosid Isopropyl- β -D-1-thiogalaktopyranosid (IPTG) die Transkription des T7-RNA-Polymerasegens (Jacob and Monod, 1961; Lewis, 2005).

Es wurde jedoch festgestellt, dass die *E. coli* RNA-Polymerase auch ohne Induktion stromaufwärts des *lacUV5*-Promotors bindet und geringe Mengen an T7-RNA-Polymerase synthetisiert. Um das Level der basalen Expression zu minimieren, wurden die Expressionssysteme BL21(DE3)pLysS und BL21(DE3)pLysE entwickelt, die das T7-Lysozym als natürlichen und selektiven Suppressor der T7-RNA-Polymeraseaktivität nutzen (Dubendorff and Studier, 1991).

2.2.4 Der Expressionsvektor pZT7 und seine funktionellen Bestandteile

Der in *E. coli* BL21 (DE3) transformierte Expressionsvektor ist eine modifizierte Version des durch die Fa. AstraZeneca, Schweden, patentierten Vektors pZT7#3.3 (Pioli *et al.*, 1998), dargestellt in Bild 2.11.

Für eine unabhängig vom bakteriellen Chromosom stattfindende Vermehrung enthält der Expressionsvektor pZT7#3.3 ein ColE1-Replikon mit dem oriV⁴ als Replikationsursprung sowie eine kurze „ColE1 resolution“-Erkennungssequenz (*cer*, 283 bp), die kodierend für ein ortsspezifisches Multimer-Resolutionssystem zur Plasmidstabilität beiträgt. Das System gewährleistet, dass das durch homologe Rekombination entstandene und durch Replikation akkumulierte Plasmidmultimer vor der Zellteilung in Monomere aufgelöst und so gleichmäßig auf die Tochterzellen segregiert werden können (Summers und Sherratt, 1984; Summers *et al.*, 1985).

Als Selektionsmarker enthält der Expressionsvektor pZT7#3.3 die auf zwei divergent angeordneten Genen *tetA* und *tetR* induzierbare Tetrazyklinresistenz. Durch die Insertion der *tet*-Gene wurde das auf dem Ursprungsvektor pAT153 (Twigg and Sherratt, 1980) vorhandene *bla*-Gen, das für das Enzym β -Lactamase kodiert, bis auf einen ca. 320 bp großen Part ersetzt.

⁴ oriV - origin of vegetative replication

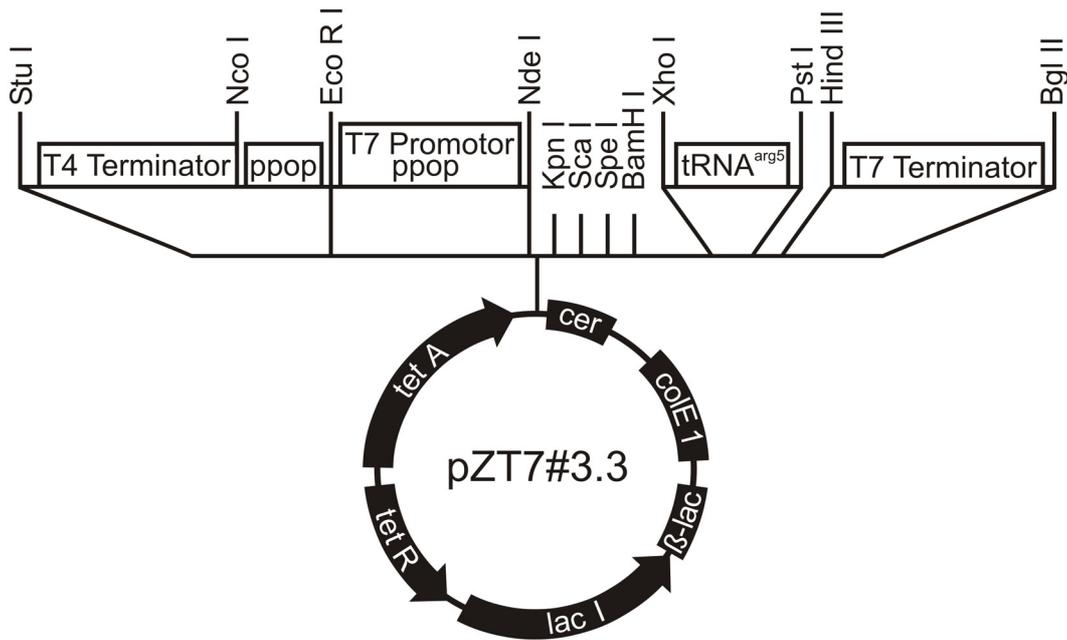


Bild 2.11: Der Expressionsvektor pZT7#3.3 mit Vektor-Elementen sowie T7-Expressionskassette mit multiple cloning site (MCS)

Tetrazyklin gelangt durch passive Diffusion über die Cytoplasmamembran in die Zelle (Sigler *et al.*, 2000) und entfaltet seine antibiotische Wirkung durch Komplexbildung mit einem zweiwertigen Metallion. Dieser Chelat-Komplex konkurriert mit der Aminoacyl-tRNA um die gleiche Bindungsstelle an der 30S-Untereinheit des prokaryontischen Ribosoms und hemmt so die Peptidelongation während der Proteinbiosynthese.

Das strukturelle Resistenzgen *tetA* kodiert für einen membranständigen Proton-Tetrazyklin-Antiporter, der den Tetrazyklin-Metall-Komplex im Austausch gegen ein Proton aktiv aus der Zelle exportiert. Das Regulatorgen *tetR* kodiert ein Repressorprotein, das die Expression von *tetA* in Abwesenheit von Tetrazyklin unterdrückt (Hillen and Berens, 1994; Chopra and Roberts, 2001).

Da die *tetA*-Sequenz keinen erkennbaren Terminator aufweist, wurde eine T4-Terminatorsequenz stromaufwärts von der *EcoRI* Schnittstelle kloniert, um ein potentielles transkriptionales Durchlesen von *tetA* in die T7-Expressionskassette zu verhindern.

Der für die T7-RNA-Polymerase hochspezifische T7-Promotor ist stromaufwärts und stromabwärts von einer perfekten palindromischen *lac*-Operatorsequenz (ppop-Sequenz) umgeben, die den Lac-Repressor im Gegensatz zur nativen *lac*-Operatorsequenz achtmal stärker binden soll (Simons *et al.*, 1984). Im erfindungsgemäßen Aspekt soll durch diesen Tandem-ppop-Operator die Basalexpression auf ein bisher nicht erreichtes Niveau reduziert werden können, ohne dabei die Menge an hergestelltem Zielprodukt negativ zu beeinflussen.

Diesen Aspekt unterstützend ist zusätzlich zum chromosomal vorhandenen *lacI*-Gen eine weitere Kopie des Gens für den Lac-Repressor auf dem Plasmid vorhanden, um eine vollständige Repression des chromosomalen *lacUV5*-Promotors sowie sämtlicher plasmidkodierten T7-Promotoren zu erreichen.

Die molekularbiologischen Arbeiten zur Herstellung der Expressionstämme wurden von der Fa. AstraZeneca, UK, durchgeführt, während die Produktion der hEphB2-Fragmente in vollautomatischen Kultivierungsprozessen im Rahmen dieser Arbeit erfolgte.

2.2.5 Expression des Fusionsproteins in Form von Inclusion Bodies

Die T7-RNA-Polymerase ist mit ca. 200-260 Polymerisationsschritten pro Sekunde bei 37 °C ungefähr fünfmal schneller als die *E. coli* RNA-Polymerase mit ca. 30-60 Polymerisationsschritten pro Sekunde, so dass das Insert ZsGreen-hEphB2-6His unter Kontrolle des T7-Promotors überexprimiert werden kann (Jahn and Jahn, 2006; Knippers, 2006; Heddle and Mazaleyrat, 2007).

Wie bei vielen rekombinanten Proteinen resultiert die Überexpression jedoch zum Teil in der intrazellulären Akkumulation von Proteinaggregaten – den Inclusion Bodies, die für *E. coli* mit einer maximalen Länge von 1 µm und einem Volumen von 0.6 µm³ charakterisiert werden (Taylor *et al.*, 1986; Bowden *et al.*, 1991; Carrió *et al.*, 1998).

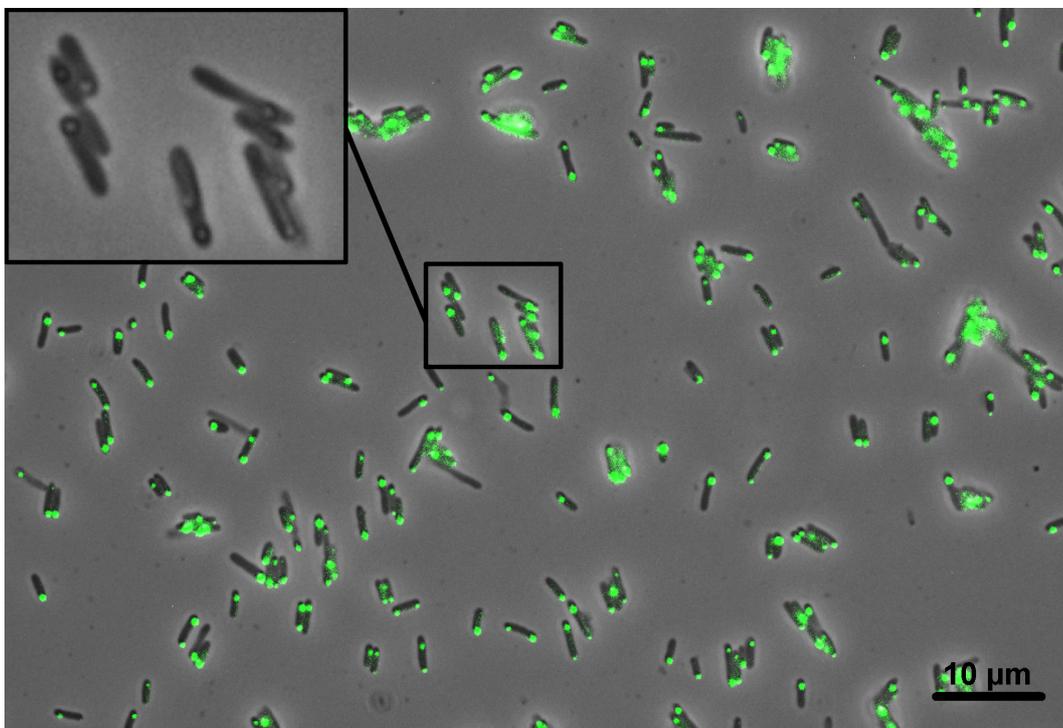
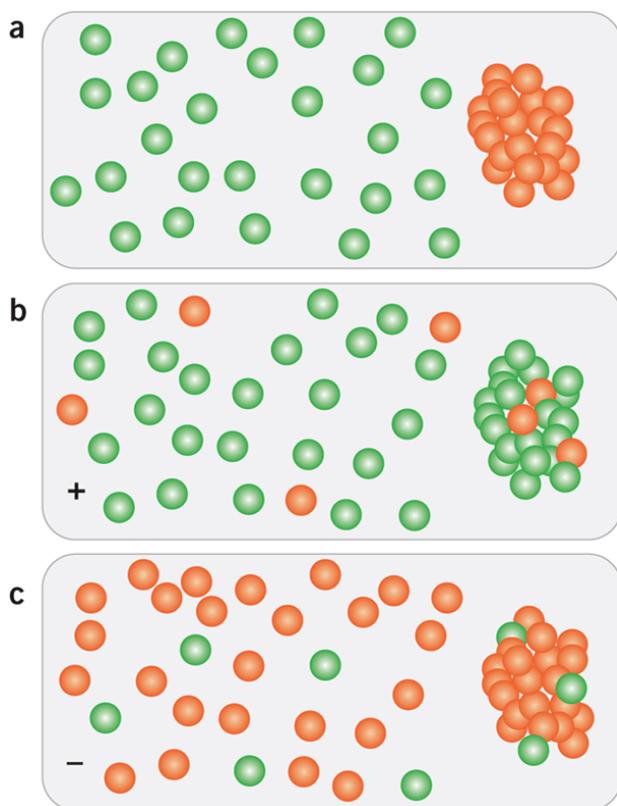


Bild 2.12: Fluoreszenzmikroskopische Aufnahme von *E. coli* MSD 5247 im induzierten Zustand bei 630-facher Vergrößerung und Anregung mit 450 - 490 nm

Bild 2.12 zeigt eine fluoreszenzmikroskopische Aufnahme von *E. coli* MSD 5247 mit fluoreszierenden Inclusion Bodies. Der vergrößerte Bildausschnitt verdeutlicht die vorwiegend unipolare Lage der annähernd eiförmig aggregierten Proteine im Zytoplasma der Zellen.

Da die fluoreszierende Eigenschaft des Chromophors erhalten ist, weist darauf hin, dass Inclusion Bodies entgegen der bisherigen Meinung zu einem signifikanten Anteil aus korrekt gefalteten und biologisch aktiven Proteinen bestehen (Martínez-Alonso *et al.*, 2009; González-Montalbán *et al.*, 2008; García-Fruitós *et al.*, 2007; de Groot and Ventura, 2006).

Um den Zusammenhang zwischen biologischer Aktivität und Proteinlöslichkeit zu verdeutlichen hilft Bild 2.13 (González-Montalbán *et al.*, 2007).



Nach der konventionellen Auffassung (a) besteht der lösliche Proteinanteil nur aus korrekt gefalteten, biologisch aktiven Proteinen (grüne Kugeln). Fehlgefaltete, biologisch inaktive Polypeptide (rote Kugeln) werden als Inclusion Bodies in der Zelle abgelagert. In diesem Modell würde die Proteinlöslichkeit ein direkter Indikator für eine korrekte Proteinkonformation und somit biologische Aktivität sein.

Nach dem neuen Modell (b, c) koexistieren Mischungen aus biologisch aktiven Proteinen und nicht-funktionalen Polypeptiden sowohl in der löslichen als auch unlöslichen Fraktion unter Bedingungen, die eine korrekte Faltung begünstigen (+) bzw. nicht begünstigen (-).

Bild 2.13: Biologische Aktivität von rekombinanten Proteinen

Die in dieser Arbeit etablierte Offline-Analytik des Zielproteins basiert auf der Fluoreszanzmessung des löslichen Proteinanteils sowie renaturierter Inclusion Bodies nach erfolgtem Zellaufschluss. Es wurde kein Vergleich der Proteinquantifizierung über elektrophoretische Methoden mit der Fluoreszenzmessung durchgeführt, und somit keine Korrelation zwischen biologischer Aktivität und Proteinlöslichkeit hergestellt.

3 Monitoring der Prozesse

3.1 Einbindung einer Bioprozessanalytik

Die Bioprozessanalytik beschäftigt sich mit Analyseverfahren und -techniken, um biotechnologische Prozesse detailliert zu beobachten, zu kontrollieren und hinsichtlich der Quantität und der Qualität des Produktes zu optimieren (Endres *et al.*, 2009). Die dabei eingesetzten Sensoren lassen sich nach verschiedenen Gesichtspunkten klassifizieren, wobei sich in der biotechnologischen Prozessanalytik eine Einteilung nach physikalischen, chemischen und biologischen Messgrößen durchgesetzt hat (Fiechter *et al.*, 1994; Rehbock *et al.*, 2008).

Besondere Anforderungen werden an den Einsatz von Prozessanalytik in der (bio-) pharmazeutischen Industrie gestellt. Neben rechtlichen Rahmenbedingungen, die ein GMP-Umfeld mit sich bringt, sind insbesondere günstige Anschaffungspreise sowie niedrige Instandhaltungskosten, einfache und vereinheitlichte Bedienkonzepte, robuste und wartungsfreie Gerätekonzepte, CIP- und SIP-Möglichkeit der Sensoren sowie steriltechnisch geeignete Probenahmesysteme von Bedeutung (Scott and McLeod, 2010).

Als Folge dieser erhöhten Anforderungen an Wirtschaftlichkeit, Zuverlässigkeit, Robustheit und Einfachheit beschränken sich Inprozesskontrollen bei einer Vielzahl von industriellen GMP-Prozessen noch auf die etablierte Standardtechnik. Dazu zählen die Messung und ggf. Regelung von Temperatur, pH- und pO₂-Wert, Druck, Schaumhöhe, Rührergeschwindigkeit und Leistungseintrag, Begasungs- und Zufütterraten, sowie Gewicht und Abgaszusammensetzung (Sonnleitner, 2000; Becker *et al.*, 2007; Chattaway *et al.*, 2008).

Derzeit wird jedoch in großen, global agierenden Unternehmen der wissenschaftliche, technische und regulatorische Rahmen für die Umsetzung von PAT/QbD-Inhalten gestaltet (Schmidt-Bader, 2010). Aus dem risikobasierten Ansatz resultiert ein Monitoring von kritischen Prozessparametern (CPP¹) über den Einsatz geeigneter Inline-, Online- und Atline-Analytik zur Messung von Metaboliten, Substraten, Medienbestandteilen, Biomasse oder Zielprodukt. Zusätzlich soll über die Anwendung multivariater Tools und modellbasierter Ansätze das Prozessverständnis erhöht und eine gesicherte Wissensbasis für eine kontinuierlichen Qualitätssicherung, Prozessvalidierung und -weiterentwicklung geschaffen werden.

Bild 3.1 zeigt die Klassifizierung von Messverfahren in Anlehnung an die Definitionen der FDA (FDA, 2004). Demnach erfolgt die Einteilung im Wesentlichen nach der Frequenz der Einzelmessung sowie dem Messort. Die hohe Abtastrate prädestiniert die Inline- und Online-Analytik für eine quasistetige Prozessregelung. Jedoch kann auch eine Atline- bzw. zeitnahe Offline-Analytik für Monitoring und Control Zwecke eingesetzt werden (Schügerl, 2001; Peuker *et al.*, 2004; Klockowa *et al.*, 2008). Generell bestimmt die Dynamik eines Prozesses, ob Daten mit einer entsprechenden zeitlichen Verzögerung oder auch als diskrete Messwerte für eine Regelung genutzt werden können oder nicht (Friehs *et al.*, 2006).

¹ CPP - Critical Process Parameter

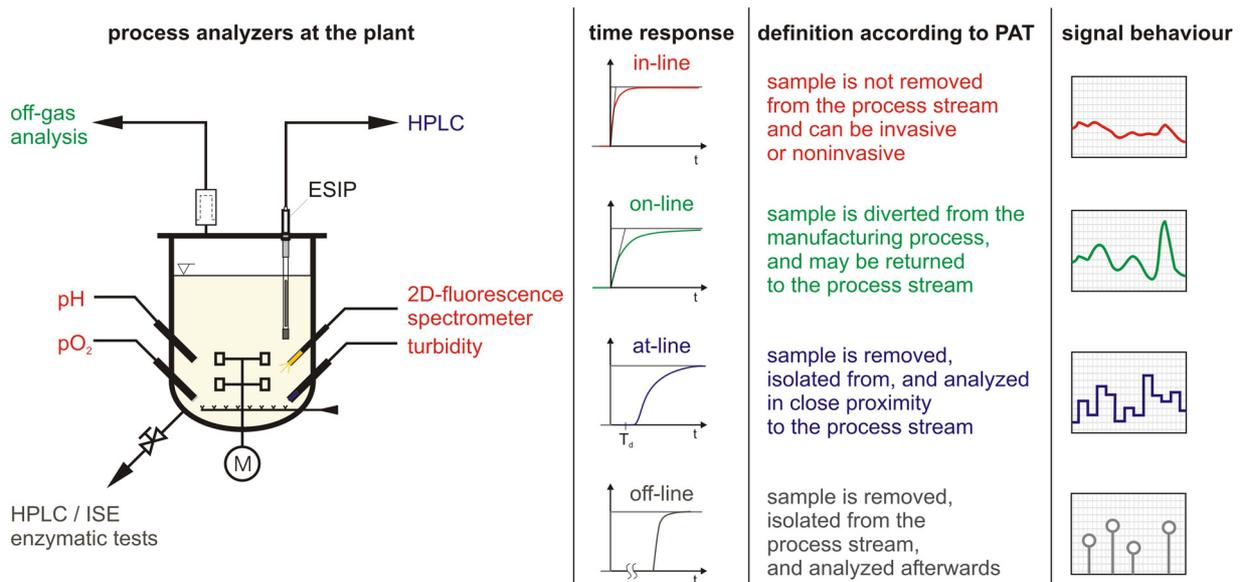


Bild 3.1: Definition der Art der Messverfahren (Peuker, 2006)

Neben den bereits genannten Anforderungen ist eine direkte Anbindung der Analysensysteme an das Prozessleitsystem (PLS) notwendig, damit einerseits die Messgrößen unmittelbar zur Steuerung oder Regelung genutzt werden können, andererseits aus aufgezeichneten Werten bzw. daraus abgeleiteten Informationen die Erkennung und Kontrolle von kritischen Prozesszuständen und Qualitätsparametern möglich wird. In vielen Fällen existiert jedoch noch keine „Plug&Play“-Implementierung der unterschiedlichen Analysensysteme, so dass trotz proprietärer Treiber bzw. offener Schnittstellen-Standards wie z. B. OPC spezielle Entwicklungsarbeiten geleistet werden müssen.

Im weiteren Verlauf dieses Kapitels werden die wichtigsten eingesetzten Messsysteme mit Bezug auf den jeweiligen Kultivierungsprozess von *P. pastoris* bzw. *E. coli* vorgestellt. Dabei wird auch auf die Einbindung der Signale in das verwendete Prozessleitsystem MFCS/win der Fa. Sartorius Stedim Systems eingegangen.

3.2 Prozessüberwachung mit Inline-, Online- und Atline-Messtechnik

3.2.1 Bestimmung der Zelldichte durch Trübungsmessung

Die Zelldichte gehört zu den wichtigsten Variablen einer Kultivierung. Neben den konventionellen Verfahren zur Offline-Bestimmung ist es möglich, diesen Prozesswert inline über direkt in den Prozess eingebrachte *in situ*-Sensoren zu erfassen. Dazu zählt insbesondere die Trübungsmessung zur Abschätzung der Zelldichte, gelegentlich auch die Impedanzspektroskopie zur Bestimmung der Lebendzellzahl oder die *in situ*-Mikroskopie zur Zellzahlbestimmung bzw. gleichzeitigen Beurteilung von Zellgröße und Zellzustandsverteilung (Olsson and Nielsen, 1997; Kiviharju *et al.*, 2008; Höpfner *et al.*, 2010).

Die Trübungsmessung wurde mit den in Bild 3.2 gezeigten Sonden ASD25-BT-N sowie ASD19-N-EC-05 der Fa. optek-Danulat durchgeführt. Das Messprinzip der Sonden basiert auf einer quantitativen Lichtmessung außerhalb des sichtbaren Spektrums, bei der die Schwächung der Lichtintensität beim Durchstrahlen der Kulturbrühe mit einer Leuchtdiode über eine Photodiode erfasst wird. Durch die Messung im festgelegten NIR-Bereich von 840 bis 940 nm verringert sich der Farbeinfluss des Mediums auf den Trübungswert (Steinke *et al.*, 2009).



Bild 3.2: Die eingesetzten Trübungssonden ASD25-BT-N (links) sowie ASD19-N-EC-05 (rechts) mit einer optischen Pfadlänge von 10 mm bzw. 5 mm

Die Trübungssonden wurden entweder über einen 25 mm Port im unteren Stutzenkranz (ASD25-BT-N) bzw. über einen 19 mm Port im Deckel (ASD19-N-EC-05) in den Bioreaktor eingebaut und das Messsignal über einen analogen Spannungseingang der lokalen Reaktorbedieneinheit DCU (Digital Control Unit) zugeführt. Das spektrale Absorptionsmaß S_{turb} [AU] wird von der DCU an das Prozessleitsystem MFCS/win übertragen und dort über die Implementierung eines exponentiellen Ansatzes in eine Zelldichte $c_{\text{XL,turb}}$ [gl^{-1}],

$$P. \text{ pastoris: } c_{\text{XL,turb}}(t) = K_1 \cdot e^{K_2 \cdot S_{\text{turb}}(t)}, \quad (3.1)$$

bzw.

$$E. \text{ coli: } c_{\text{XL,turb}}(t) = K_1 \cdot (e^{K_2 \cdot S_{\text{turb}}(t)} - 1), \quad (3.2)$$

umgerechnet. Die Zelldichte kann somit online beobachtet und für eine automatisierte Prozessführung herangezogen werden.

Die Identifikation der stammspezifischen Parameter K_1 und K_2 ,

$$P. \text{ pastoris: } K_1 = 3.9 \text{ gl}^{-1}, K_2 = 1.15 \text{ AU}^{-1},$$

$$E. \text{ coli: } K_1 = 0.119 \text{ gl}^{-1}, K_2 = 2.60 \text{ AU}^{-1},$$

erfolgte postexperimentell aus mehreren Kultivierungen durch Ermittlung der jeweiligen statischen Kennlinie über eine nichtlineare Regression.

In Bild 3.3 ist beispielhaft die Kennlinie für die Trübungssonde ASD19-N-EC-05, die ausschließlich bei *E. coli* Kultivierungen eingesetzt wurde, dargestellt. Aus den Rohdaten S_{turb} konnte eine gute Abschätzung für c_{XL} im Bereich [10 - 60 g l^{-1}] mit $R^2 = 0.958$ gefunden werden.

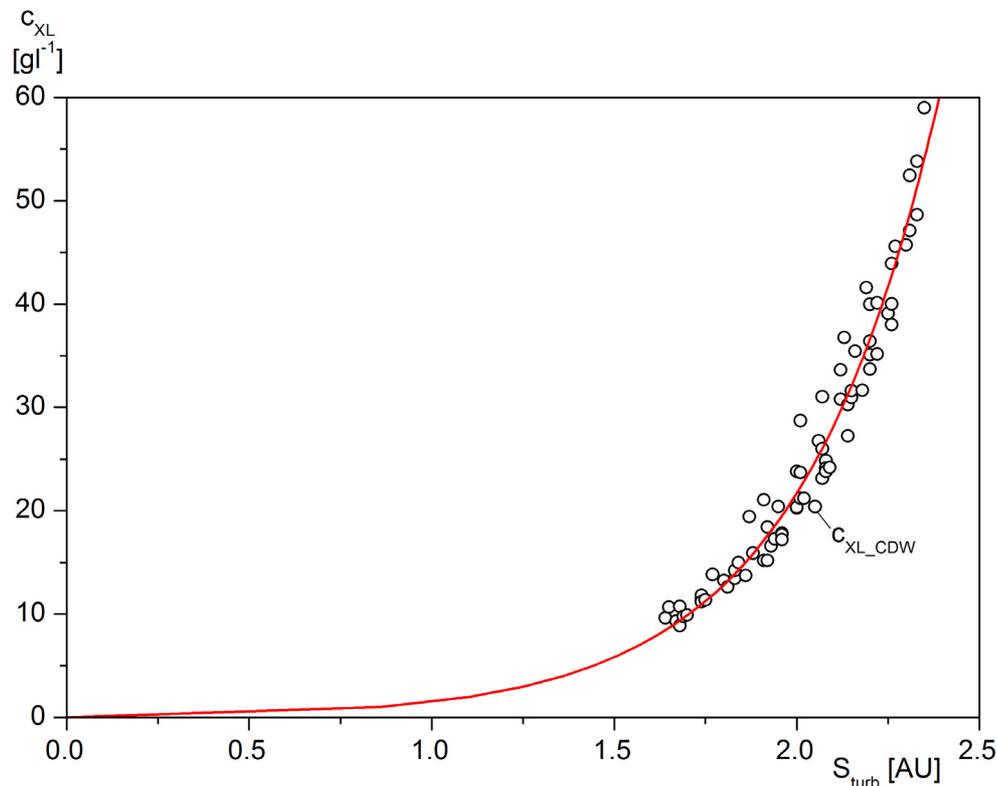


Bild 3.3: Kennlinie der Trübungssonde ASD19-N-EC-05 (*E. coli* Kultivierungen)

3.2.2 Abgasmessung für O_2 und CO_2

Die Messung von Sauerstoff und Kohlenstoffdioxid in der Abgasstrecke von Bioreaktoren ermöglicht nicht nur eine direkte Beurteilung der biologischen Aktivität bei Kultivierungen, sondern findet auch eine indirekte Verwendung über bilanzierte Variablen als maßstabsunabhängige Kriterien für ein späteres Scale-up des Prozesses sowie als Steuer- oder Regelgrößen für eine automatisierte Prozessführung (Luttmann *et al.*, 1990; Levisauskas *et al.*, 1996; Kaiser *et al.*, 2008; Herwig, 2010).

Es wurden prozessabhängig unterschiedliche Geräte für die Abgasanalytik eingesetzt. Der für *P. pastoris* Kultivierungen eingesetzte Gasanalysator MULTOR 610 der Fa. Maihak ermittelt den Stoffmengenanteil von O_2 über ein paramagnetisches Messprinzip und den CO_2 -Anteil optisch über ein nicht-dispersives Infrarotverfahren (NDIR). Das Gerät misst beide Komponenten simultan und sendet die Signale über einen CAN-I/O-Knoten an den Prozessleitcomputer, der mit einem PCI-CAN-Interface ausgestattet ist. Der Wertetransfer in MFCS/win erfolgt über eine OPC-Server/Client Kombination, wobei die Prozesswerte durch den CAN OPC-Server (Fa. esd) an den MFCS/win OPC-Client (MFCSOPCC) weitergegeben werden.

Des Weiteren wurden Sensoren vom Typ BCP-O₂ und BCP-CO₂ der Fa. BlueSens (Bild 3.4) für *E. coli* Kultivierungen verwendet. Hier basiert die Messung von O₂ auf einer elektrochemischen Sauerstoffpumpzelle aus Zirkoniumdioxid, die CO₂-Sensoren beinhalten dagegen eine Infrarot-Strahlungsquelle.

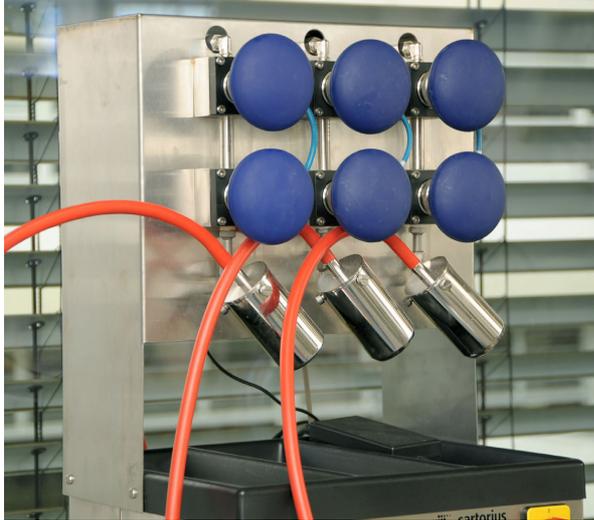


Bild 3.4: O₂- und CO₂-Abgassensoren der Fa. BlueSens

Die für die Abgasanalytik eingesetzten O₂- und CO₂-Sensoren sind über ein RJ-45 Ethernetkabel an eine Multiplexerbox BC-BACCom der Fa. BlueSens angeschlossen. Die Box bündelt die Signale der Sensoren und ermöglicht eine serielle Datenübertragung zum Prozessleitrechner über eine einzige RS-232-Schnittstelle. Für die Kommunikation mit MFCS/win wird der optional erhältliche, PLS-interne BACCom-Treiber verwendet (MFCSSBAC).

3.2.3 Verfahren zur Methanolbestimmung

Die Proteinexpression wird bei der methylotrophen Hefe *P. pastoris* über die Zugabe von Methanol ins Medium induziert. Bei einer zu niedrigen Konzentration ist der AOX1-Promotor deaktiviert, eine zu hohe Konzentration ist dagegen schädlich für die Kultur. Eine effiziente und kontinuierliche Herstellung von rekombinanten Proteinen wird daher über die Regelung der Methanolkonzentration während der Produktionsphase möglich (Gellissen, 2000; Zhang *et al.*, 2002; Cornelissen, 2004; Li *et al.*, 2007).



Bild 3.5: Flammenionisationsdetektor 3006

In den durchgeführten Kultivierungen kamen zwei verschiedene Messsysteme zum Einsatz. Bild 3.5 zeigt den im Abgasweg installierten Flammenionisationsdetektor (FID) Typ 3006 der Fa. SICK zur indirekten Bestimmung der Methanolkonzentration im Medium. Die Flammenionisationsmessung beruht auf der Messung des Ionenstroms einer Wasserstoff-Luft-Flamme in einem elektrischen Feld (Oesterle, 1994).

Im Brennsystem werden Wasserstoff, Luft und Messgas in einer offenen Flamme verbrannt. Zur Ableitung des Ionenstroms ist über der Flamme eine Elektrode angeordnet. Die reine Wasserstoffflamme hat eine geringe elektrische Leitfähigkeit. Wird methanolhaltige Abluft hinzugemischt, so steigt die Leitfähigkeit stark an. Der Ionenstrom ändert sich dabei proportional zum Methanolstoffmengenanteil x_{S2G} im Abgas.

Der FID wurde zur externen Weiterverarbeitung seiner Messwerte an einen analogen Spannungseingang der Reaktorbedieneinheit DCU angeschlossen. Das Rohsignal x_{S2G} [ppm] wird von der DCU an das Prozessleitsystem übertragen und dort über die Implementierung eines linearen Ansatzes in eine Konzentration c_{S2M} [$g\ l^{-1}$],

$$c_{S2M}(t) = K_{S2} \cdot x_{S2G}(t), \quad (3.3)$$

mit

$$K_{S2} = 1.212 \cdot 10^{-3} \text{ g l}^{-1} \text{ ppm}^{-1},$$

umgerechnet.

Das in Bild 3.6 dargestellte Messsystem FMC-MINI 03 (Fermentation Micro Computer) der Fa. Kempe Biotechnologie kann flüchtige organische Substanzen in der Flüssigphase durch eine membranbedeckte Sonde, die in den Bioreaktor eingesetzt wird, über das Tubingverfahren messen (Wagner *et al.*, 1997; Straková *et al.*, 1998).

Da die Messung temperaturabhängig ist, wird neben der eigentlichen Silicone-Tubing-Probe eine Pt-100 Temperatursonde über einen 25 mm Normstutzen am unteren Stutzenkranz in den Bioreaktor eingebracht. Für den Betrieb des Messsystems wird eine Druckluftversorgung mit 6 bar Vordruck, die frei von Kohlenwasserstoffen sein muss, benötigt. Das Messsystem ist nicht selektiv für mehrere flüchtige Substanzen und kann somit z. B. nicht zwischen Ethanol und Methanol unterscheiden.



Bild 3.6: Benchtop-Version des FMC mit Silicone-Tubing-Probe und Pt-100 (links) sowie SnO_2 -Sensor und Silikonmembran in Nahaufnahme (rechts)

Da der Trägergasstrom am Eingang des Systems kein Methanol enthält, ist die Methanoleintrittsrate \dot{m}_{S2in} gleich Null.

Die Methanolaustrittsrate \dot{m}_{S2out} ,

$$\dot{m}_{S2out}(t) = F_D \cdot c_{S2D}(t), \quad (3.5)$$

ist über den konstanten Trägergasstrom F_D , der nun die zu bilanzierende Komponente enthält, definiert.

Der Methanolstofftransport aus dem Medium in den Detektionsraum \dot{m}_{S2T} ,

$$\dot{m}_{S2T}(t) = k_{S2T} \cdot A_D \cdot [x_{S2M}(t) - x_{S2D}(t)], \quad (3.6)$$

beinhaltet neben der Stoffmengenanteildifferenz auch die Membranfläche zwischen Detektionskammer und Flüssigphase A_D und den resultierenden Methanol-Stoffübergangskoeffizienten k_{S2T} .

Die Korrelation zwischen Methanolmasse m_{S2D} und Methanolstoffmengenanteil x_{S2D} im Detektionsraum wird mit Hilfe des idealen Gasgesetzes,

$$p_D \cdot V_D \cdot x_{S2D}(t) = \frac{R \cdot T_D}{M_{S2}} \cdot m_{S2D}(t), \quad (3.7)$$

beschrieben, wobei der Druck p_D , das Volumen V_D und die Temperatur T_D der Detektionskammer als konstant angesehen werden.

Durch Umstellung der Gleichung (3.7) kann die unbekannte Methanolkonzentration in der Detektionskammer c_{S2D} ,

$$c_{S2D}(t) = \frac{m_{S2D}(t)}{V_D} = \frac{p_D \cdot M_{S2}}{R \cdot T_D} \cdot x_{S2D}(t), \quad (3.8)$$

notiert werden.

Mit Einführung der maximal möglichen Methanolkonzentration im Detektionsraum c_{S2Dmax} ,

$$c_{S2Dmax} = \frac{p_D \cdot M_{S2}}{R \cdot T_D}, \quad (3.9)$$

kann nach Einsetzen der Teilbilanzen in Gleichung (3.4) die Stoffmengenbilanz von Methanol in der Detektionskammer,

$$V_D \cdot c_{S2Dmax} \cdot \dot{x}_{S2D}(t) = k_{S2T} \cdot A_D \cdot [x_{S2M}(t) - x_{S2D}(t)] - F_D \cdot c_{S2Dmax} \cdot x_{S2D}(t), \quad (3.10)$$

hergeleitet werden.

Der Methanolstoffmengenanteil in der Medienphase x_{S2M} ,

$$x_{S2M}(t) = \frac{c_{S2M}(t)}{\rho_{S2}}, \quad (3.11)$$

kann bei bekannter Methanolkonzentration in der Medienphase c_{S2M} über die Dichte von reinem Methanol ρ_{S2} berechnet werden.

Nach Einsetzen der Gleichung (3.11) in die Gleichung (3.10) und anschließender Variablentrennung kann die Stoffmengenbilanz von Methanol in der Detektionskammer,

$$V_D \cdot c_{S2Dmax} \cdot \dot{x}_{S2D}(t) + [k_{S2T} \cdot A_D + F_D \cdot c_{S2Dmax}] \cdot x_{S2D}(t) = k_{S2T} \cdot A_D \cdot x_{S2M}(t), \quad (3.12)$$

als inhomogene lineare Differentialgleichung 1. Ordnung notiert werden.

Mit der Methanoldetektionszeitkonstante T_{S2D} ,

$$T_{S2D} = \frac{V_D \cdot c_{S2Dmax}}{k_{S2T} \cdot A_D + F_D \cdot c_{S2Dmax}}, \quad (3.13)$$

und dem Methanolverstärkungsfaktor K_{S2D} ,

$$K_{S2D} = \frac{k_{S2T} \cdot A_D}{k_{S2T} \cdot A_D + F_D \cdot c_{S2Dmax}} < 1, \quad (3.14)$$

kann die Messcharakteristik des Tubingsystems in einer vereinfachten Form über ein klassisches P-T1-Verhalten,

$$T_{S2D} \cdot \dot{x}_{S2D}(t) + x_{S2D}(t) = K_{S2D} \cdot x_{S2M}(t), \quad (3.15)$$

beschrieben werden.

Das Aufsuchen der Lösung der inhomogenen Differentialgleichung (3.15) führt auf die Beschreibung des unbekanntes Stoffmengenanteils x_{S2D} von Methanol im Detektionsraum,

$$x_{S2D}(t) = x_{S2D}(t_i) \cdot e^{-\frac{t-t_i}{T_{S2D}}} + K_{S2D} \cdot x_{S2M}(t_i) \cdot \left(1 - e^{-\frac{t-t_i}{T_{S2D}}}\right); t \geq t_i, \quad (3.16)$$

mit der Annahme, dass ab dem Messzeitpunkt $t = t_i$ der Stoffmengenanteil von Methanol in der Medienphase x_{S2M} konstant ist.

Mit

$$x_{S2M}(t_i) = \bar{x}_{S2M}, \quad (3.17)$$

liegt das stationäre Verhalten des Systems,

$$\bar{x}_{S2D} = K_{S2D} \cdot \bar{x}_{S2M}, \quad (3.18)$$

mit $K_{S2D} < 1$ ebenfalls fest.

Die Sondensprungantwort,

$$x_{S2D}(t) = h_{S2D}(t) \cdot \bar{x}_{S2M} = K_{S2D} \cdot \left(1 - e^{-\frac{t-t_i}{T_{S2D}}}\right) \cdot \bar{x}_{S2M}, \quad (3.19)$$

beschreibt das dynamische Verhalten bei Annahme eines bei t_i energiefreien Systems,

$$x_{S2D}(t_i) = 0, \quad (3.20)$$

und einer zeitinvarianten Eingangsgröße \bar{x}_{S2M} ,

$$x_{S2M}(t) = \bar{x}_{S2M}. \quad (3.21)$$

Der Widerstandswert des Methanoldetektors R_D ,

$$R_D(t) = R_{D0} + a_1 \cdot x_{S2D}(t)^{b_1}, \quad (3.22)$$

mit

R_{D0}	:= Detektorwiderstand bei Methanolabwesenheit	[Ω]
a_1	:= Widerstandsänderung bei Methanoltotalbeladung des Detektionsraumes	[Ω]
b_1	:= Korrelationskoeffizient	[-]

ist quasistationär mit dem Methanolstoffmengenanteil im Detektionsraum x_{S2D} verbunden.

Eine stationäre Betrachtung des Messverhaltens führt zur nichtlinearen Kennlinie des Messaufnehmers mit

$$\bar{R}_D = R_{D0} + a_1 \cdot \bar{x}_{S2D}^{b_1} = R_{D0} + a_1 \cdot (K_{S2D} \cdot \bar{x}_{S2M})^{b_1}. \quad (3.23)$$

Eine Linearisierung des Messaufnehmers, d. h. eine Rückberechnung des Messwertes des Methanolstoffmengenanteils im Medium \tilde{x}_{S2M} ,

$$\tilde{x}_{S2M}(t) = \frac{\left(\frac{R_D(t) - R_{D0}}{a_1}\right)^{\frac{1}{b_1}}}{K_{S2D}} = \sqrt[b_1]{\frac{R_D(t) - R_{SD0}}{a_1 \cdot K_{S2D}^{b_1}}}, \quad (3.24)$$

erfolgt durch den Messverstärker der FMC-Grundeinheit mit Hilfe der aufgenommenen Kalibrierungsfunktion, um den auf dem LCD-Display des Gerätes angezeigten Messwert \tilde{x}_{S2M} in eine Spannung U_{S2M} umzuwandeln. Die Spannung U_{S2M} verhält sich somit proportional zum Messwert \tilde{x}_{S2M} und zur Messgröße x_{S2M} .

Der FMC wurde zur externen Weiterverarbeitung seiner Messwerte an einen analogen Spannungseingang der Reaktorbedieneinheit DCU angeschlossen. Die DCU berechnet als Messwert die Methanolkonzentration in der Medienphase \tilde{c}_{S2M} ,

$$\tilde{c}_{S2M}(t) = c_{S2Mlow} + \frac{c_{S2Mhigh} - c_{S2Mlow}}{U_{S2Mhigh} - U_{S2Mlow}} \cdot U_{S2M}(t), \quad (3.25)$$

mit

\tilde{c}_{S2M} := Messwert der Methanolkonzentration in der Medienphase [gl⁻¹]

$c_{S2Mhigh}$:= obere Messgrenze der Methanolkonzentration in der Medienphase [gl⁻¹]

c_{S2Mlow} := untere Messgrenze der Methanolkonzentration in der Medienphase [gl⁻¹]

U_{S2M} := Spannungssignal des Methanoldessgerätes [V]

$U_{S2Mhigh}$:= maximales Spannungssignal des Methanoldessgerätes [V]

U_{S2Mlow} := minimales Spannungssignal des Methanoldessgerätes [V]

aus dem Spannungseingang U_{S2M} . Dieser Wert wird von der DCU an das Prozessleitsystem mit der SCADA-Software MFCS/win übertragen.

In Bild 3.8 sind die Kennlinien der Signalverarbeitung bei der Methanoldessung dargestellt.

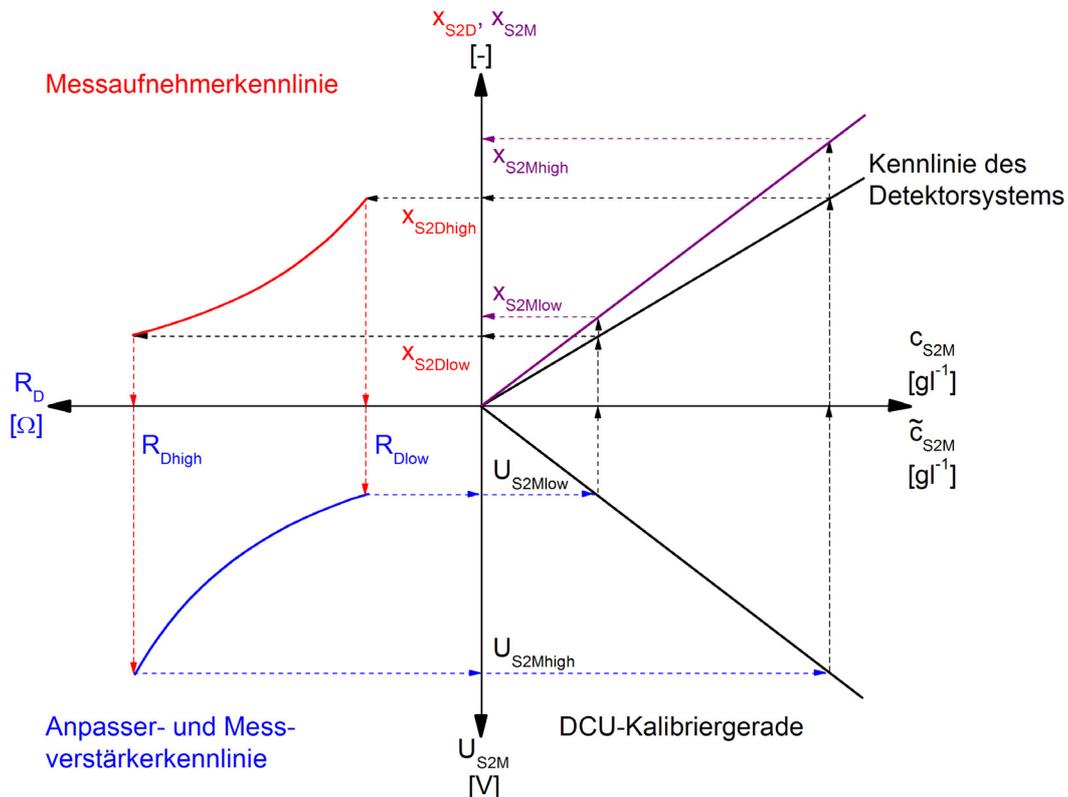


Bild 3.8: Zusammenspiel der Kennlinien in einer Methanoldesskette

In Bild 3.9 ist eine Vergleichsmessung der beiden Analysensysteme FID und FMC gezeigt. Neben der relativen Methanoldzufütterrate F_{R2rel} ist die UV-Absorption A_{P1M} des Zielproteins

I-TAC aufgetragen. Als Regelgröße des Methanol PI-Reglers diente während der anfänglichen Produktionsphase zunächst das Signal des FID bei einem Sollwert von 1 gl^{-1} . Für die nachfolgenden sequentiellen Produktionsphasen ab $t = 92.7 \text{ h}$ wurde die Regelung beim gleichen Sollwert mit dem FMC-Signal durchgeführt. Die Regelung verläuft mit beiden Analysensystemen sehr gut, jedoch zeigt sich in den entsprechenden Phasen eine starke positive bzw. negative Drift des jeweils redundanten Systems.

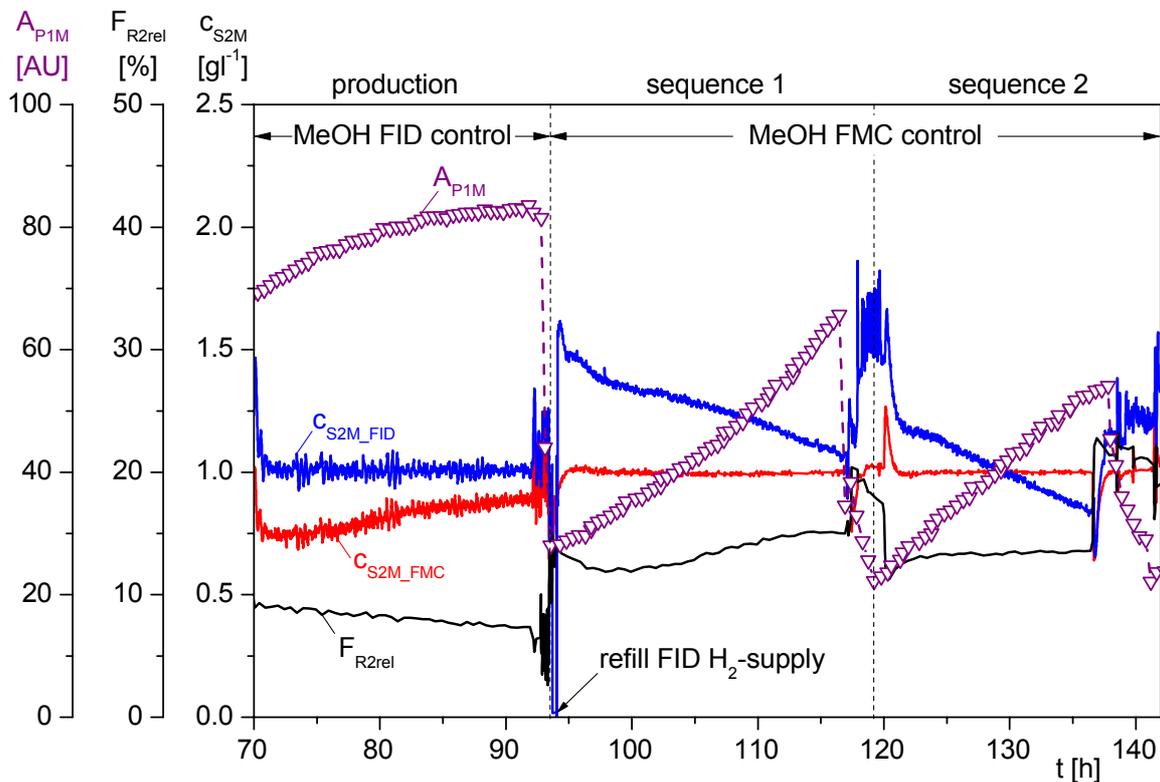


Bild 3.9: Vergleich von FID- und FMC-Methanolvormessung während einer Kultivierung von *P. pastoris* in Sequentiell Integrierter Prozessführung

3.2.4 HPLC-Analytik der Proteinkonzentration in der Flüssigphase

Die wichtigste Information eines Herstellungsprozesses rekombinanter Proteine ist der Verlauf der Produktbildung. Aufgrund unterschiedlicher physikalischer, chemischer und biologischer Eigenschaften der Proteine, intra- oder extrazellulärer Expression sowie Milieubedingungen muss für nahezu jedes Protein eine eigene Nachweismethode entwickelt werden. Die Chromatographie gehört dabei zu den am häufigsten angewandten analytischen Techniken.

Für die Aufreinigung mit *P. pastoris* sekretierter Proteine wurde zunächst eine Offline-FPLC-Methode (Hahn, 2001) und später für die Beobachtung der Zielproteinexpression eine Atline-HPLC-Analytik (Cornelissen, 2004) laborintern etabliert. In dieser Arbeit wurde ein Merck-Hitachi HPLC-System der Fa. VWR International, wie in Bild 3.10 gezeigt, direkt an den Kultivierungsprozess angebunden.

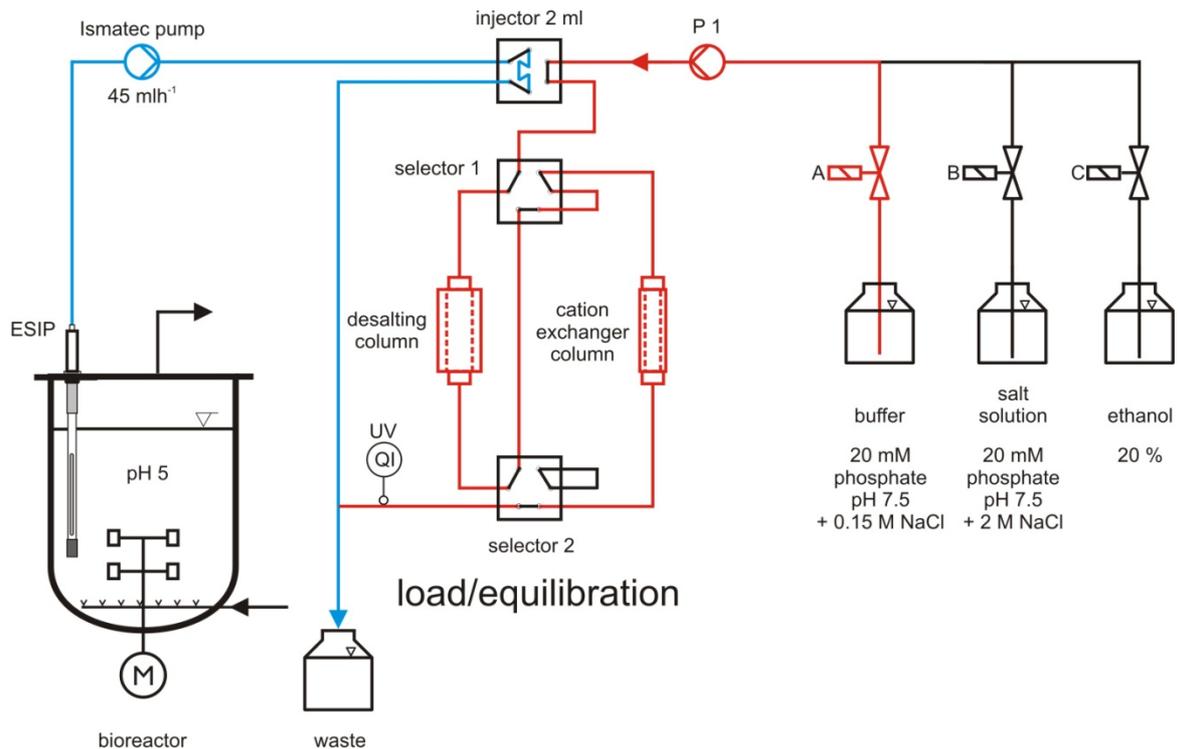


Bild 3.10: Anbindung des eingesetzten HPLC-Systems an den Bioreaktor

Der zellfreie Probenstrom wird mit einer Pumpe über eine sterilisierbare Filtrationssonde ESIP kontinuierlich aus dem Bioreaktor entnommen. Die zu analysierende Probe wird zu einem Injektionsventil gefördert und dort über eine 2 ml-Probenschleife geleitet. Die HPLC ist signaltechnisch mit dem Injektor verbunden und schaltet diesen zu Beginn eines Chromatogramms in die Position „inject“, so dass die Probe auf eine 5 ml HiTrap™ Desaltingsäule der Fa. Amersham Biosciences gegeben wird. Hier werden die Proteine von den Salzen durch Größenausschluss getrennt und vom Kultivierungs-pH 5 auf den Chromatographie-pH 7.5 umgepuffert.

Aufgrund des hohen isoelektrischen Punktes pI von 10.79 des herzustellenden Chemokins 1-3del I-TAC trägt es in einem Phosphatpuffersystem von pH 7.5, im Gegensatz zu fast allen anderen im Medium vorkommenden Proteinen, eine positive Nettoladung und bindet an einen Kationenaustauscher (Cole *et al.*, 1998; Fiedler, 1993). Die im vorherigen Trennungsschritt umgepufferte Probe wird daher auf eine 1 ml Kationenaustauschersäule HiTrap™ SP-FF der Fa. Amersham Biosciences gegeben. Das Chemokin bindet dort und wird anschließend mit einem Salzgradienten eluiert. Die Proteinabsorption wird mit einem UV-Detektor bei einer Wellenlänge von 215 nm photometrisch detektiert. Ethanol (20 %) dient zum Reinigen und Lagern der eingesetzten Säulen.

Am Ende eines Chromatogramms wird von der HPLC-Software automatisch ein Exportfile generiert, welches die Daten der Peakauswertung enthält. Eine in Visual Basic programmierte

Applikation liest die Daten aus dem Textfile aus und stellt sie dem Prozessleitsystem MFCS/win zur Verfügung.

Das Signal der Absorption wird über eine lineare Kennlinie in die Proteinkonzentration c_{P1M} ,

$$c_{P1M}(t) = K_{P1/A} \cdot A_{P1M}(t), \quad (3.26)$$

mit

$$K_{P1/A} = 3 \cdot 10^{-3} \text{ g l}^{-1} \text{ AU}^{-1},$$

umgerechnet. Der Proportionalitätsfaktor $K_{P1/A}$ wurde über einen ELISA² in Vorversuchen bestimmt.

Bild 3.11 zeigt die Beobachtung einer Produktionsphase sowie drei darauf folgender Sequenzen einer Integrierten Prozessführung durch Atline-Proteinmessung über HPLC. Nach der Induktion bei $t = 46.3$ h steigt die Produktkonzentration im Medium in etwa analog einer Verzögerung erster Ordnung bis auf 0.25 g l^{-1} bei einer ermittelten Biotrockenmasse c_{XL_CDW} von 108.0 g l^{-1} . Das HPLC-System liefert sowohl in der Produktionsphase als auch in den nachfolgenden Sequenzen über lange Zeit ein stabiles Messsignal.

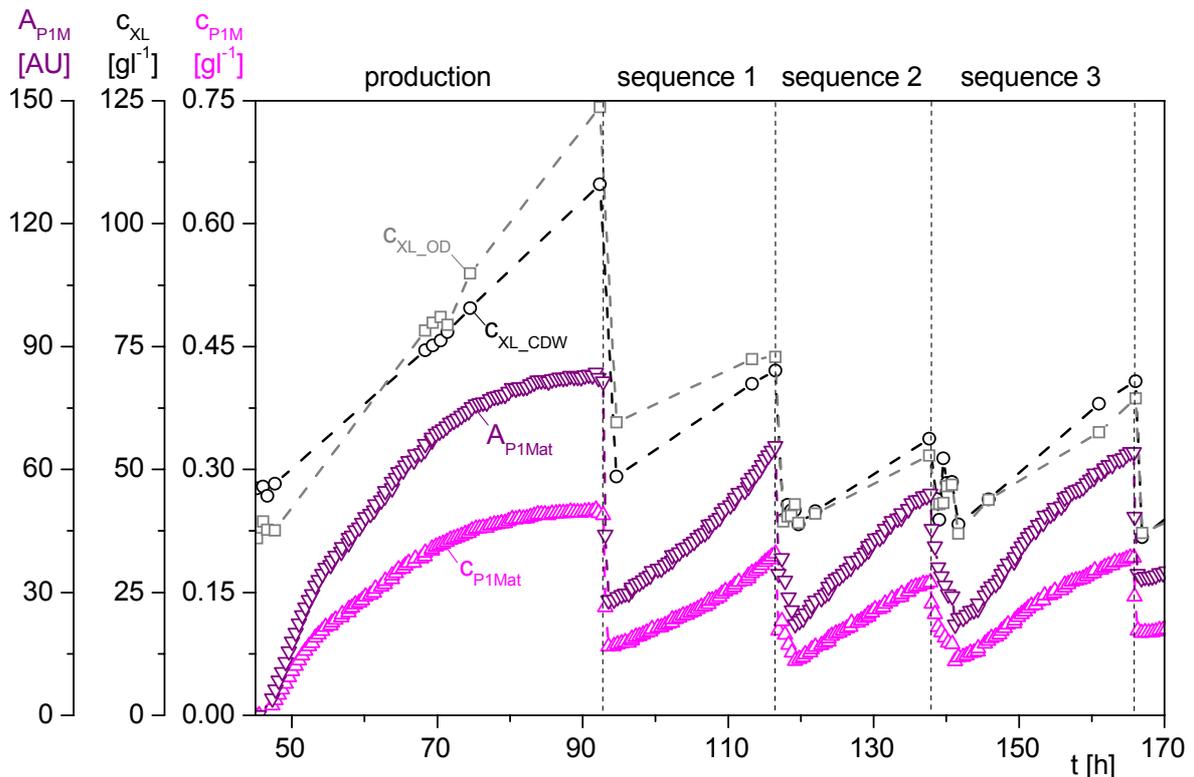


Bild 3.11: Beobachtung der Produktionsphase mit drei Sequenzen einer Integrierten Prozessführung durch Atline-Proteinmessung über HPLC

² ELISA - Enzyme Linked Immuno Sorbent Assay

3.2.5 Produktmessung mittels 2D-Fluoreszenzspektroskopie

Nicht-invasiv arbeitende optische Sensoren werden immer häufiger zum Monitoring von biotechnologischen Prozessen eingesetzt. Im Bereich der *in situ*-Fluoreszenzsensoren ermöglichen vor allem 2D-Prozessfluorometer eine simultane Detektion und in Verbindung mit chemometrischen Methoden eine Quantifizierung von Proteinen, Vitaminen und Koenzymen sowie Biomasse, Glukose und Metaboliten in Fermentationsbrühen (Hantelmann *et al.*, 2006; Rehbock *et al.*, 2008; Kaiser, 2009).

Der Expressionsverlauf des mit *E. coli* hergestellten fluoreszierenden Fusionsproteins ZsGreen-hEphB2-6His konnte durch die Anbindung des 2D-Fluoreszenzspektrometers BioView[®] der Fa. DELTA Light & Optics, Dänemark, beobachtet werden. Der Sensorkopf des Spektrometers wird über einen optischen Adapter mit Quarzglasfenster nicht produktberührend in den Bioreaktor eingebracht. Die Montage erfolgt dabei über einen 25 mm Sicherheitsstutzen im unteren Stutzenkranz (Bild 3.12a).

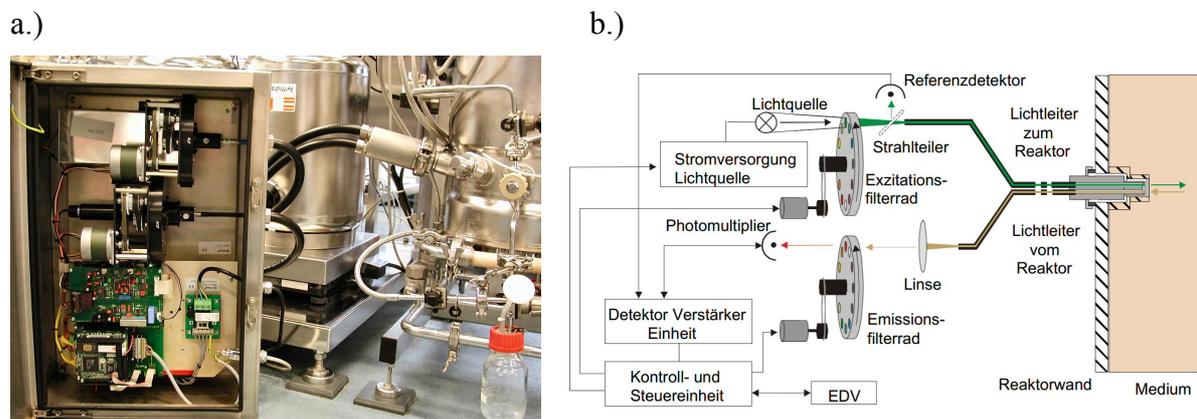
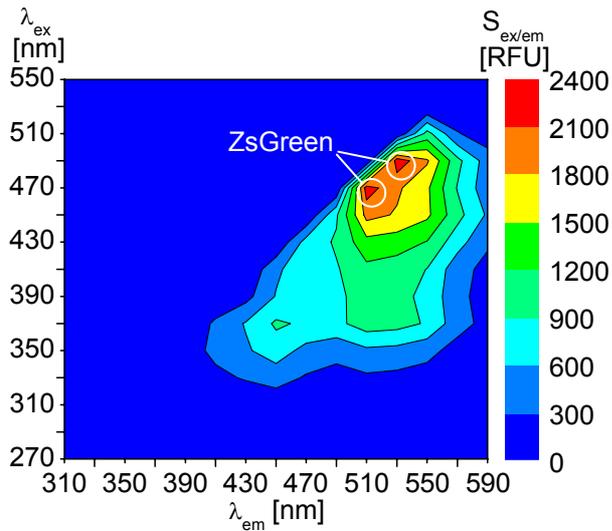


Bild 3.12: Einsatz des BioView[®] 2D-Fluoreszenzspektrometers am BIOSTAT[®] C-30

Die Messungen des Fluoreszenzlichts erfolgen nach dem „Open-end“-Prinzip im Winkel von 180° zum Anregungslicht. Durch Anregung der Oberfläche vor dem Sensorkopf wirken sich Veränderungen in der Eindringtiefe des Anregungslichts und der „Inner-Filter“-Effekt weniger stark aus (Stärk, 2002). Der eigentliche Sensor mit der Lichtquelle, Photomultiplier, Filterrädern für die Exzitation und Emission sowie der zugehörigen Mechanik und Elektronik, sind in einer geschützten Edelstahlbox untergebracht (Bild 3.12b; Lindemann, 1998).

Die 2D-Fluoreszenzspektren werden ca. im Minutentakt aufgenommen, indem die Extinktions- und Emissionswellenlängen mit Farbfiltern im Abstand von 20 nm durchgescannt werden. Das Anregungslicht wird von einer gepulsten Xenon-Blitzlichtlampe erzeugt. Das Spektrometer hat einen Anregungsbereich von 270 - 550 nm und einen Emissionsbereich von 310 - 590 nm. Die daraus resultierenden 2D-Fluoreszenzspektren bestehen aus insgesamt 150 Filterkombinationen.



Das in Bild 3.13 exemplarisch dargestellte Spektrum zeigt die Produktionsphase einer Kultivierung von *E. coli* MSD 5248. Das Exzitationsmaximum von ZsGreen liegt bei 496 nm, das Emissionsmaximum bei 506 nm. Die zeitliche Änderung der Fluoreszenzintensität kann aufgrund der Nähe der Spektralmaxima sowie der fixen 20 nm Bandbreite faktisch nur bei der Wellenlängenkombination $\lambda_{ex/em}$ von 470/510 nm bzw. 490/530 nm erfasst werden.

Bild 3.13: Fluoreszenzspektrum einer *E. coli* Kultivierung

Bild 3.14 zeigt den Verlauf der Produktionsphase mit unterschiedlich eingestellten Wachstumsraten während der Kultivierung von *E. coli* MSD 5248. Neben der Zellmasse m_{XL} sind die beiden BioView[®] Inline-Signale $S_{47/51}$ und $S_{49/53}$ aufgetragen. Die Induktion mit IPTG bei $t = 17.2$ h bewirkt einen exponentiellen Anstieg der Fluoreszenzsignale, der jedoch nach Änderung des μ -Sollwerts bei $t = 21.7$ h in einen linearen Verlauf übergeht.

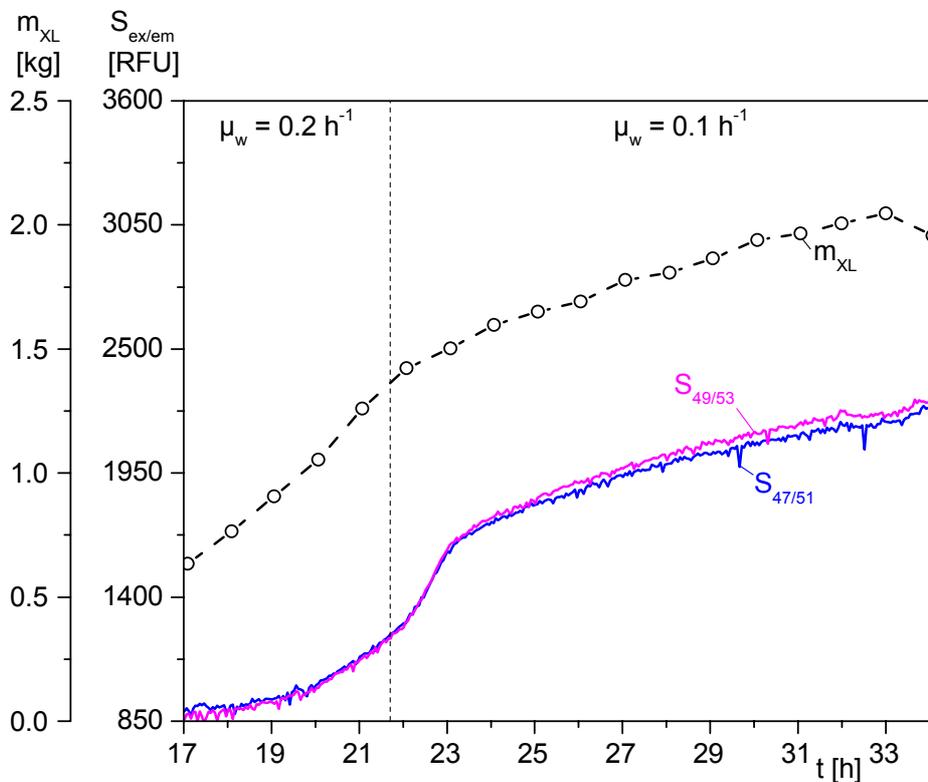


Bild 3.14: Beobachtung der Produktionsphase durch 2D-Fluoreszenzspektroskopie

3.3 Erweitertes Prozessmonitoring

3.3.1 Einsatz von Software-Sensorik an biotechnologischen Prozessen

Trotz der Entwicklung und Bereitstellung innovativer Online-Sensorik, sind viele prozessrelevante biotechnologische Größen *in vitro* nur indirekt messbar und *in vivo* nur mit erheblichem Aufwand bzw. gar nicht (Sonnleitner, 2000). Eine bedeutende Technologie auf dem Gebiet der indirekten Messung ist der Einsatz sog. Software-Sensoren (syn. Softsensoren) zur Prozessbeobachtung und darauf aufbauender Automatisierung (Luttmann *et al.*, 1990; Becker *et al.*, 2007; Becker and Krause, 2010).

Softsensoren sind virtuelle Sensoren, die aus einer primären physikalischen/(bio-)chemischen Messung über mathematische Ansätze/Modelle eine gewünschte sekundäre Variable bzw. einen Parameter berechnen (Chéruy, 1997). Softsensoren werden in zwei unterschiedliche Gruppen unterteilt. Die erste Gruppe erzeugt, ohne explizites Prozesswissen zu nutzen, mit Hilfe datengetriebener oder Blackbox-Modelle eine gewünschte sekundäre Variable. Die zweite Gruppe, auch Beobachter genannt, integriert direktes a-priori Prozesswissen zur Schätzung der sekundären Zielgröße über Whitebox-Modelle (James *et al.*, 2002).

Die Wahl des Modellierungsansatzes steht somit im Zentrum bei der Konstruktion eines Softsensors. Modelle der ersten Gruppe vergleichen und korrelieren Ein- und Ausgangsgrößen miteinander, so dass Fuzzy Logic, Evolutionäre Algorithmen, Neuronale Netze und multivariate Regressionsansätze der klassischen Chemometrie die Grundlage bilden. Basis für die zweite Gruppe sind dagegen Zustandsdifferentialgleichungen, die das dynamische Verhalten des Prozesses beschreiben und in denen typischerweise Transportvorgänge und Reaktionskinetiken berücksichtigt werden. Da diese Differentialgleichungen nicht analytisch gelöst werden können, werden sie online numerisch integriert (Rehbock *et al.*, 2008).

In der Literatur sind verschiedene Softsensorsysteme zu finden, die sowohl im akademischen als auch industriellen Bereich erfolgreich zur Prozessüberwachung bzw. Steuerung eingesetzt wurden (Luttmann *et al.*, 1990; Farza and Chéruy, 1994; Simutis *et al.* 1997; Schmidt, 2005; Gnoth *et al.*, 2008). Hierbei wurden vor allem mathematisch-naturwissenschaftliche Programmpakete wie Mathematica[®] (Wolfram Research) oder MATLAB[®] (The MathWorks) verwendet.

3.3.2 Modellbasiertes Monitoring mit MATLAB[®] Applikationen

In der vorliegenden Arbeit wurde das modulare Programm MAC4MFCS³ entwickelt, das die Beobachtung, Führung und Auswertung biotechnologischer Prozesse in bis zu zehn Bioreaktoren erleichtern soll. Es basiert auf der Entwicklungsumgebung MATLAB[®] (Vers. 6.5) und wurde spezifisch für *E. coli* Produktionsprozesse mit dem an der HAW realisierten verfahrenstechnischen Prozess-Setup erstellt (Ellert *et al.*, 2010).

³ MAC4MFCS - Monitoring And Control for MFCS/win

MATLAB[®] ist ein einheitliches interaktives numerisches System zur Berechnung, Simulation, Visualisierung oder Programmierung mathematischer Ausdrücke. Das System kann durch proprietäre Toolboxes, die für einen bestimmten Anwendungsbereich spezialisiert sind, einfach erweitert werden. In dieser Arbeit wurden die Control System Toolbox (Vers. 5.2), die Optimization Toolbox (Vers. 2.2) und die Symbolic Math Toolbox (Vers. 2.1) genutzt.

MATLAB[®] ist über eine OPC Server/Client-Schnittstelle an das Prozessleitsystem MFCS/win angebunden und mit den vorliegenden klassischen Online-Daten und der vorgestellten erweiterten Inline-Messtechnik und Atline-Analysentechnik verknüpft.

OPC ist eine von der OPC-Foundation definierte, standardisierte Software-Schnittstelle für den herstellernerneutrale Daten- und Informationsaustausch von Softwarekomponenten in der Prozesssteuerung und Automatisierungstechnik (Lange, 2005).

Der MFCS/win OPC-Server (MFCSOPCS, Vers. 3.0) wird entweder über COM oder, wie in Bild 3.15 dargestellt, bei vernetzten Rechnern über DCOM gestartet, sobald der OPC-Client (Vers. 3.0) der Fa. IPCOS, Belgien, eine Verbindung initialisiert. COM/DCOM⁴ ist ein von Microsoft entwickeltes objektorientiertes RPC⁵-System, das den Fernaufruf von Funktionen ermöglicht. Der Einsatz der OPC-Technologie beschränkt sich daher auf Windows[®] Betriebssysteme.

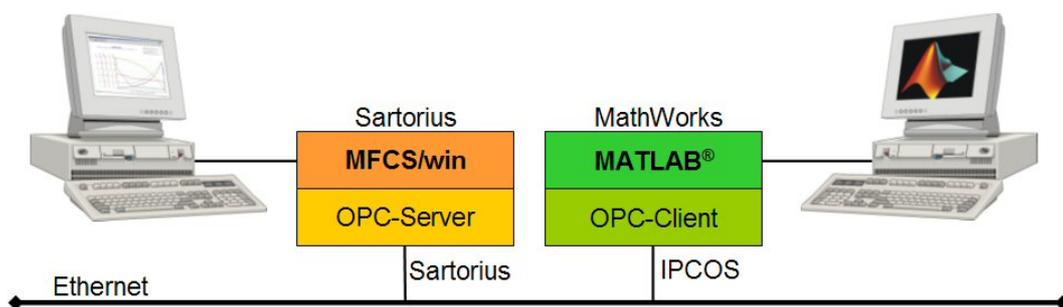


Bild 3.15: Prinzipieller Datentransfer zwischen MATLAB[®] und MFCS/win

Die obere Hälfte der Benutzeroberfläche (GUI⁶) der MATLAB[®] Applikation MAC4MFCS in Bild 3.16 zeigt eine Übersicht der physikalischen Prozesseinheiten in tabellarischer Form. Jede Zeile visualisiert die in MFCS/win definierte Batch ID, den aktuellen Monitoring- und Control-Status einer Prozesseinheit sowie die für die Automatisierung der Kultivierung geladene Methode. Eine Methode enthält spezifische Informationen über das verfahrenstechnische Prozess-Setup, die biologischen Eigenschaften des kultivierten Mikroorganismus und die Bedingungen zum Start des Prozesses.

⁴ (D)COM - (Distributed) Component Object Model

⁵ RPC - Remote Procedure Calls

⁶ GUI - Graphical User Interface

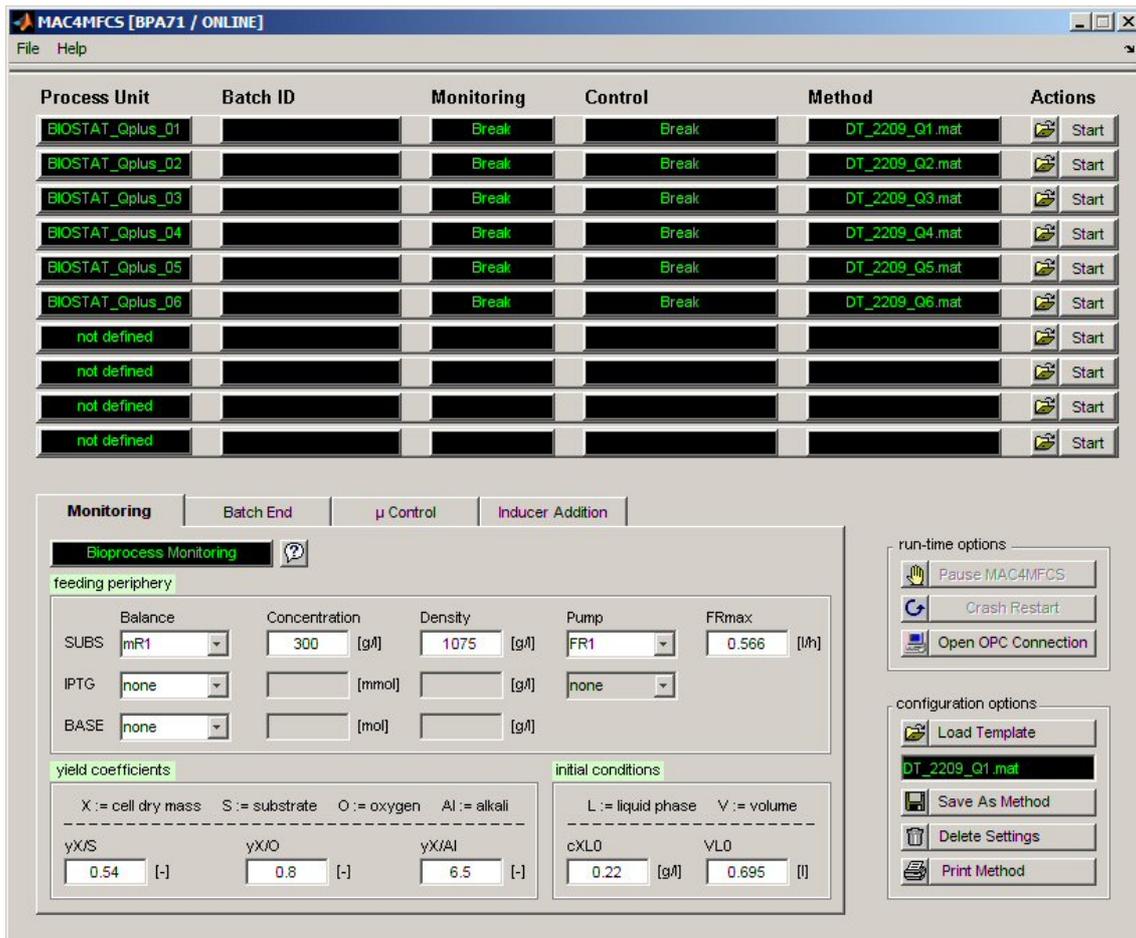


Bild 3.16: Benutzeroberfläche der MATLAB[®] Applikation MAC4MFCS

Für die Konfiguration einer Methode stehen dem Benutzer die vier Registerkarten Monitoring, Batch End, μ -Control und Inducer Addition in der unteren Hälfte des GUI zur Verfügung, die gleichzeitig den modularen Aufbau des Programms und dessen Einsatzmöglichkeiten widerspiegeln.

Mit dem *Monitoring-Modul* ist neben der klassischen Prozessbeobachtung über Abgasanalyse, die es dem Benutzer ermöglicht die Respirationsaktivität sowie Stofftransferbedingungen zu beobachten, eine Online-Bestimmung der zellspezifischen Wachstumsrate und Zelldichte implementiert.

Ein kritischer Schritt im Verlauf einer Kultivierung ist der Übergang von Batch- auf Fed Batch-Betrieb, da eine Fehlfunktion dieses Automatisierungsschrittes unter Umständen zu einer lang anhaltenden Substratlimitierung führen würde, die vom Organismus nicht dauerhaft toleriert werden kann. Daher ist eine in sich konsistente aber überaus effektive selbsttätige Weiterschaltung von der einen zur anderen Wachstumsphase in MAC4MFCS in Form einer automatischen Batchende-Erkennung mit dem *Batch End-Modul* realisiert. Der implementierte Algorithmus berücksichtigt den Anstieg des Sauerstoffpartialdruckes pO_2 und den davon abhängigen Abfall der Rührerdrehzahl N_{St} bei aktiver pO_2 -Regelung zum Batchende.

Die Auswahl einer geeigneten Kultivierungsstrategie ist ein entscheidender Gesichtspunkt im biotechnologischen Upstream-Prozess. Unter der Vielzahl der entwickelten Techniken, hat sich die Kultivierung zu hohen Zelldichten im substratlimitierenden Fed-Batch als robustes und effizientes Verfahren erwiesen (Lee, 1996; Riesenberg and Guthke, 1999). Mit dem μ -Control Modul hat der Benutzer die Möglichkeit, die zellspezifische Wachstumsrate μ auf einem definierbaren Sollwert μ_w zu steuern oder auch über das direkte μ -stat Prinzip zu regeln (Luttmann *et al.*, 1994).

Die Proteinexpression wird bei *E. coli* durch die Zugabe von IPTG in die Kulturbrühe initiiert. Das *Inducer Addition-Modul* wurde speziell für die ereignisgesteuerte pulsweise Zugabe von IPTG entwickelt. Der Benutzer kann die automatische vorwählbare Induktorzugabe z. B. bei Erreichen einer gewissen Zelldichte oder einer bestimmten Feeding-Zeit realisieren.

Neben MAC4MFCS wurde in dieser Arbeit zur Online-Beobachtung und Auswertung der biotechnologischen Herstellungsprozesse mit *P. pastoris* ein ebenfalls auf der Entwicklungsumgebung von MATLAB[®] basierendes Monitoring-Programm MONI eingesetzt. Die grafische Benutzeroberfläche des Programms ist in Bild 3.17 gezeigt.

Da das System flexibel an verschiedenen Prozessen mit unterschiedlichen Konfigurationen einsetzbar ist, geschieht die Zuordnung der Variablen, die vom OPC-Server empfangen werden sollen, sowie die Definition zeitinvarianter Prozess- und Wachstumsparameter sowie Anfangsbedingungen in den sechs Registerkarten auf der rechten Seite des GUI.

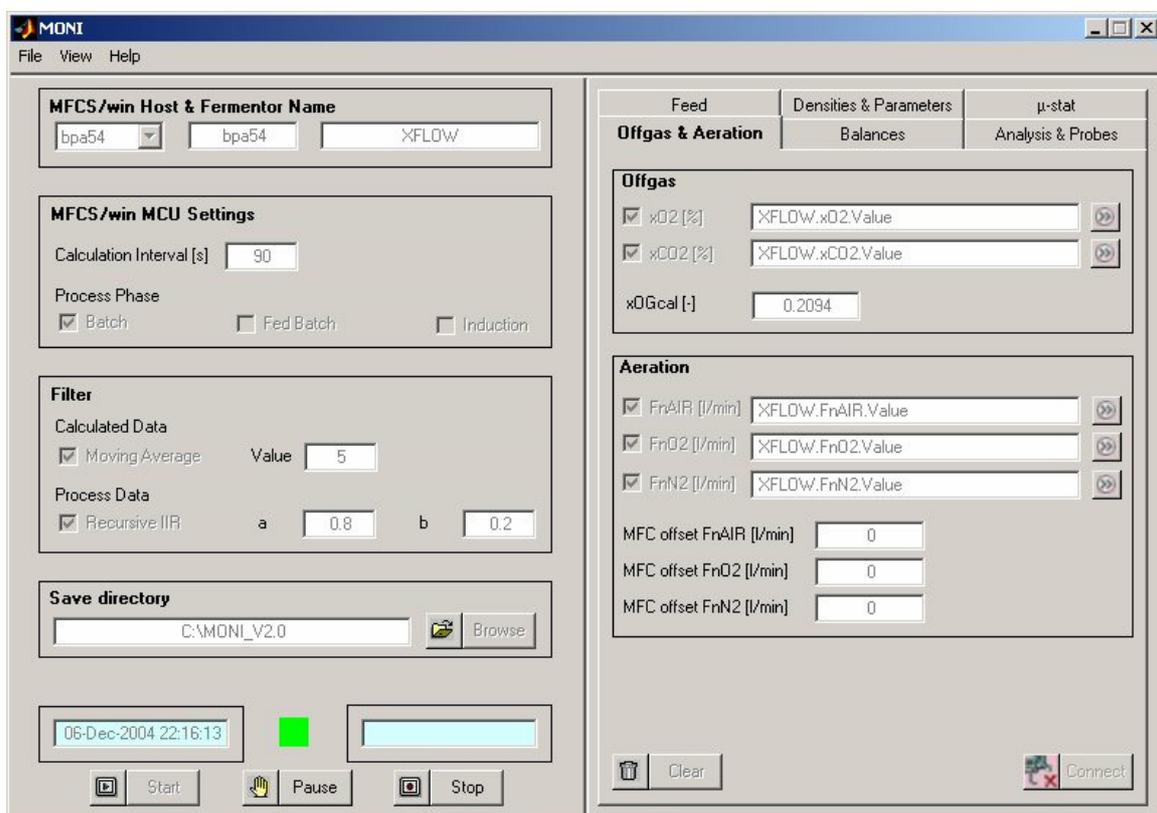


Bild 3.17: Benutzeroberfläche des Monitoring-Programms MONI

In der linken Seite des GUI wird einerseits der MFCS/win PC ausgewählt, zu dem eine Verbindung aufgebaut werden soll, andererseits wird das Zeitintervall zur Ausführung von Berechnungen sowie die aktuelle Prozessphase vorgegeben. Die Qualität der Berechnungen hängt von der Güte des mathematischen Modells sowie der Qualität des Messsignals und dem eingestellten Berechnungsintervall ab. Um eventuelles Rauschen der Messsignale zu minimieren bzw. berechnete Daten zu glätten, wurden verschiedene Filter implementiert.

Die mathematische Modellierung von Bioreaktionsprozessen kann bei pseudohomogener Betrachtung mit Einphasen-Modellen oder heterogener Betrachtung mit Mehrphasen-Modellen erfolgen (Moser, 1994). Einphasen-Modelle aber auch das klassische Zweiphasen-Modell von Gas- und Flüssigphase sind zur Beschreibung von Hochzelldichtekultivierungen ungeeignet, da einerseits bei hohen Zelldichten die feuchte Biomasse einen bedeutenden Anteil der Flüssigphase L einnimmt, andererseits die messtechnische Erfassung relevanter Substrate und Produkte in der zellfreien Medienphase M stattfindet.

Peuker (2006) und Kaiser (2009) berichten, dass bei *E. coli* Kultivierungen mit Zelldichten von 60 bis 100 gl^{-1} die feuchte Biomasse zum Ende einer Kultivierung 30 bis 55 % der Flüssigphase L einnimmt. Daher wird zur mathematischen Beschreibung der durchgeführten Kultivierungen mit *E. coli* und *P. pastoris* das Dreiphasen-Modell mit Gas-, Medien- und Biofeuchtphase verwendet.

Die MATLAB[®] Applikationen MAC4MFCS und MONI beinhalten neben den Algorithmen zur übergeordneten Steuerung der Prozesse mit *E. coli* bzw. *P. pastoris* ein mathematisches Modell für das Online-Monitoring. Eine ausführliche Beschreibung der implementierten Gleichungen ist in Cornelissen (2004) und Peuker (2006) zu finden.

4 Herstellung von pharmazeutischen Proteinen im dreiphasigen Produktionsablauf

4.1 Der klassische dreiphasige Produktionsablauf

Eine effiziente Produktion therapeutischer Proteine wird durch viele Faktoren beeinflusst, wie z. B. die Konstruktion eines geeigneten Expressionssystems, die Beachtung von Sicherheits- bzw. Qualitätsaspekten (FDA-Zulassung) oder die Entwicklung eines kostengünstigen Produktionsverfahrens. Sowohl für *E. coli* als auch für *P. pastoris* hat sich das Fed Batch-Verfahren bewährt, bei dem die Zellen zu hohen Zelldichten gezüchtet werden können. Damit wird dem primären Anliegen einer hohen volumetrischen Produktivität bzw. Produktausbeute Rechnung getragen, wodurch besonders bei der Produktion unter GMP-Bedingungen Kosten- und Zeitvorteile resultieren (Shiloach and Fass, 2005, Voedisch *et al.*, 2005; Rinas, 2006).

Wie in Bild 4.1 exemplarisch an einer Kultivierung von *P. pastoris* gezeigt, kann das Herstellungsverfahren in die drei Hauptphasen Batch, Fed Batch und Produktion unterteilt werden. Im Batch wird *Pichia* auf Glycerol unter nicht limitierenden Bedingungen und im Fed Batch mit reduzierter spezifischer Wachstumsrate unter Substratlimitierung angezchtet. Glycerol wurde mit einer Anfangskonzentration c_{S1M0} von 30 g l^{-1} vorgelegt und konnte offline per HPLC-Messung bestimmt werden. Aufgrund der maximal möglichen Glycerolzufüterraterate F_{R1max} von 0.18 lh^{-1} erreichte die volumetrische Glyceroleintragsrate Q_{S1in} mit 4.42 g(lh)^{-1} bei $t = 39.3 \text{ h}$ ein Maximum.

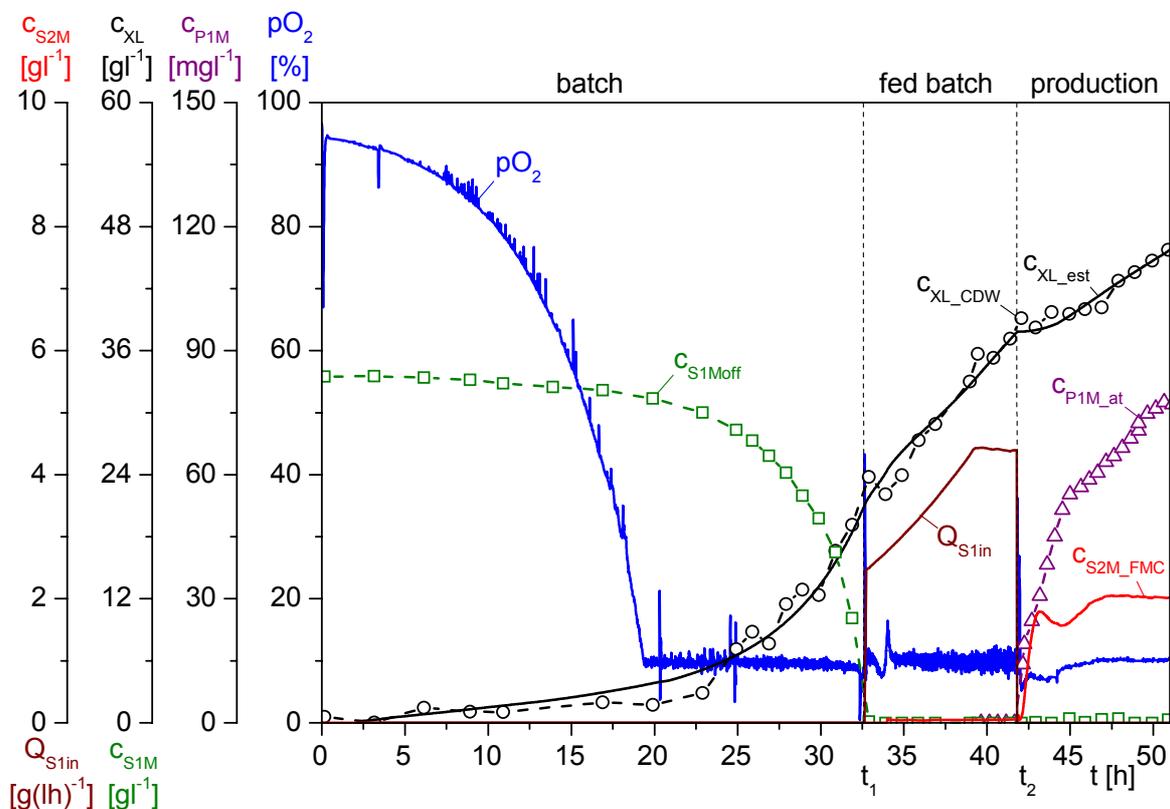


Bild 4.1: Vollautomatischer, dreiphasiger Prozessablauf zur Herstellung von 1-3del I-TAC

Durch Wechsel des Zufüttermediums auf den Induktor Methanol geht der Prozess in die Produktionsphase des Zielproteins über. Die Methanolkonzentration im Medium c_{S2M} wird bis zum Abbruch der Kultivierung auf einen Sollwert c_{S2Mw} von 2 g l^{-1} geregelt. Bei dieser Konzentration wird eine gute Proteinexpression mit einer hohen Produktbildungsrate erwartet.

Die UV-Absorption A_{P1M} des Zielproteins wurde atline per HPLC gemessen und im Prozessleitsystem MFCS/win nach Gl. 3.26 in eine Konzentration c_{P1Mat} umgerechnet. Die Messung der Methanolkonzentration c_{S2M} erfolgte inline über das FMC-System. Die Zelldichte c_{XLest} wurde mit dem MATLAB[®] Programm MONI über Abgasbilanzen online berechnet. Nach der Kultivierung wurde die Zelldichte c_{XL_CDW} über die Biotrockenmasse bestimmt.

4.2 Der Aufbau der Bioreaktoranlage BIOSTAT[®] C

Der in Bild 4.2 gezeigte BIOSTAT[®] C der Fa. B. Braun Biotech International wurde für die Kultivierung von *P. pastoris* eingesetzt. Der Reaktor kann mit einem Arbeitsvolumen von 7 - 30 l betrieben werden. Für die Durchmischung der Kulturbrühe sorgen drei höhenverstellbare 6-Blatt-Scheibenrührer sowie vier Strömungsbrecher. Das Rührersystem wird von einem Gleichstrom-Servomotor der Fa. Kollmorgen/Seidel Servo Drives mit einem Regelungsbe- reich der Drehzahl N_{St} von $10 - 800 \text{ min}^{-1}$ angetrieben. Die Wellendurchführung im Reaktor- deckel erfolgt über eine doppelte Gleitringdichtung.

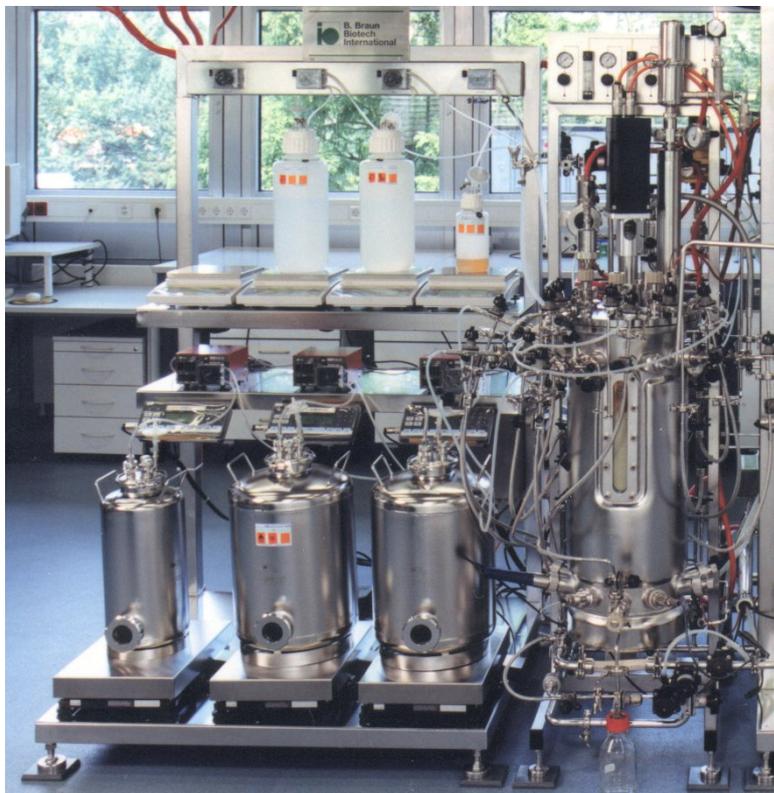


Bild 4.2: Bioreaktor BIOSTAT[®] C und Vorlagenstation

Im oberen Stutzenkranz wird über 25 mm Normstutzen Glycerol für den Fed Batch sowie Methanol für die Produktionsphase zugegeben. Der Anschluss für die Methanolzufütterung ist mit einem Steigrohr versehen. Der untere Stutzenkranz besteht aus vier radialen 25 mm Sicherheitsstutzen und zwei radialen 25 mm Normstutzen zur Aufnahme von Reaktorsonden, einem geraden 25 mm Sicherheitsstutzen für das Probenahmeventil und einem 12 mm Normstutzen für die Temperatursonde.

Neben den Anschlüssen für das CIP-System sind im Deckel die Stutzen für Lauge, Säure und Antischaummittelzugabe sowie eine Schaumsonde (Fa. Sartorius) in 19 mm Normstutzen eingelassen. Außerdem ist ein Abluftkühler zur Verminderung von Flüssigkeitsverlusten über die Abluft auf dem Deckel installiert. Der konduktive High Foam Alarmsensor (Fa. Sartorius) befindet sich zwischen dem Abluftkühler und dem *in-situ* sterilisierbaren Abluft-Membran-Sterilfilter. Dort wird auch der Druckaufnehmer (Fa. Philips) eingesetzt. Das Zuluft-Membran-Sterilfilter mit von der Fa. Sartorius patentiertem Adapter für die *In-situ*-Sterilisation ist ebenfalls auf dem Deckel befestigt.

Oberhalb des Bioreaktorkessels ist eine Gasmischstation mit Mass-Flow-Controllern (Fa. Bronkhorst HI-TEC, Niederlande) für Luft, Sauerstoff und Stickstoff installiert. Die Begasung erfolgt mit Luft (20.94 % Sauerstoff, 0.03 % Kohlenstoffdioxid, 79.03 % Stickstoff) über einen Begasungsring am Boden des Bioreaktors mit einer Luftbegasungsrate von bis zu 50 lmin^{-1} . Bei Bedarf kann zusätzlich mit reinem Sauerstoff (max. 20 lmin^{-1}) bzw. Stickstoff (max. 10 lmin^{-1}) begast werden.

Das Temperiersystem besteht aus einem elektrischen Heizaggregat (Fa. Sartorius) und einem Wärmetauscher (Fa. Swep International, Schweden) sowie einem Zirkulationssystem mit Umwälzpumpe (Fa. Grundfos), das an den Reaktordoppelmantel angeschlossen ist. Die Sterilisation des Kessels erfolgt ebenfalls über dieses System.

Der Bioreaktor ist auf drei DMS-Wägezellen des Typs PR 6207/51D1 (Fa. GWT) aufgehängt. Die Aufhängung an drei Wägezellen minimiert die Störeinflüsse von Pendelbewegungen und Verdrehungen. Es wird eine gleichmäßige Lastverteilung erreicht, über die eine Bilanzierung der Kultivierungssuspension möglich ist. Die Wägezellen sind als Vollbrücken geschaltet und haben eine jeweilige Nennlast von 50 kg.

Im oberen Regal der Vorlagenstation befinden sich NalgeneTM Vorratsflaschen für die Säurezugabe (H_3PO_4 , $C_{\text{ActofT1}} = 2 \text{ mol l}^{-1}$), für die Laugezugabe (NH_3 , $C_{\text{AltotT2}} = 6.67 \text{ mol l}^{-1}$), für Antischaummittel (Struktol[®] J673, Fa. Schill+Seilacher) und eine optionale Vorlage. Die Peristaltikpumpen (Typ 102 R, Fa. Watson-Marlow) für diese Vorlagen sind in den Ständerrahmen integriert. Die Zugabemengen werden über Waagen der Master^{pro} Serie des Typs LP12000S der Fa. Sartorius bilanziert.

Die Edelstahlgefäße für die Glycerol- und Methanolvorlage befinden sich im unteren Teil der Vorlageneinheit. Die Zufütterung der Substrate Glycerol und Methanol erfolgt über Einkanalpumpen des Typs 101 U/R der Fa. Watson-Marlow. Eine Bilanzierung der Zufütterung wird mit Waagen des Typs QC65EDE-S der Fa. Sartorius realisiert. Alle Behälter sind autoklavierbar.

4.3 Mess-, Steuer- und Regelungsaufgaben für eine Kultivierung mit *P. pastoris*

Während des Standby-Betriebs, der Inokulation des Bioreaktors und der drei Hauptphasen sind eine Reihe von Automatisierungsaufgaben zu lösen, um reproduzierbare Produktionsabläufe und somit vergleichbare Ergebnisse zu erhalten. Für die vollautomatische Umsetzung der Prozessführungsstrategie sind neben der pH-, Druck- und Temperaturregelung über DCU3-interne PI- bzw. PID-Regler die in Bild 4.3 aufgeführten Automatisierungselemente erforderlich. Auf dieser Grundstruktur und der in Kapitel 3.2 beschriebenen erweiterten Messtechnik basieren die entwickelten Prozessführungsstrategien.

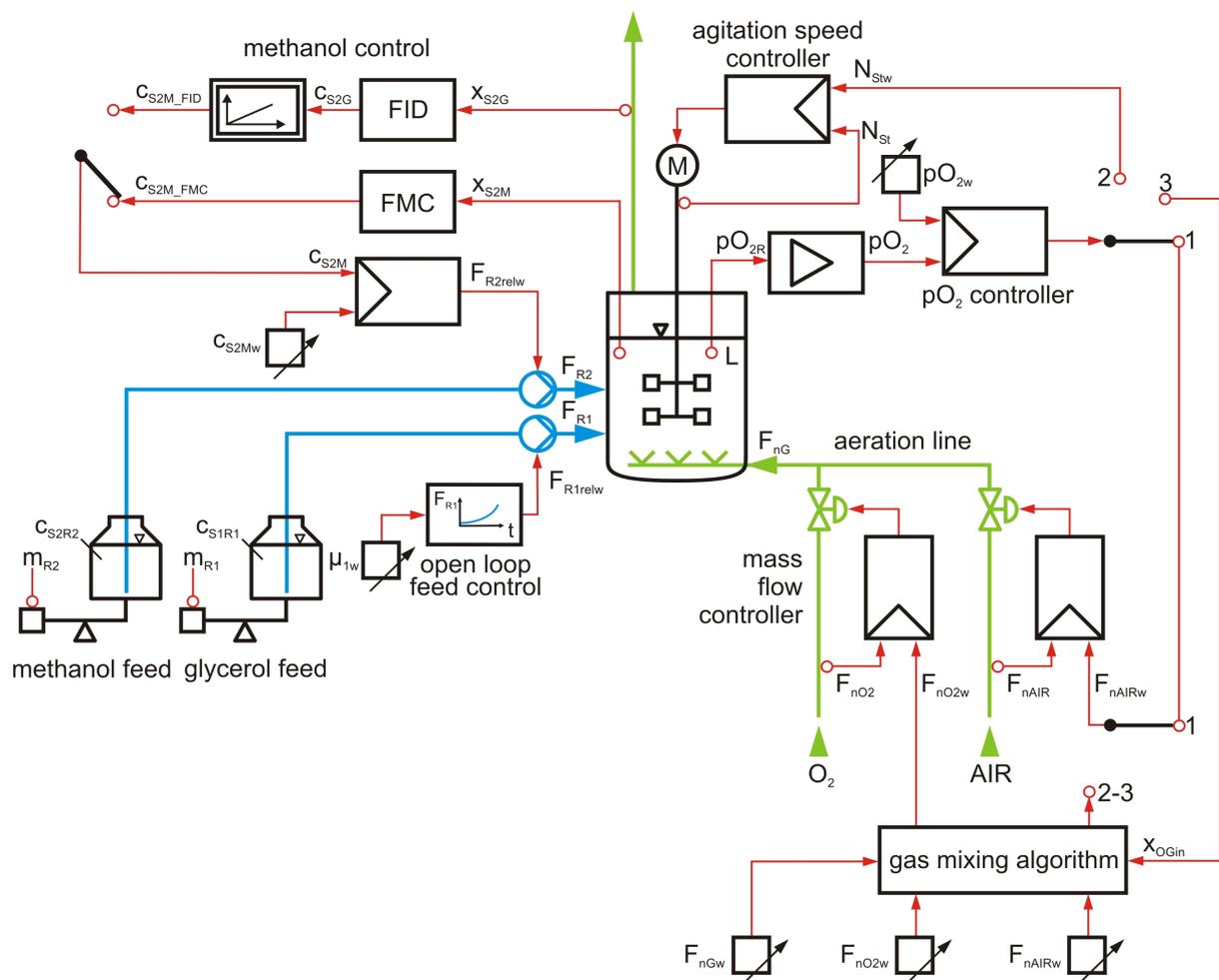


Bild 4.3: μ -Steuerung sowie Regelung des pO_2 und der Methanolkonzentration

Der Sauerstoffpartialdruck pO_2 wird nach Abfall auf seinen Sollwert im weiteren Verlauf der Kultivierung zunächst über die pO_2 /AIR-aeration-Methode, dann über die Regelungsart pO_2 /agitation und nachfolgend über pO_2 /gasmix auf 10 % geregelt. Zur Messung des pO_2 -Werts kommt eine Elektrode mit integriertem 22 k Ω NTC zur Temperaturkompensation der Fa. Mettler-Toledo zum Einsatz. Bei der pO_2 /AIR-aeration- und der pO_2 /agitation-Regelung dient die Stellgröße des pO_2 -Reglers als Sollwert für den jeweiligen Folgereger.

Die Realisierung der pO_2 /gasmix-Regelung, bei der die Stellgröße des pO_2 -Reglers den O_2 -Eingangsmolenbruch und somit eine AIR/ O_2 -Mischbegasung vorgibt, erfordert die Implementierung eines zusätzlichen Algorithmus zur Berechnung des Sollwertes F_{nAIRw} für die Luftbegasungsrate,

$$F_{nAIRw}(t) = \frac{1 - x_{OGin}(t)}{1 - x_{OAIR}} \cdot F_{nGw}(t), \text{ mit } x_{OGin}(t) > x_{OAIR}, \quad (4.1)$$

und des Sollwertes F_{nO2w} für die Sauerstoffbegasungsrate,

$$F_{nO2w}(t) = F_{nGw}(t) - F_{nAIRw}(t), \quad (4.2)$$

wobei der einstellbare Sollwert F_{nGw} für die Gesamtbegasungsrate der festgelegten maximalen Luftbegasungsrate $F_{nAIRmax}$ von 18 lmin^{-1} ($\hat{=} 1.5 \text{ vvm}$) entspricht. Die Mass Flow Controller arbeiteten dabei als Folgereger, die den geforderten Eingangsmolenbruch x_{OGin} erfüllen.

Das Ende der Batch-Phase wird automatisch über die Beobachtung des pO_2 -/ N_{St} -Verlaufes identifiziert. In der anschließenden Fed Batch-Phase wurde die exponentielle Glycerolzufütterung so gesteuert, dass die zellspezifische Wachstumsrate einen vorgebbaren Wert μ_{1w} annimmt (Gregory and Turner, 1993; Cornelissen, 2004). Die μ -Steuerung ist in Bild 4.3 als open loop feed control der stetig ansteuerbaren Zufütterpumpe 1 dargestellt.

Der Sollwert der relativen Glycerolzufüterraterate F_{R1relw} ,

$$F_{R1relw}(t) = F_{R1rel}(t_j) \cdot e^{\mu_{1w}(t_j)(t-t_j)}, \quad (4.3)$$

ist bei Vorgabe des Sollwertes μ_{1w} und der Zufüterraterate F_{R1rel} ,

$$F_{R1rel}(t_j) = \frac{\left[\mu_{1w}(t_j) + y_{X/S1gr} \cdot q_{S1/Xm} \right] \cdot c_{XLturb}(t_j) \cdot V_L(t_j)}{y_{X/S1gr} \cdot c_{S1R1}} \cdot \frac{100\%}{F_{R1max}}, \quad (4.4)$$

zum Start der Zufütterung bei $t_j = t_1$ oder späterem μ -Wechsel online berechenbar. Mit dem Ziel eine O_2 -Begasung während der Fed Batch-Phase zu vermeiden, wird ein kleines μ_{1w} von 0.1 h^{-1} gewählt.

Die unbekanntenen Reaktionsparameter $y_{X/S1gr}$ und $q_{S1/Xm}$ werden aus Vorversuchen ermittelt. Der zeitinvariante Zuwachsausbeutekoeffizient $y_{X/S1gr}$ wurde mit 0.75 gg^{-1} und die Substratmaintenancerate $q_{S1/Xm}$ mit 0.011 h^{-1} definiert (Cornelissen, 2004). Die maximale Zufüterraterate F_{R1max} und die Glycerolvorlagenkonzentration c_{S1R1} stellen feste Prozessparameter dar. Die über Trübungsmessung berechnete Zelldichte c_{XLturb} steht ebenso wie das gemessene Reaktionsvolumen V_L online zur Verfügung.

Sobald eine Zelldichte c_{XLturb} von 30 gl^{-1} am Ende der Fed Batch-Phase erreicht ist, wird die Zufütterung mit Glycerol gestoppt und die Zielproteinexpression der Hefezellen durch Zuga-

be von Methanol eingeleitet. Das Inline-Messsystem FMC-MINI 03 dient als Istwertgeber (c_{S2M_FMC}) für die Regelung der Methanolkonzentration im Medium. Alternativ kann der über eine redundante Online-FID-Messung ermittelte Prozesswert c_{S2M_FID} zur Regelung eingesetzt werden.

Im Anhang sind die Automatisierungsaufgaben während des dreiphasigen Produktionsablaufs tabellarisch zusammengefasst. Die automatische Batchende-Erkennung, die μ -Steuerung im Fed Batch und die sequentielle pO_2 -Regelung erfolgte mit dem Visual Basic[®] Programm MFCS-Tools. Das Monitoring nicht-direkt messbarer Größen erfolgte mit dem MATLAB[®]-Tool MONI. Die Initialisierung der Regelungen und des MATLAB[®]-Tools erfolgte über ein MFCS/win S88-Rezept.

4.4 Erforderliche Infrastruktur zur automatischen Prozessführung

4.4.1 Vor-Ort-Automatisierung mit dem Prozesscontroller DCU

Generell wird angestrebt, alle Regelungsaufgaben vor Ort durchzuführen, um die Wege der Datenübertragung möglichst kurz zu halten und eine eventuelle manuelle Prozessführung ohne Leitrechner zu ermöglichen. Der BIostat[®] C ist daher mit dem validierbaren, digitalen Mess- und Regelsystem DCU3 (Digital Control Unit 3) ausgestattet.

Das 32-Bit VMEbus mikroprozessorgesteuerte DCU3-System erlaubt neben der Grundsteuerung die Erfassung von Messwerten, die Kalibrierung von Sonden sowie die Regelung von Hauptprozessgrößen. Die Bedienung erfolgt über ein konfigurierbares grafisches Interface auf einem Split-Screen Display mit Berührungseingabe.



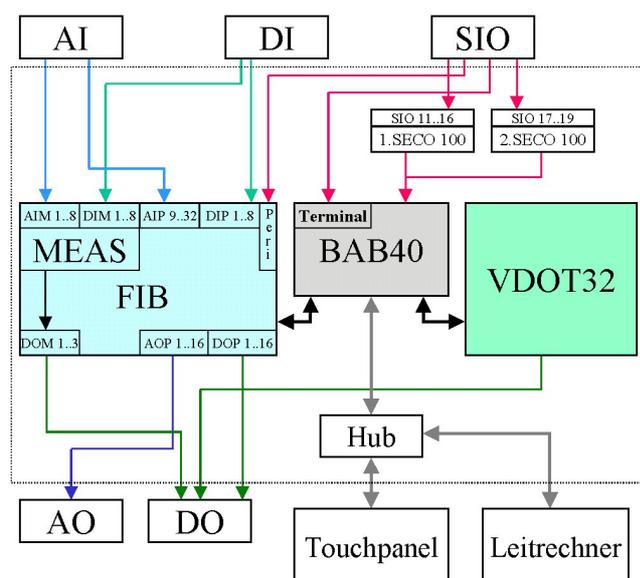
Bild 4.4: DCU3 Tower

Das DCU3-System kann in Remote-Control betrieben werden, so dass alle Mess- und Regelfunktionen von dem Prozessleitsystem MFCS/win beobachtet und gesteuert werden können.

Die DCU3 ermöglicht neben der seriellen Datenübertragung zu einem in das Automatisierungssystem eingebundenen Leitrechner im Vergleich zu ihren Vorgängerversionen auch einen Netzwerkbetrieb über eine Ethernetschnittstelle. Hierdurch sind höhere Datenübertragungsraten und geringere Abtastzeiten als im seriellen Betrieb möglich. Zusätzlich erlaubt die Netzwerkanbindung über FTP (File Transfer Protocol) einen schnellen Austausch der DCU-Konfiguration.

Bild 4.5 zeigt die interne Struktur der DCU3. Der gepunktete Rahmen deutet das DCU-Gehäuse an. Im Zentrum der DCU3 steht das Basic Automation Board BAB[®] 40 der Fa. Eltec Elektronik, das die Arbeit der beteiligten I/O-Karten koordiniert und das System entsprechend der Konfiguration steuert. Die Software (Version 4.62) und die Konfiguration (Version C_X501I) sind gemeinsam auf einer 16.1 MB PCMCIA-Card gespeichert, die über einen Steckplatz auf der BAB[®] 40 installiert ist. Die internen Programme laufen auf dem Echtzeit-Betriebssystem Microware OS-9[®].

Der FIB-Einschub (Fermentation Interface Board) ist eine Standardkarte, die bereits eine Grundausstattung von Ein- und Ausgängen zur Verfügung stellt. Jeweils die ersten acht analogen (AIM) und digitalen (DIM) Eingänge sind für Messverstärkerausgänge (MEAS) reserviert. Eine Erweiterungskarte des Typs VDOT-32 der Fa. Janz Automationssysteme wurde installiert, da die 16 digitalen Ausgänge (DOP) des FIB aufgrund der hohen Zahl ansteuerbarer Ventile nicht ausreichen. Dieses VMEbus I/O-Board erweitert die DCU um zusätzlich 32 digitale Kanäle, die als Ein- oder Ausgänge konfiguriert werden können.



Die Waagensignale werden seriell über zwei Schnittstellenkarten vom Typ SECO 100 der Fa. Eltec Elektronik eingelesen, die jeweils sechs konfigurierbare Kanäle bereitstellen. Die Verbindung zum Touchpanel erfolgt über ein Netzwerk, das mit Hilfe eines Hubs zum Leitreechner verzweigt ist. Das Bedienpanel hat seine eigene Rechereinheit, die auf einer Linux-Basis arbeitet.

Bild 4.5: DCU3 I/O Struktur

4.4.2 Datenerfassung mit dem SCADA-System MFCS/win

Das Prozessleitsystem MFCS/win ist eine batchorientierte, modular aufgebaute SCADA¹-Software der Fa. Sartorius Stedim Systems, die in der vorliegenden Arbeit für die Überwachung, Steuerung und Datenerfassung der durchgeführten Kultivierungen im BIOSTAT[®] C in der Version 2.1 und im BIOSTAT[®] Qplus in der Version 3.0 eingesetzt wurde.

MFCS/win wurde nach dem Client-Server-Modell konzipiert. Die Server-Module werden bei Systemstart automatisch durch den MFCS-Service (LOD) geladen und laufen fortan als Hin-

¹ SCADA - Supervisory Control And Data Acquisition

tergrunddienste auf dem MFCS/win Server. Die entweder auf dem gleichen PC oder einer Windows® Workstation im Vordergrund laufenden Client-Module können durch den Benutzer manuell über die in Bild 4.6 gezeigte MFCS/win Shell gestartet werden.

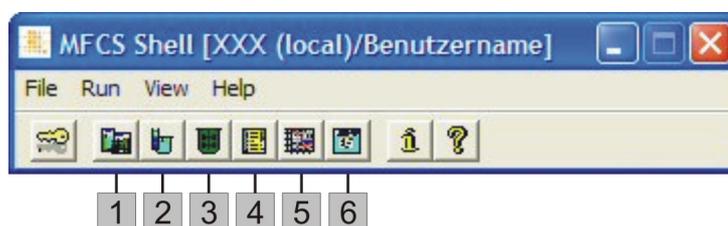


Bild 4.6: Die MFCS/win Shell mit sechs Client-Modulen

MFCS/win ist in der Lage bis zu 16 Kultivierungen gleichzeitig zu steuern, wobei die Messwerte der angeschlossenen Reaktoren mit dem Memory Resident Datapool (MRD) als zentrales Server-Modul erfasst und anderen Modulen zur Verfügung gestellt werden. Tabelle 4.1 zeigt die Beschreibung der zugehörigen Client-Module für die Konfiguration des Systems (CMA), die Verwaltung der Kultivierungen (BMA), die Durchführung und Überwachung der Kultivierungen (OPR, SDM, PLO) und die Dokumentation der Prozesse und Systemeinstellungen (BRP).

Tabelle 4.1: MFCS/win Client-Module und deren Funktion

Modul	Beschreibung
OPR (1)	Operator Service; Bedienung des Systems, Anzeige und Änderung von Variablen, Trends, Alarmen und Reglern
SDM (2)	Sample Data Management; Eingabe und Speicherung von Ergebnissen aus Off-line-Analysen
BMA (3)	Batch Management; Archivierung, Prüfung, Freigabe sowie anderweitige Verwaltung von Kultivierungen
BRP (4)	Batch Report; Erstellung von alphanumerischen Berichten zu erfolgten Kultivierungen
PLO (5)	Plot Management; Grafische Darstellung von bis zu sechzehn Variablen einer laufenden oder abgeschlossenen Kultivierung
CMA (6)	Configuration Management; Erstellung und Modifikation von diversen Systemobjekten wie Prozesseinheiten, Variablen, Displays und Benutzern

Die Entwicklung der Software erfolgt unter einem Qualitätssicherungssystem und entspricht den Anforderungen zur Validierung von computergestützten Systemen, die in der GAMP²-Richtlinie der International Society for Pharmaceutical Engineering (ISPE) festgehalten sind. Über ein optionales Modul unterstützt MFCS/win die technologischen Anforderungen der FDA-Richtlinie 21 CFR Part 11 für elektronische Aufzeichnungen und Unterschriften.

² GAMP - Good Automated Manufacturing Practice

4.4.3 Prozessautomatisierung mit der Visual Basic® Applikation MFCS-Tools

MFCS-Tools ist ein modular aufgebautes Visual Basic® Programm, das dem Anwender Werkzeuge zur Automation von Fermentationsprozessen zur Verfügung stellt (Schultz, 2001). Bei den in dieser Arbeit mit *P. pastoris* durchgeführten Kultivierungen wurden die Module *pO2-control*, *exponential feeding* (Abk.: *expon. feed*) und *batch end* eingesetzt. Die Module werden über zyklische Variablen aus MFCS/win heraus an- und abgeschaltet. Das konfigurierbare Zeitintervall für die Datenübertragung mit MFCS/win beträgt 15 s.

Das Modul *pO2-control* hat die Aufgabe, den pO_2 -Wert während des gesamten Prozesses auf einem konstanten Niveau zu halten. Hierbei sollen die bereits angesprochenen Regelungsstrategien durchlaufen werden. Das Modul wird über die MFCS-Variable *sw_pO2* gestartet [1] und gestoppt [0]. Bei zurückgehendem O_2 -Bedarf der Zellen ermöglicht die programmtechnische Umsetzung der pO_2 -Regelungsstrategie eine stoßfreie Umschaltung in die vorhergehende Kaskadierungsart.

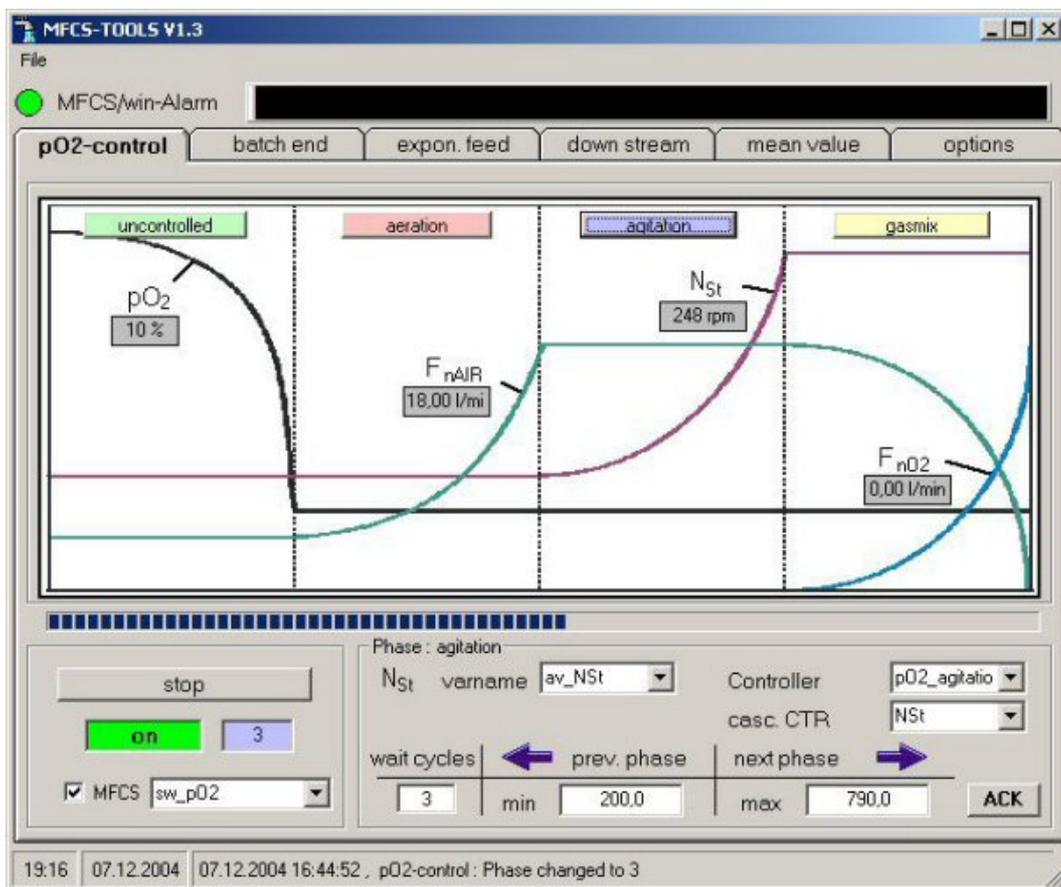


Bild 4.7: Das Visual Basic® Modul *pO2-control*

Das Modul *batch end* dient der vollautomatischen Erkennung des Batchendes und ermöglicht somit ein selbsttätiges Weiterschalten vom Batch zum Fed Batch. Das Modul wird über die zyklische MFCS-Variable *sw_BatchEnd* aktiviert [1] und deaktiviert [2].

Zur Batchende-Erkennung werden der pO_2 und die Drehzahl beobachtet. Nach Verbrauch des vorgelegten Glycerols fällt der Sauerstoffbedarf der Zellen plötzlich ab und der pO_2 steigt stark an. Der pO_2 -Regler versucht weiterhin den Sollwert zu halten und senkt die Rührerdrehzahl stark ab. Diese beiden Aktionen werden erkannt und der Prozess wird dann in den Glycerol-Fed Batch überführt.

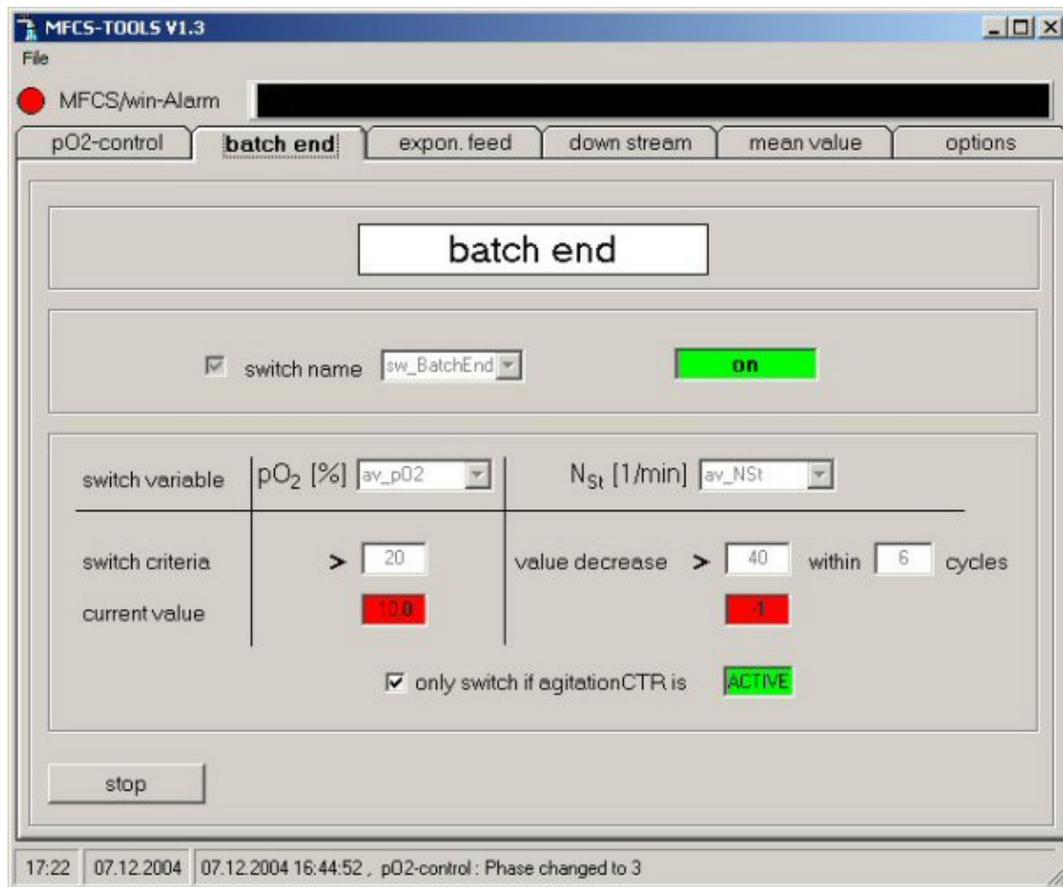


Bild 4.8: Das Visual Basic Modul *batch end*

Das in Bild 4.9 dargestellte Modul *exponential feed* basiert auf dem in Kapitel 4.3 vorgestellten Algorithmus zur Steuerung der Glycerolzufütterung in der Fed Batch-Phase. Das Modul wird über die Variable $sw_ExpoFeed$ angeschaltet [1] und abgeschaltet [0]. Die für den Formelapparat notwendigen Parameter werden sowohl manuell vom Benutzer beim Start des Programms eingegeben als auch automatisch aus MFCS/win eingelesen.

Um einen vollautomatischen Betrieb zu gewährleisten, muss für die Berechnung der relativen Glycerolzufütterrate F_{Rrel} zum Startzeitpunkt t_1 ein zeitdiskreter Wert für die Zelldichte vorliegen. In der Maske des Programms wird zu diesem Zweck die Hilfsvariable $c_{XLstart}$ ausgewählt, für die im S88-Rezept eine Online-Formel implementiert ist. In dieser Formel wird der aktuelle Wert der kontinuierlich neu berechneten Zelldichte c_{XLturb} der Hilfsvariablen $c_{XLstart}$ über das MFCS/win Calculation-Modul (MFCSCAL) zugewiesen.

Da MFCS/wins während eines laufenden Batches nur Online-Berechnungen auswertet die angeschaltet sind, wird die hinterlegte Formel im Rezept kurz vor dem Beginn der Zufütterung gestoppt, so dass die Hilfsvariable einen für den weiteren Kultivierungsverlauf zeitinvarianten Wert annimmt.

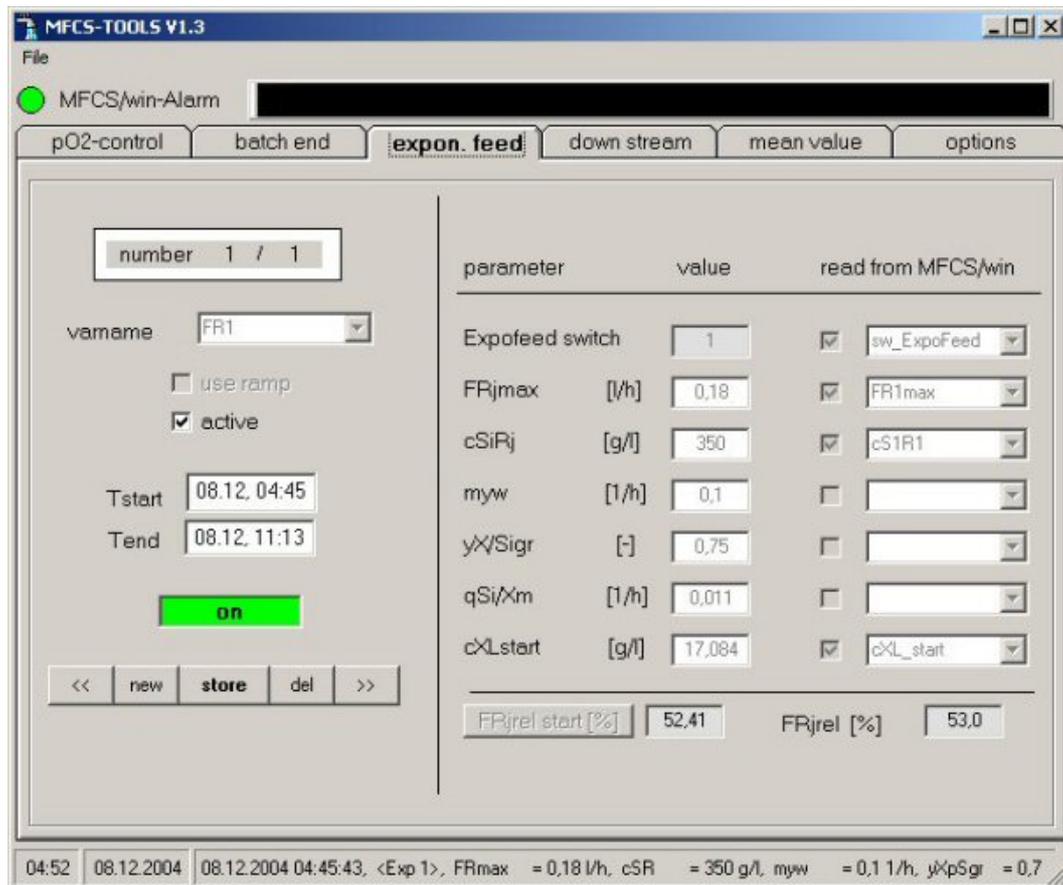


Bild 4.9: Modul *exponential feed*

4.4.4 Prozessführung mit Hilfe von Rezepten nach ANSI/ISA-88.01

Die Umsetzung der beschriebenen Prozessführung erfolgte einerseits über eine verfahrenstechnisch orientierte Prozedursteuerung in Anlehnung an ANSI³/ISA⁴-88.01 mit dem für die Prozessleitsoftware MFCS/win optional erhältlichen Batch Management Modul S88C, andererseits unter Verwendung der in Kapitel 4.4.3 vorgestellten VB-Applikation MFCS-Tools.

Der ANSI/ISA Standard 88.01-1995 „Batch Control Part 1: Models and Terminology“ wurde 1995 durch ein internationales Komitee mit dem Ziel entwickelt, verfahrenstechnische Automatisierungskonzepte für Batchprozesse weltweit zu vereinheitlichen. Bei der Teilintegration dieses Standards in MFCS/win wurden alle notwendigen Aspekte für Batch, Fed-Batch und kontinuierliche Prozessführungen übernommen.

³ ANSI - American National Standards Institute

⁴ ISA - The International Society of Automation

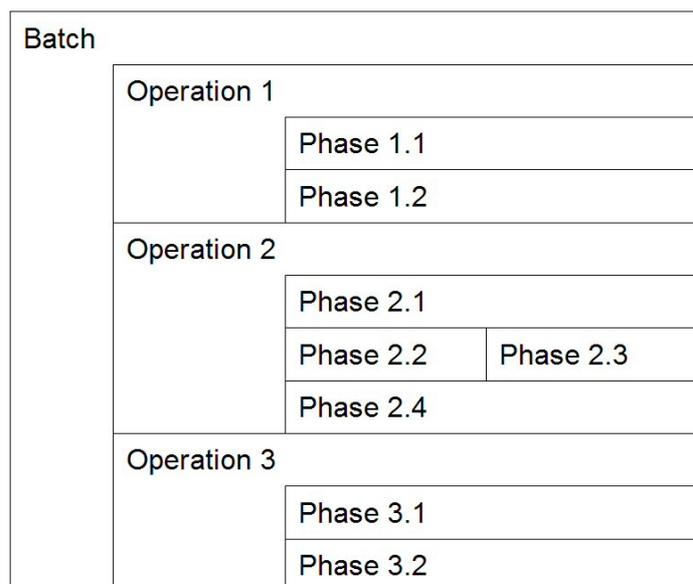


Bild 4.10 zeigt das ISA-Modell der Prozedursteuerung, wonach eine Teilprozedur, die in MFCS/win „Batch“ genannt wird, aus einer hierarchisch geordneten Menge von Prozedurelementen, sog. Operationen und Phasen, besteht. Diese werden in einer Bibliothek gruppiert, durch einen einmaligen Namen identifiziert und mit dem MFCS/win Rezepteditor in einem Ablaufdiagramm (SFC⁵) untereinander verknüpft.

Bild 4.10: Modell der Prozedursteuerung in MFCS/win

Ein Grundrezept kann in MFCS/win aus mehreren Operationen (max. 30) und jede Operation aus mehreren Phasen bestehen. Operationen können nur sequentiell, Phasen dagegen sequentiell (max. 20) oder parallel (max. 10) angeordnet werden. In sequentieller Anordnung sind die Phasen über eine logische Bedingung, die sog. Transition, miteinander verbunden.

Die Transition kann wahr [1] oder falsch [0] sein und gibt an, unter welchen Bedingungen eine Weiterschaltung von der aktiven Phase zur nächsten Phase vorgenommen wird. Der Übergangsindikator kann Manual, Always oder eine benutzerdefinierte Ausführung sein. Die benutzerdefinierte Ausführung ist das Ergebnis einer Berechnung, in der jede konfigurierte Prozess- oder Systemvariable verwendet werden kann.

Bild 4.11 (links) zeigt die Run-Time Version des verwendeten Grundrezepts, das sog. Steuerrezept, mit den Operationsblöcken *STANDBY*, *INOCULATION*, *BATCH*, *FED BATCH* und *INDUCTION*. Bild 4.11 (rechts) zeigt die zwei parallel ablaufenden Pfade der Operation *STANDBY*, die vor dem Animpfen des Bioreaktors abgearbeitet werden. Abgeschlossene Prozedurelemente sind in cyan hinterlegt, aktive Elemente in grün, noch nicht gestartete Elemente sind grau.

Im linken Pfad werden Prozessparameter festgelegt und Startbedingungen für die externen Programme MFCS-Tools und MONI vorgegeben. Im rechten Ablaufpfad werden die benötigten Regler, wie in Tabelle 9.5 aufgeführt, initialisiert. Die letzte Phase der S88-Operation *STANDBY* muss manuell vom Benutzer beendet werden. Das Rezept springt daraufhin in die nächste Operation *INOCULATION*.

⁵ SFC - sequenial function chart

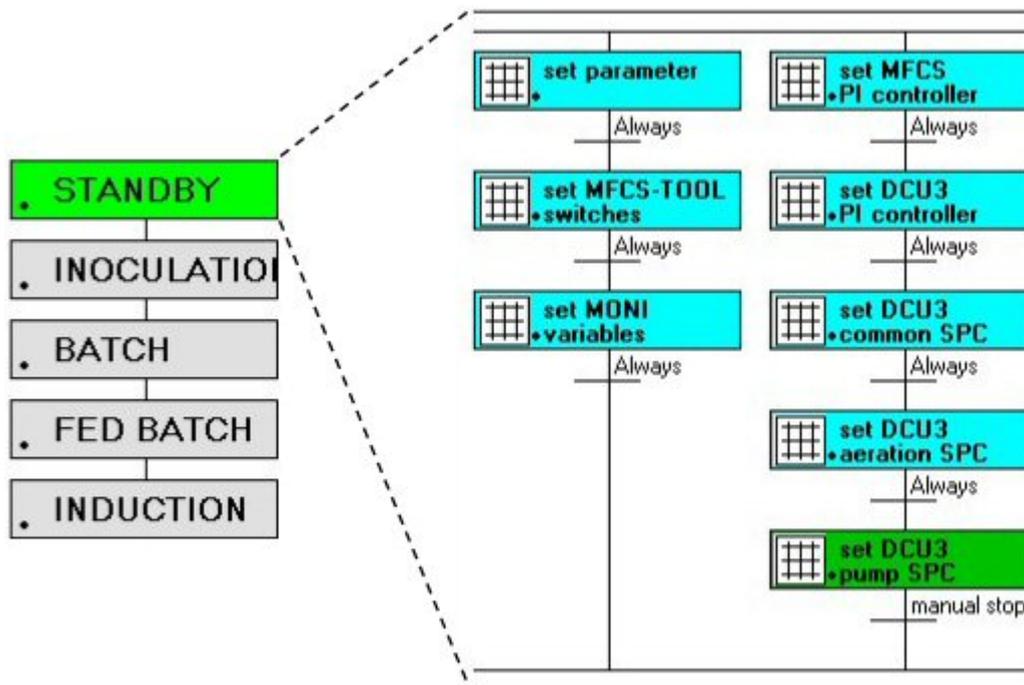
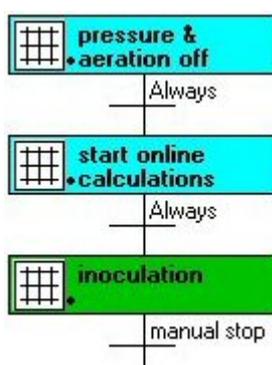


Bild 4.11: Operationspfad des S88-Rezepts (links) mit den Phasen der Operation *STANDBY*

In der ersten Phase der Operation *INOCULATION* wird zunächst der Sollwert für die Luftbegasung F_{nAIRw} von 12.0 lmin^{-1} auf 0.0 lmin^{-1} herabgesetzt. Der Regler für den Druck im Bioreaktor wird in Handbetrieb genommen und die Stellgröße y_{pG} auf 100 % gesetzt und somit der Abgasweg vollständig geöffnet.

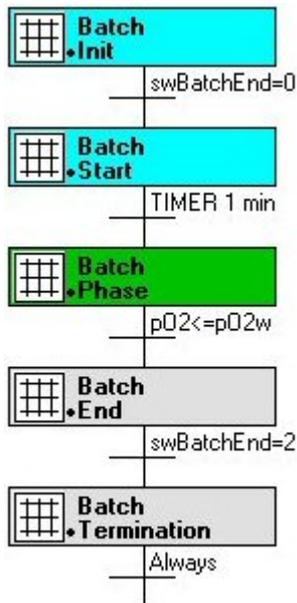
In der zweiten Phase werden Online-Formeln aktiviert, die zur Berechnung von nicht messbaren, abgeleiteten Parametern des Kultivierungsprozesses genutzt werden. Dazu zählen z. B. die Sauerstoffeintragsrate Q_{O_2} , die Kohlenstoffdioxidaustragsrate Q_{CO_2} und der aus diesen Größen ermittelte Respirationsquotient RQ. Des Weiteren werden hier die im Kapitel 3 beschriebenen Formeln zur Berechnung der Zelldichte sowie der Methanol- und der Produktkonzentration im Medium gestartet.



In der dritten Phase wird die Batchzeit synchronisiert, d. h. auf null zurückgestellt. Daten die zuvor gespeichert wurden, bekommen eine negative Zeit. Nachdem der Druck im Bioreaktor einen Wert von 10 mbar unterschritten hat, wird die Vorkultur mit der Harvestpumpe F_H mit zuvor eingestellter inverser Drehrichtung in den Bioreaktor überführt. Nach dem Animpfen wird die Phase manuell vom Benutzer beendet. Das Rezept springt in die nächste Operation *BATCH*.

Bild 4.12: Der Operationsblock *INOCULATION*

Der Operationsblock *BATCH* wird durch fünf Phasen codiert. Die Bezeichnungen *Init*, *Start*, *Phase*, *End* und *Termination* werden für die folgenden Operationsblöcke des S88-Rezepts beibehalten.



In der *Init*-Phase wird ein Timer-Reset durchgeführt. Der Status der zyklischen Variable *TIMER* wird zunächst auf *off* dann auf *on* geschaltet. In der Übergangsbedingung wird die Schaltervariable *sw_BatchEnd* auf ihren Zustand überprüft. Der Anfangszustand dieser Variablen [0] wurde in der Operation *STANDBY* gesetzt. Ohne manuellen Benutzereingriff ist die Transition erfüllt und das Rezept wechselt in die nächste Phase.

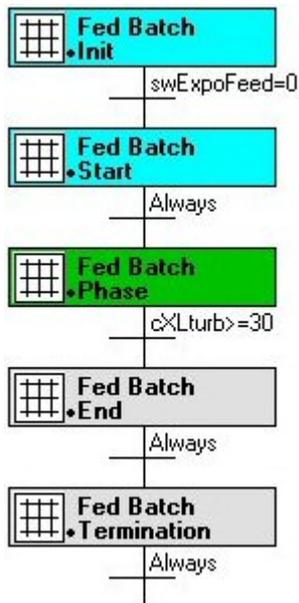
Während der *Start*-Phase wird zunächst der Sollwert der Luftbegasung F_{nAIRw} wieder auf 12 lmin^{-1} angehoben. Der Regler für den Druck im Bioreaktor wird erneut in Automatikbetrieb mit einem Sollwert p_{Gw} von 500 mbar gesetzt. Anschließend werden die pH-Regelung und der Timer gestartet.

Bild 4.13: Der Operationsblock *BATCH*

Die Maßnahmen während der Inokulation bewirken ein Sinken des pO_2 -Wertes. Um zu vermeiden, dass die Transition $pO_2 \leq pO_2w$ der *Batch Phase* ungewollt wahr wird, findet der Übergang in diese Phase erst nach Ablauf einer Minute statt.

Der Timer wird in der *Batch Phase* deaktiviert. Sobald der pO_2 seinen Sollwert erreicht hat, wird in der Phase *Batch End* die pO_2 -Regelung und die automatische Batchende-Erkennung gestartet. Ist das Batchende erkannt, wird die Schaltervariable *sw_BatchEnd* durch das Programm MFCS-Tools auf den Wert [2] gesetzt. Da das Umschaltkriterium erfüllt ist, geht das Rezept zum nächsten Operationsblock *FED BATCH* über.

Der Operationsblock *FED BATCH* beinhaltet fünf sequentiell ablaufende Phasen, in denen die exponentielle Glycerolzufütterung initialisiert und beendet wird. Die *Init*-Phase enthält keine Befehle, sondern lediglich eine Übergangsbedingung, bei der die Schaltervariable *sw_ExpoFeed* des MFCS-Tools auf ihren Zustand überprüft wird. Der Anfangszustand dieser Variablen [0] wurde in der Operation *STANDBY* gesetzt. Ohne manuellen Benutzereingriff ist die Transition erfüllt.

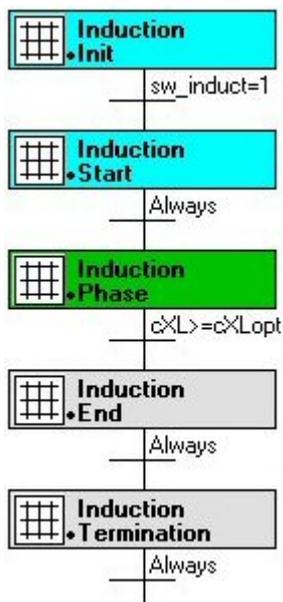


Das Rezept wechselt in die Start-Phase. Hier wird der Wert der Schaltervariablen auf [1] gesetzt und damit die exponentielle Glycerolzufütterung eingeleitet. Der Übergang in die nächste Phase findet automatisch statt, da keine Transition definiert ist – der Übergangsindikator steht auf *Always*.

Nachdem eine Zelldichte c_{XLturb} von 30 gl^{-1} erreicht ist, wird die Glycerolzugabe beendet. Dafür wird die Schaltervariable *sw_ExpoFeed* in der End-Phase wieder auf den Wert [0] gesetzt. Der Übergang in die Termination-Phase findet automatisch statt. Hier wird die Schaltervariable *sw_induction* auf den Wert [1] gesetzt.

Bild 4.14: Phasen des Operationsblocks *FED BATCH*

Der Operationsblock *INDUCTION* beinhaltet fünf sequentiell ablaufende Phasen. Die Init-Phase enthält keine Befehle. Die Schaltervariable *sw_induction* wird auf ihren Zustand überprüft. Ohne manuellen Benutzereingriff ist die Transition erfüllt.



In der Start-Phase wird die Methanolregelung mit einem Sollwert c_{S2Mw} von 2 gl^{-1} initialisiert. Der Übergang in die *Induction Phase*, die keine Befehle enthält, findet automatisch statt. Ist die für eine Filtration optimale Zelldichte c_{XLopt} von 40 gl^{-1} erreicht, wird in der Termination-Phase die Schaltervariable *sw_caking* auf den Wert [1] und *sw_induction* auf den Wert [0] gesetzt.

Die End-Phase enthält keine Befehle, wurde aus Gründen der Einheitlichkeit jedoch beibehalten.

Bild 4.15: Phasen des Operationsblocks *INDUCTION*

4.5 Vollautomatischer Prozessablauf

4.5.1 S88-gesteuerte Inokulation

In Bild 4.16 ist der S88-gesteuerte Animpfvorgang des zunächst in Stand-By befindlichen Bioreaktors zu sehen. Am Anfang der Inokulation wurde der Sollwert für die Luftbegasung F_{nAIRw} von 12.0 lmin^{-1} auf 0.0 lmin^{-1} herabgesetzt. Des Weiteren wurde der Regler für den Druck im Bioreaktor in Handbetrieb genommen und die Stellgröße y_{pG} auf 100 % gesetzt. Anschließend erfolgte eine Synchronisation der Batchzeit.

Bei $t = 0.03 \text{ h}$ erreichte der Druck im Bioreaktor einen Wert von 10 mbar. Die Vorkultur wurde über den Anschluss für die Zellernte in den Bioreaktor überführt. Die dafür genutzte Harvestpumpe F_H wurde mit zuvor eingestellter inverser Drehrichtung aus MFCs/win mit einem Sollwert F_{Hrelw} von 100 % angesteuert. Während des Animpfvorganges war eine Abnahme des pO_2 auf 67.0 % zu verzeichnen.

Nach dem Transfer der Vorkultur in den Bioreaktor wurde der Sollwert für die Luftbegasung F_{nAIRw} bei $t = 0.07 \text{ h}$ wieder auf 12 lmin^{-1} angehoben. Der Regler für den Druck im Bioreaktor wurde erneut in Automatikbetrieb mit einem Sollwert p_{Gw} von 500 mbar gesetzt.

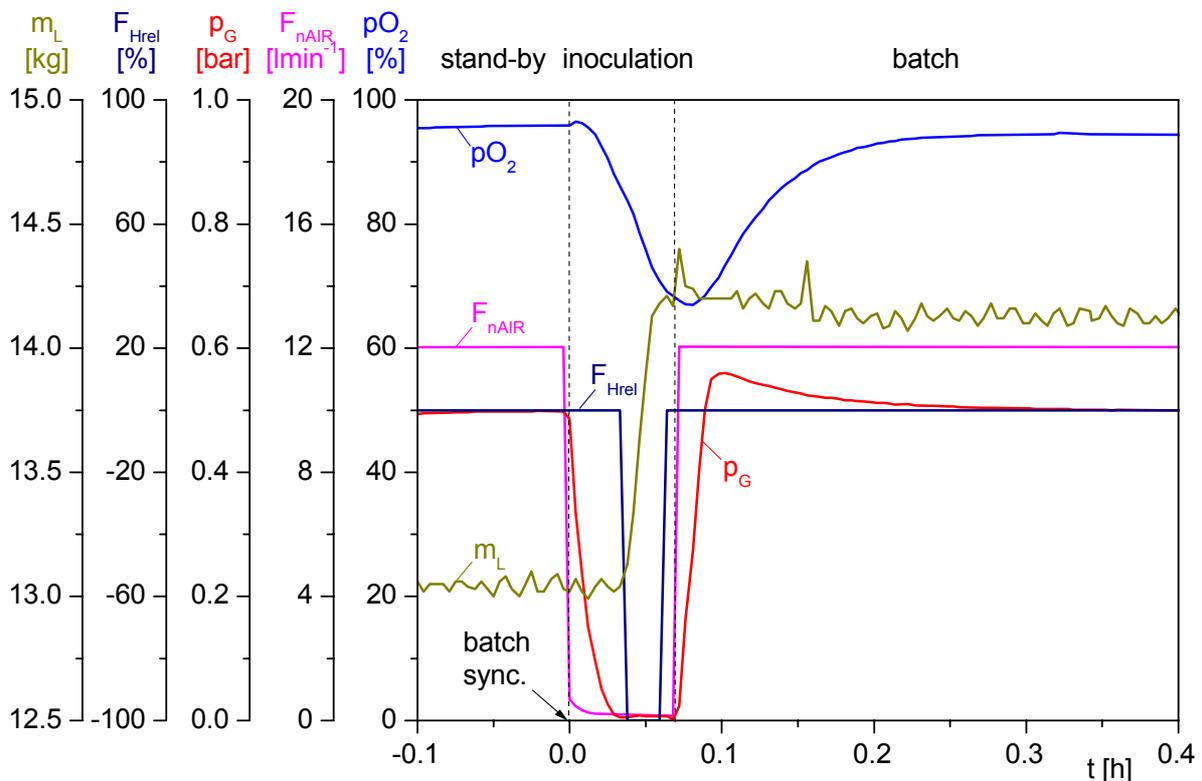


Bild 4.16: S88-gesteuertes Animpfen des Bioreaktors

4.5.2 Automatische Batchende-Erkennung

Das Modul *batch end* des Visual Basic[®] Programms MFCS-Tools wurde zur automatischen Batchende-Erkennung eingesetzt.

Der pO_2 wurde bei dem in Bild 4.17 dargestellten Experiment auf einen Sollwert von 10 % geregelt. Als Kriterien für die Batchende-Erkennung wurden ein pO_2 -Anstieg von 10 % sowie ein Absinken der Rührerdrehzahl N_{St} von 40 min^{-1} innerhalb von 90 s (= 0.025 h) festgelegt.

Nach Verbrauch des vorgelegten Glycerols zum Ende der Batchphase nahm der Sauerstoffbedarf der Zellen ab. Die Sauerstoffeintragsrate Q_{O_2} blieb zunächst konstant bei 2.91 g(lh)^{-1} und bewirkte ein schnelles Ansteigen des relativen Sauerstoffpartialdrucks in der Flüssigphase. Die pO_2 /agitation-Regelung versuchte weiterhin den Sollwert zu halten und senkte daher die Drehzahl ab. Zum Zeitpunkt $t = 32.642 \text{ h}$ wurde die Schaltervariable $sw_BatchEnd$ von [1] auf [2] erhöht und damit das Batchende detektiert.

Nach der automatischen Batchende-Erkennung wurde das Visual Basic[®] Modul *exponential feed* zum Zeitpunkt 32.646 h über das Setzen der Schaltervariable $sw_ExpoFeed$ von [0] auf [1] initialisiert. Die exponentielle Zufütterung startete bei $t = 32.651 \text{ h}$ mit einer relativen Zufütterrate F_{R1rel} von 52.41 %.

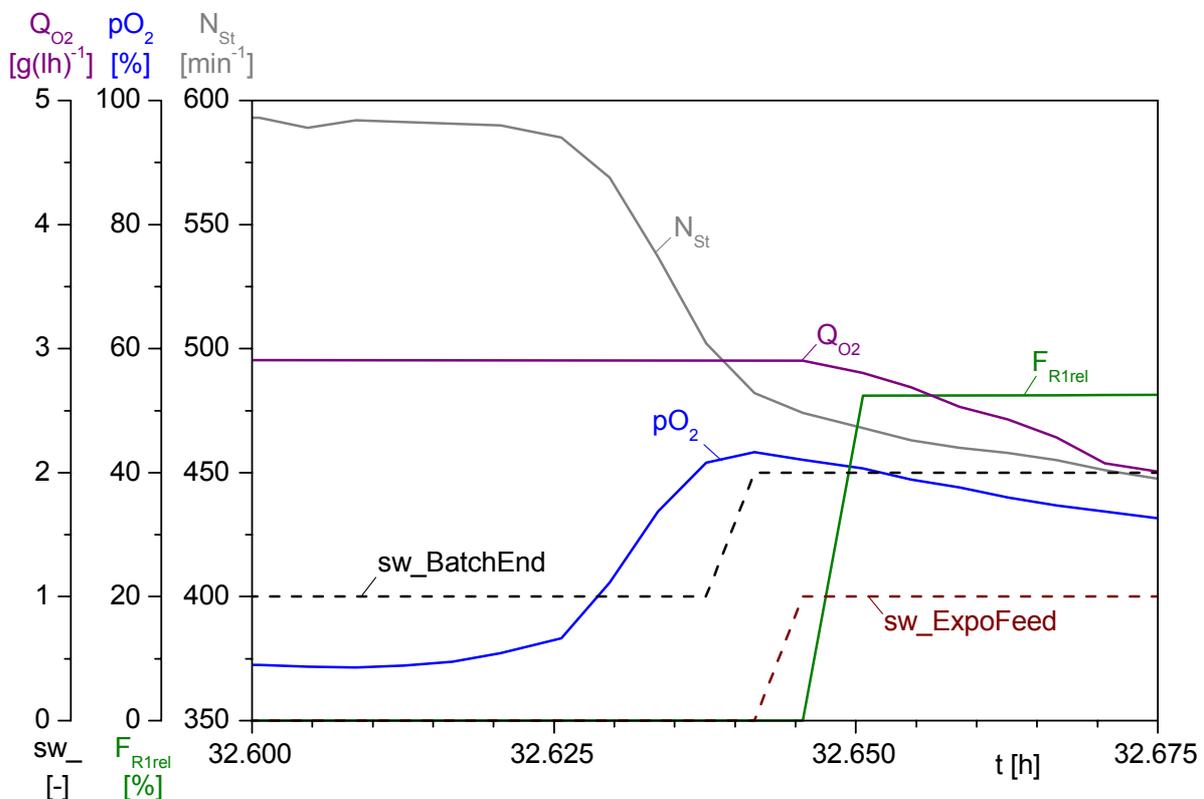


Bild 4.17: Automatische Batchende-Erkennung mit Start der Zufütterung

4.5.3 Steuerung der zellspezifischen Wachstumsrate

In Bild 4.18 ist das Ende der Batchphase, die Steuerung der zellspezifischen Wachstumsrate in der Fed Batch-Phase sowie der Beginn der Induktionsphase dargestellt. Durch eine gezielte limitierte Substratzugabe sollte eine Steuerung der zellspezifischen Wachstumsrate auf einen Sollwert μ_{1w} von 0.1 h^{-1} erfolgen.

Die Glycerolzufütterung begann mit einer relativen Zufüterraterate F_{R1rel} von 52.4 %. Zur Berechnung dieses Startwertes nach Gleichung (4.4) wurde die über Trübungsmessung geschätzte Zelldichte $c_{XL,turb}$ mit 17.1 gl^{-1} und das online gemessene Reaktionsvolumen V_L mit 13.3 l verwendet. Die Glycerolreservoirkonzentration wurde auf 350 gl^{-1} eingestellt, die maximale Zufüterraterate F_{R1max} mit 0.18 h^{-1} vorgegeben.

Der Verbrauch des vorgelegten Substrats zum Ende der Batchphase bewirkte einen sofortigen Rückgang des Sauerstoffbedarfs der Mikroorganismen, erkennbar an der zellspezifischen Sauerstoffaufnahme $q_{O/X}$. Resultierend nahm die beobachtbare Wachstumsrate $\hat{\mu}_{O_2}$, wie in Bild 4.18 dargestellt, abrupt ab. Derselbe Effekt war beim Übergang von der Fed Batch-Phase in die Induktionsphase bei $t = 41.8 \text{ h}$ zu erkennen. Nach dem Start der Zufütterung war ein erhöhter Sauerstoffbedarf zu verzeichnen, wodurch die Wachstumsrate zunächst anstieg, dann im weiteren Verlauf der μ -Steuerung allerdings unterhalb des Sollwerts μ_{1w} von 0.1 h^{-1} lag.

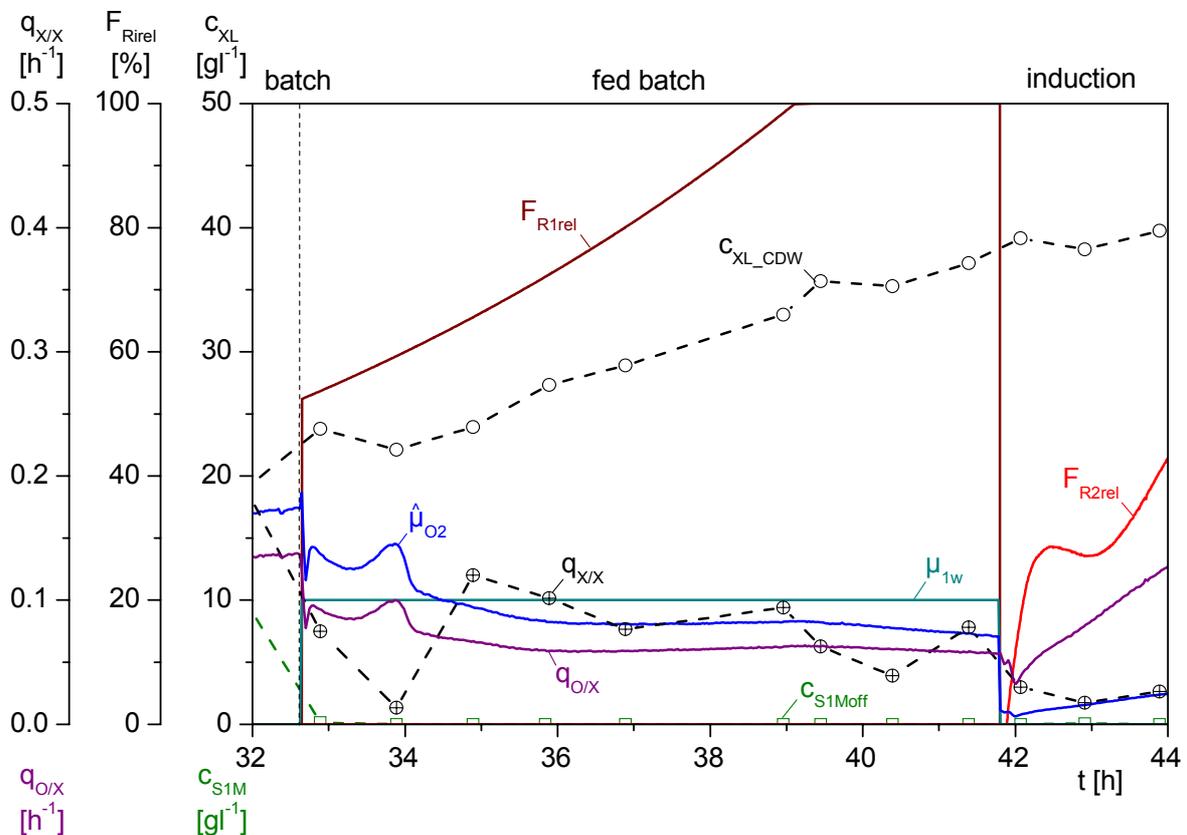


Bild 4.18: Steuerung der zellspezifischen Wachstumsrate

4.5.4 Induktion der Produktbildung

In Bild 4.19 ist das Ende der Fed Batch-Phase, die Induktion sowie der Verlauf einer darauf folgenden Produktionsphase dargestellt. Die Methanolkonzentration c_{S2M} wurde über die Sonde des FMC-Systems detektiert und ab $t = 41.8$ h auf einen Sollwert von 2 g l^{-1} geregelt. Mit Beginn der Methanolregelung stieg die volumetrische Methanoleintragsrate Q_{S2in} bis auf $7.05 \text{ g l}^{-1} \text{ h}^{-1}$. Ab $t = 46.0$ h ging der Methanolbedarf der Zellen jedoch zurück.

Die Proteinexpression setzte unmittelbar nach der Induktion ein. Nach einem zunächst steilen Anstieg, flachte die Produktbildung ab $t = 45.0$ h jedoch stark ab. Zum Ende der Produktionsphase war bei $t = 51.2$ h eine Proteinkonzentration im Medium c_{P1M} von 77.6 mg l^{-1} erreicht.

Während der Produktionsphase war eine Zunahme der Zelldichte c_{XL_OD} von 30.5 g l^{-1} bis auf 40.6 g l^{-1} zu verzeichnen. Die aus den Offline-Werten ermittelte zellspezifische Wachstumsrate $q_{X/X}$ lag für diesen Abschnitt durchschnittlich bei 0.031 h^{-1} . Die über den O_2 -Massentransferstrom berechnete Wachstumsrate verlief im Durchschnitt bei 0.029 h^{-1} .

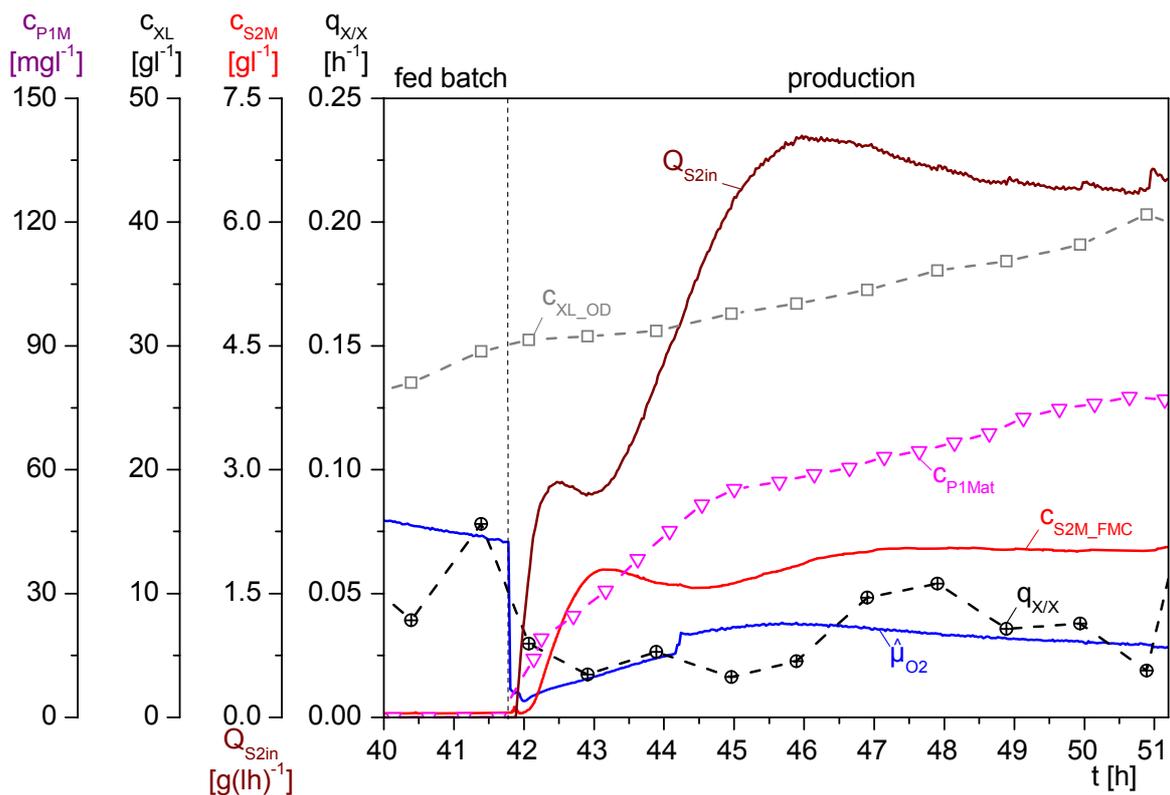


Bild 4.19: Induktion der Proteinexpression durch Zufütterung von Methanol

4.5.5 Sequentielle pO_2 -Regelung

Während der in Bild 4.20 dargestellten Batchphase verlief der pO_2 zunächst unregelt. Bei Erreichen des Sollwertes pO_{2w} von 10 % bei $t = 19.4$ h wurde eine Vermeidung der O_2 -Limitierung durch Anhebung der Luftbegasungsrate F_{nAIR} sichergestellt. Die pO_2 /Drehzahl-Regelung wurde initialisiert, nachdem die Luftbegasungsrate bei $t = 20.5$ h die Schaltschwelle $F_{nAIRmax}$ von 18 lmin^{-1} erreichte.

Zu Beginn der Zugabe von Methanol bei $t = 41.8$ h war ein starker Anstieg der Rührerdrehzahl zu verzeichnen. Dies ist auf ein verändertes Koaleszenzverhalten der Blasen, d. h. einen starken Anstieg des k_La -Wertes von 122.4 h^{-1} auf 815.2 h^{-1} bei $t = 44.2$ h zurückzuführen. Zu diesem Zeitpunkt erreichte die Rührerdrehzahl ihren Maximalwert von 800 min^{-1} , so dass trotz der geringen Zelldichte von 39.75 g l^{-1} eine Beimischung von reinem Sauerstoff, beobachtbar in F_{nO_2} , in die Zuluft notwendig wurde.

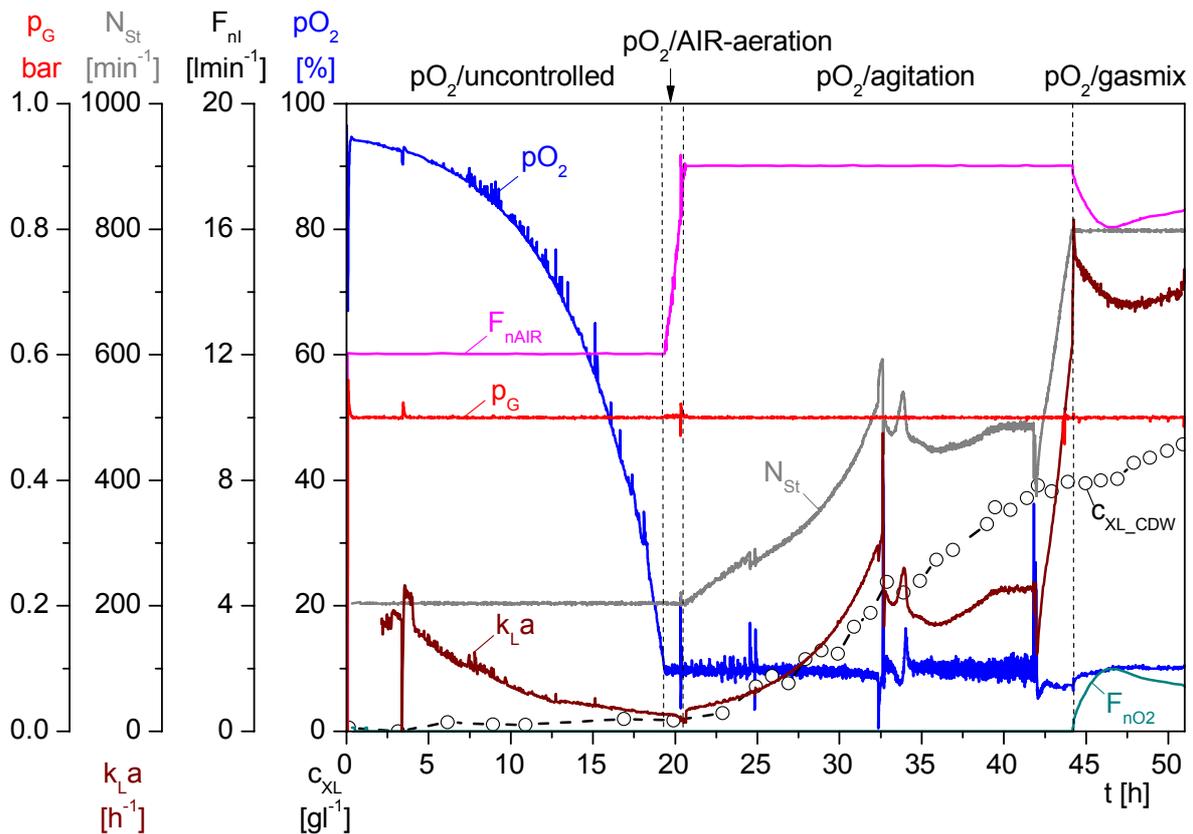


Bild 4.20: Sequentielle Regelung des Sauerstoffpartialdrucks

5 Integrierte Prozessführung – die Verbindung von Up- und Downstream

5.1 Das Konzept der Integrierten Prozessführung

5.1.1 Gewählter Prozessaufbau

Der klassische Produktionsablauf kann, wie in Kapitel 4 vorgestellt, in die unlimitierte Batch-Phase, die glycerollimitierte Fed Batch-Phase und die methanolgeregelte Produktionsphase unterteilt werden. Da das herzustellende Chemokin 1-3del I-TAC durch *P. pastoris* in das Kulturmedium sekretiert wird, ist es möglich, den Upstream-Prozess mit primären Downstream-Schritten in einem Integrierten Bioprozess zu verknüpfen.

Unter Integrierter Bioprozessführung versteht man die direkte prozesstechnische Verknüpfung von Kultivierungs- und Primäraufarbeitungsschritten bis hin zu Feinreinigungsmethoden bei der Herstellung wertschöpfender biotechnologischer Produkte (Cornelissen *et al.*, 2003).

In Bild 5.1 ist der konzeptionelle Aufbau einer Anlage zur Integrierten Prozessführung gezeigt, bei der mittels einer direkt an den Bioreaktor angekoppelten Querstrom-Mikrofiltration das sekretierte Zielprotein im Diafiltrationsmodus abgetrennt wird. Das Permeat wird in einen kühlbaren Produkttank überführt und kann über chromatographische Methoden unter Einsatz des ÄKTA™ purifier 100 aufgereinigt werden (Hahn, 2001).

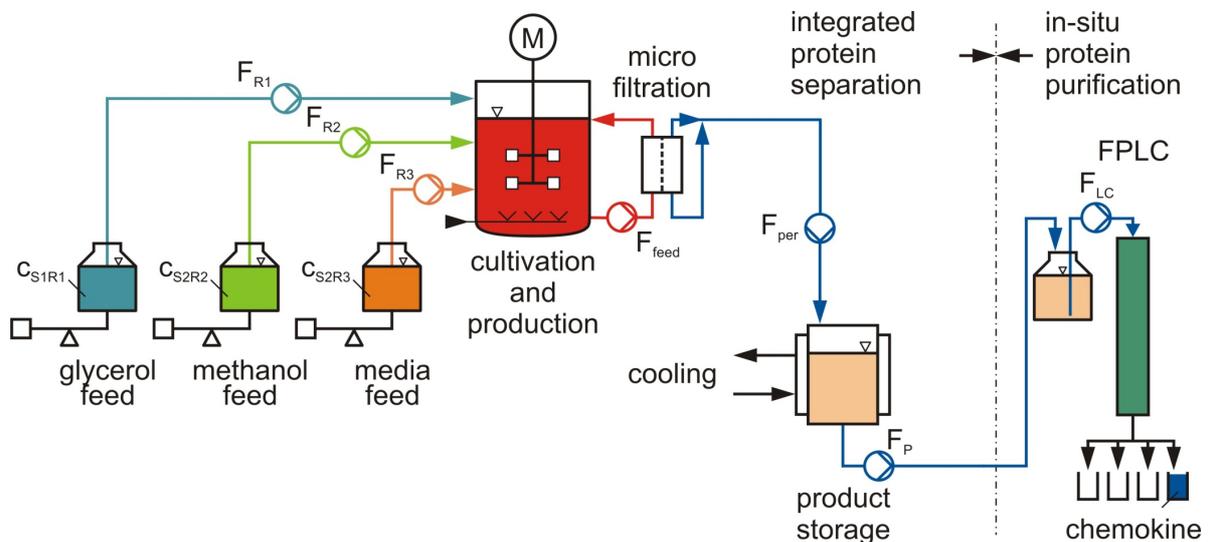


Bild 5.1: Integrierte Prozessführung zur Herstellung von sekretierten Pharmaproteinen

Der in Bild 5.2 dargestellte BIOSAT® C-Xflow der Fa. B. Braun Biotech International wurde für die beschriebene Integrierte Bioprozessführung entwickelt. Das Bioreaktorsystem besteht aus der bereits in Kapitel 4.2 vorgestellten Kultivierungseinheit mit einer Zufütterstation über sieben Vorlagen, einem Zellerntebehälter, einer Crossflow-Filtrationseinheit mit dazugehöriger Feedpumpe und entsprechender Einbindung in den Reaktionsprozess sowie einem kühlbaren Produkttank.

Die zugehörige Vorortautomatisierung DCU3, das übergeordnete SCADA-System MFCS/win sowie ein integriertes CIP-System ergänzen den Aufbau. Des Weiteren sind die bereits beschriebenen Analysensysteme zur Substrat- und Produktmessung angeschlossen.

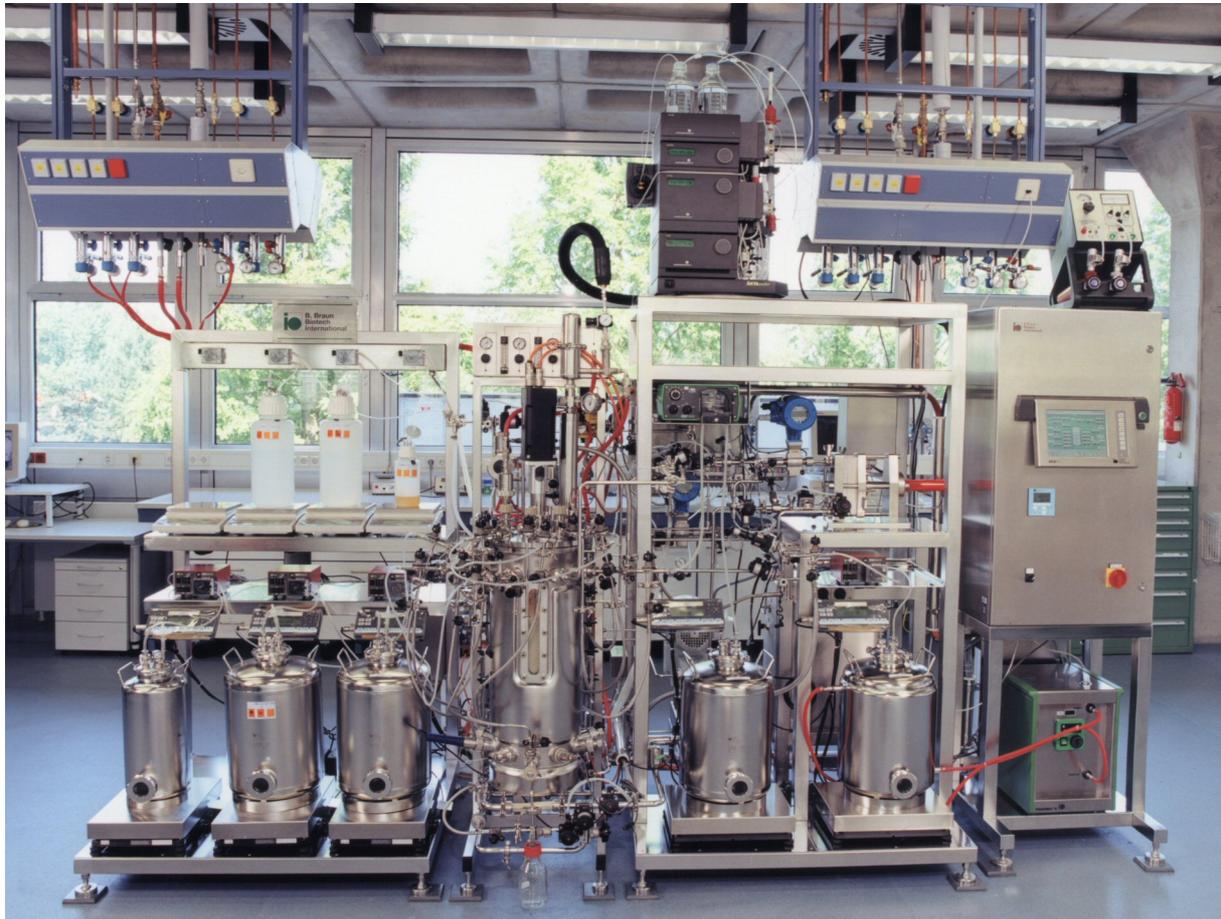


Bild 5.2: Die BIostat[®] C-Xflow Anlage

Wie in Bild 5.3 gezeigt, wird bei der Crossflow-Filtration die Membran parallel von der Kulturbühe überströmt, so dass Überströmungs- und Filtratflussrichtung kreuzweise zueinander stehen. Nach dem Anfahren der Filtration lagern sich zurückgehaltene Partikel ähnlich einem Filterkuchen auf der Membran ab. Bis auf eine dünne Grundschicht werden diese allerdings durch Scher- und Auftriebskräfte an der Membranoberfläche wieder in die Kernströmung zurückgeführt. Im Idealfall bleibt die Deckschichtdicke nach einer Einfahrzeit konstant, so dass daraus eine stationäre Betriebsphase mit konstantem Permeatflux J_{per} und Transmembrandruck p_{TM} resultiert (Kraume, 2003).

Der Aufbau von Deckschichten (Sekundärschichten) und Porenverstopfung (fouling) beeinträchtigen sehr stark die Leistung einer Filtration (Boerlage *et al.*, 1997; Meyer *at al.*, 1998; Wang and Song, 1999). Da sich zu Beginn der ersten Mikrofiltration noch keine Deckschicht auf der Membranoberfläche gebildet hat, ist die Geschwindigkeit, mit der das Retentat an der Membran vorbeiströmt, noch nicht konstant und ausreichend für die Filtrationsbedingungen.

Daher findet einmalig vor der ersten Mikrofiltration bzw. nach jedem Reinigungszyklus der Membranen eine Kuchenbildung statt. Da die Membranen einen signifikanten Kostenfaktor darstellen, wird durch die mehrfache Wiederverwendung der Prozess wirtschaftlicher. Für diesen Zweck wurden die Filterkassetten mit $1 \text{ mol}^{-1} \text{ NaOH}$ bei $40 \text{ }^\circ\text{C}$ in einer speziellen Apparatur gereinigt.

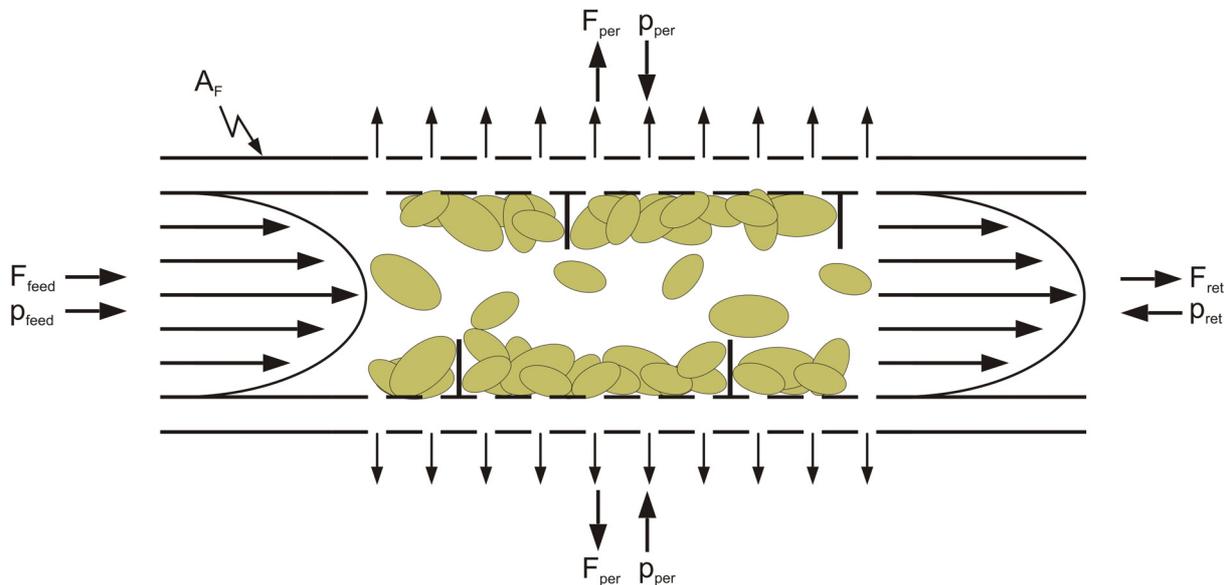


Bild 5.3: Schematische Darstellung der Crossflow-Filtration

Mit Bezug auf die eingesetzte Filterfläche A_F erzeugt jeder Volumenstrom F_{flow} einen flächenbezogenen Flux J_{flow} ,

$$J_{\text{flow}}(t) = \frac{F_{\text{flow}}(t)}{A_F}, \text{ flow} = \text{feed, ret, per.} \quad (5.1)$$

Zur Charakterisierung von Membranen bzw. als Schlüsselgröße für die Modul- und Anlagenauslegung wird entweder der Permeatflux J_{per} oder die Permeabilität P_{per} ,

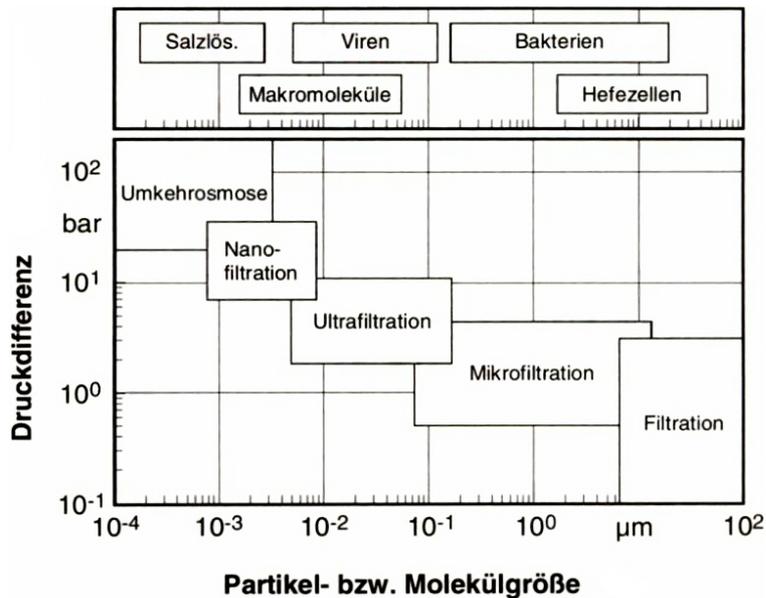
$$P_{\text{per}}(t) = \frac{J_{\text{per}}(t)}{p_{\text{TM}}(t)}, \quad (5.2)$$

angegeben, in der zusätzlich die transmembrane Druckdifferenz p_{TM} ,

$$p_{\text{TM}}(t) = \frac{p_{\text{feed}}(t) + p_{\text{ret}}(t)}{2} - p_{\text{per}}(t), \quad (5.3)$$

berücksichtigt wird. Der Transmembrandruck p_{TM} beschreibt die mittlere Druckdifferenz von der Feedseite zur Permeatseite.

In dieser Arbeit wurde mit einem Eingangsdruck von 2.0 bar, einem Retentatdruck von 1.0 bar sowie einem Permeatdruck von 0.5 bar gearbeitet.



Der Transmembrandruck ist die Triebkraft für die in Bild 5.4 gezeigten druckgetriebenen Membranverfahren. Mikrofiltrationen werden typischerweise mit einer Druckdifferenz zwischen 0,5 bis 5 bar durchgeführt. Die Porengröße der Membranen liegt mit ca. 0,08 bis 10 μm in einem Bereich, der es erlaubt, sowohl Bakterien als auch Hefezellen zurückzuhalten.

Bild 5.4: Zuordnung der druckgetriebenen Membranverfahren (Rautenbach, 1997)

5.1.2 Die Sequentiell Integrierte Bioprozessführung

Zur Durchführung mehrerer Produktionsphasen innerhalb eines Prozesses und als Regenerierungsmöglichkeit für die Zellen wurde eine Sequentiell Integrierte Prozessführungsstrategie in Anlehnung an Cornelissen (2004) verwendet. Aus dieser quasi-kontinuierlichen Fahrweise resultieren nicht nur Vorteile einer gleichzeitigen Abtrennung unerwünschter Nebenprodukte, sondern auch einer wirtschaftlicheren Nutzung der Anlage. Sie erfordert allerdings ein gewisses Maß an Automatisierbarkeit (Schügerl, 2000; Stark and von Stockar, 2003; Jahic *et al.*, 2006).

Die Strategie unterscheidet vier sich zyklisch wiederholende Hauptphasen, wobei eine Sequenz immer mit der Produktion beginnt und von den Phasen Mikrofiltration, Zellernte und Mediaauffrischung komplettiert wird. Die erste Produktionsphase der Sequentiell Integrierte Bioprozessführung (SIB) entspricht dabei der letzten Phase des klassischen Produktionsablaufs. Im Folgenden wird dieser Ansatz weiter ausgearbeitet und in eine vollautomatische Prozessführungsstrategie übertragen.

In Bild 5.5 sind die sich zyklisch wiederholenden Hauptphasen der SIB mit den Signalen $c_{XL_{turb}}$, m_L und m_p für die Steuerung des Ablaufs der einzelnen Sequenzen gezeigt. Die Methanolkonzentration im Medium c_{S2M} wird bis zum Abbruch der Kultivierung auf einen Sollwert c_{S2Mw} von 2 gl^{-1} geregelt.

In Vorversuchen wurde festgestellt, dass das integrierte Filtrationsverfahren bei Zelldichten im Bereich von 40 bis 60 gl^{-1} BTM noch eine ausreichende Produktivität sowie eine moderate Standzeit der Module gewährleistet (Cornelissen, 2004).

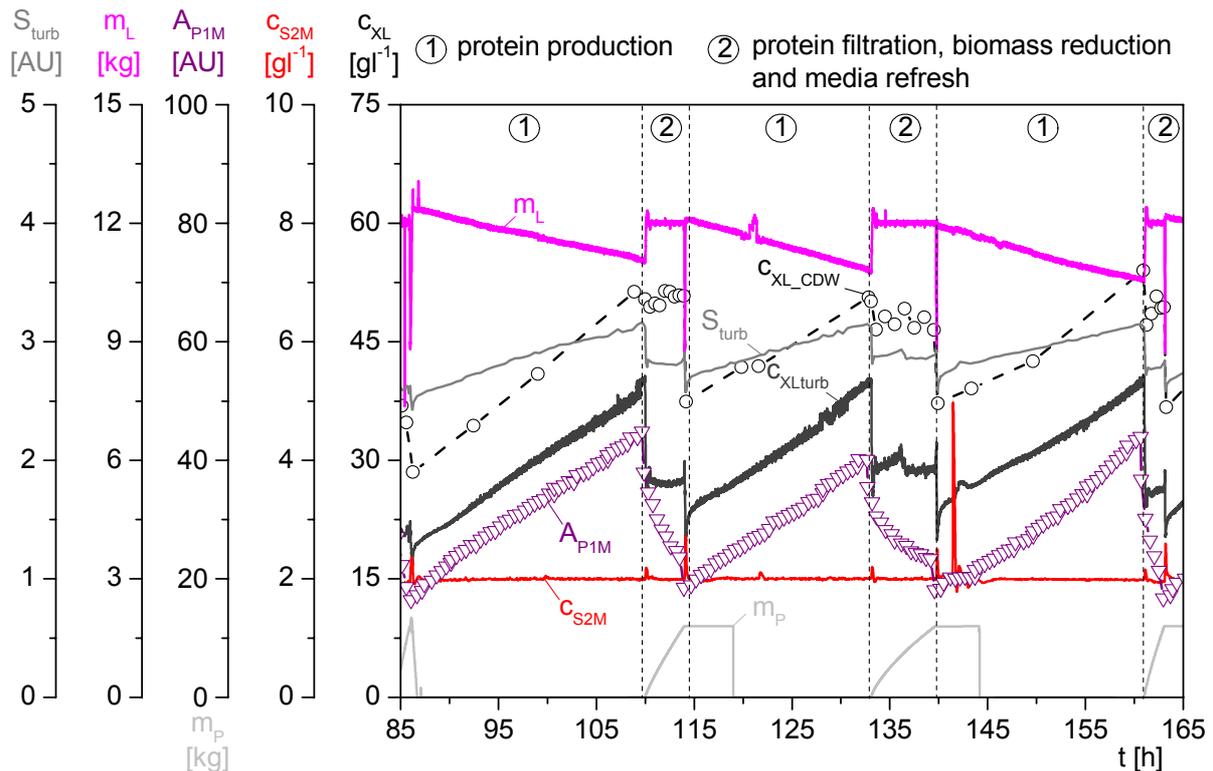


Bild 5.5: Automatisierung der Sequentiell Integrierten Prozessführung

Anhand der ausgearbeiteten Prozessführungsstrategie wurde eine Mikrofiltration immer dann gestartet, sobald die Hefezellen eine Zelldichte $c_{XL_{turb}}$ von 40 gl^{-1} erreicht hatten. Die mit Hilfe des Trübungssignals S_{turb} berechnete Zelldichte zeigte eine Abweichung um ca. 10 gl^{-1} gegenüber der Biotrockenmasse $c_{XL_{CDW}}$. Diese Differenz lässt sich durch die Wahl falscher Parameter zur Berechnung von $c_{XL_{turb}}$ erklären. Sie beeinflusst dennoch die Reproduzierbarkeit des Prozesses nicht.

Während der Filtrationsphasen war die Abweichung jedoch wesentlich größer. Da die Masse des Reaktionsvolumens m_L in den Produktionsphasen durch die Probenahme sank, wurde zu Beginn einer Mikrofiltration die Kulturbrühe durch Initialisierung der Füllstandsregelung verdünnt. Dieser Schritt erklärt aber vermutlich nur zum Teil die starke Abnahme der Zelldichte $c_{XL_{turb}}$. Durch die Zufuhr von frischem Medium geht die gelbe Eigenfärbung der Kulturbrühe zurück. Das Trübungssignal könnte durch diesen Vorgang ebenfalls beeinflusst werden.

Die Mikrofiltration zur Rückhaltung der Zellen im Bioreaktor und zur Abtrennung des Zielproteins wird als Diafiltration durchgeführt. Abgeführtes Permeat wird dabei über die Regelung der Masse des Reaktorvolumens auf einen Sollwert von 12 kg durch frisches Medium ersetzt. Sobald die Masse des produktbehafteten Mediums im Permeattank m_P den für einen Massenaustausch notwendigen Sollwert m_{Pw} von 12 kg erreichte, wurde die Filtration automatisch beendet und die Füllstandsregelung deaktiviert.

Die anschließende Konditionierung der Kulturbrühe durch die Phasen Zellernte und Mediumauffrischung ist notwendig, um die gleichen Ausgangsbedingungen zu schaffen, die zu Beginn der Induktion vorlagen. Die Refreshphase ist beendet, sobald die Masse des Reaktionsvolumens m_L wieder 12 kg erreicht hat. Der nächste Zyklus beginnt erneut mit einer Produktionsphase. Die Hefezellen wachsen wie zuvor bis zu einer Zelldichte c_{XLturb} von 40 gl^{-1} an, ehe die nächste Mikrofiltration beginnt.

Die Waagensignale m_p und m_L bilden ein sicheres Umschaltkriterium für die Steuerung des Ablaufs der einzelnen Sequenzen. Die berechnete Zelldichte c_{XLturb} hingegen ist als Indikator nur bedingt einsetzbar.

5.1.3 MSR-Aufgaben während der Integrierten Prozessführung

Bild 5.6 zeigt den erweiterten Aufbau der Reaktoranlage mit den unterschiedlichen Mess- und Regelungssystemen, um eine Integrierte Prozessführung vollautomatisch zu realisieren. Die aus der Trübungsmessung berechnete Zelldichte c_{XLturb} sowie die beiden Waagensignale m_L und m_p dienen zur Steuerung des Ablaufs der einzelnen Sequenzen.

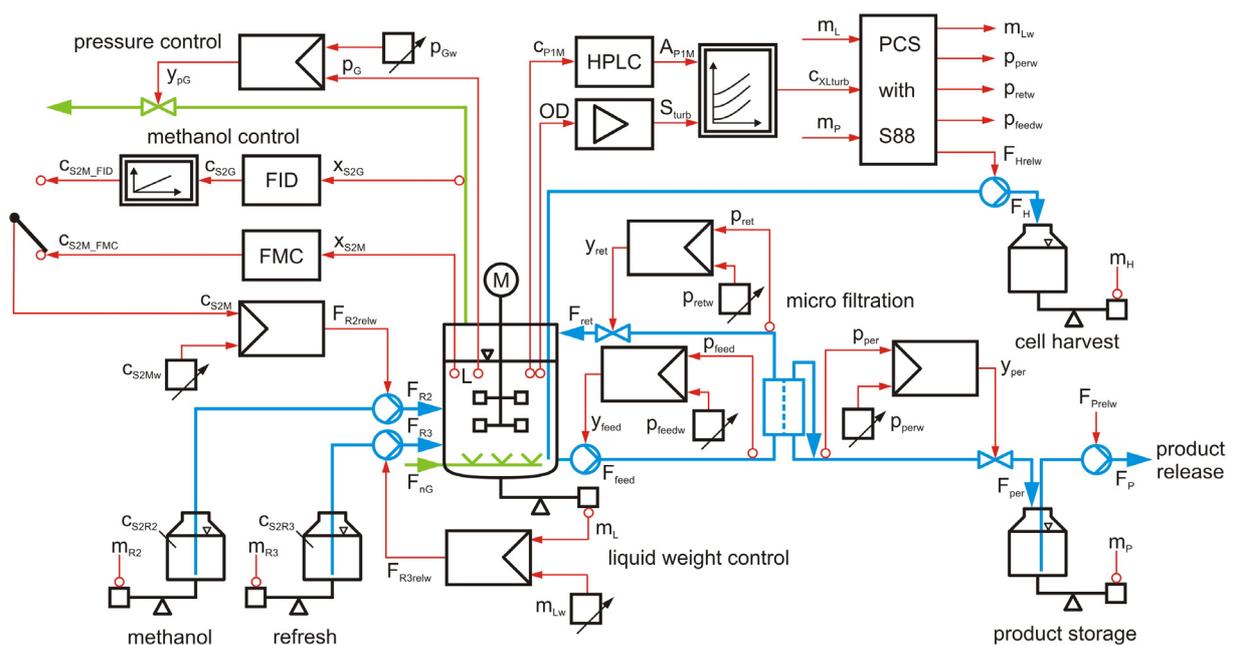


Bild 5.6: MSR-Aufgaben während der Integrierten Prozessführung

Das Inline-Messsystem FMC-MINI 03 Fa. Kempe Biotechnologie dient nur als Messverstärker und damit als Messwertgeber für die Regelung der Methanolkonzentration im Medium während der sequentiell integrierten Prozessführung. Alternativ kann der über eine redundante FID-Messung online ermittelte Prozesswert c_{S2M_FID} zur Regelung eingesetzt werden.

Die Filtrationsbedingungen werden über drei sich gegenseitig beeinflussende Druckregelungen eingestellt. Der Druck am Eingang des Filtermoduls p_{feed} wird durch die Förderleistung der Pumpe F_{feed} bestimmt. Am Filterausgang wirkt der Druck p_{ret} , der über ein stetiges Ventil im Retentatweg eingestellt wird. Der permeatseitige Druck p_{per} wird ebenfalls über ein stetiges Ventil gesteuert.

Das bei der Mikrofiltration abgeführte Permeat wird über die Regelung der Masse m_L des Reaktorvolumens durch frisches Medium ersetzt (Gasper, 2000). Daher bleibt die Zelldichte während der Filtration nahezu konstant.

Im Anhang sind die Automatisierungsaufgaben während der unterschiedlichen Phasen der Sequentiell Integrierten Bioprozessführung tabellarisch zusammengefasst.

5.1.4 *In-situ* product removal (ISPR) mit der Crossflow-Filtrationseinheit

Die in Bild 5.7 gezeigte Filtrationseinheit ist mit dem Bioreaktorkessel über das Bodenablassventil verbunden. Die Kulturbrühe wird über eine Kreiskolbenpumpe (Typ FLFN 55L, Fa. Fristam Pumpen) mit einem Spaltmaß von 0.15 mm zum Mikrofiltrationsmodul gefördert. Die Pumpe ist *in situ* sterilisierbar und hat eine maximale Förderrate von 9000 l h^{-1} für Flüssigkeiten mit spezifischem Gewicht von 1 kg l^{-1} und einer Viskosität von Wasser. Die Antriebseinheit der Pumpe ist mit einem Frequenzumrichter (VLT[®] 5000, Fa. Danfoss) verbunden, der eine stufenlose Anpassung der Flussrate an die Anforderungen des Filtrationsprozesses erlaubt.

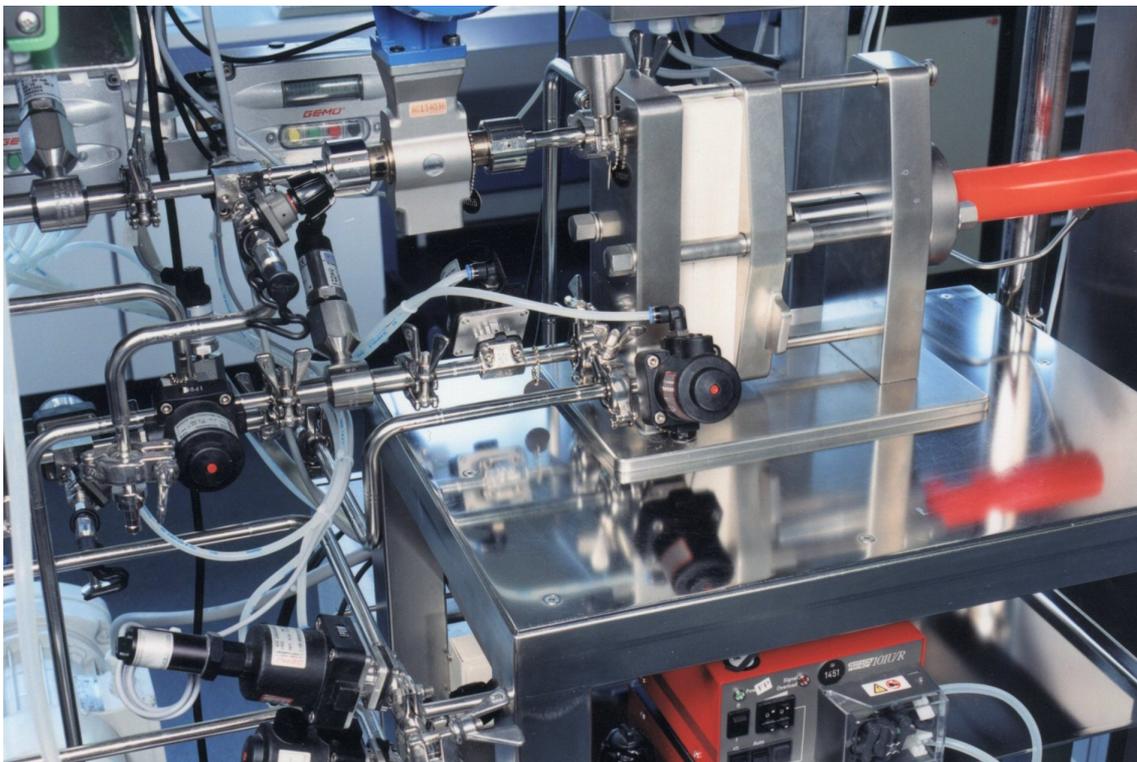


Bild 5.7: Die Crossflow-Filtrationseinheit der BIOSTAT[®] C-Xflow Anlage

Nach Überströmung der Membran gelangt das Retentat über einen Anschluss im oberen Stutzenkranz zurück in den Bioreaktor. Das Permeat wird in einen Produktbehälter geleitet und dort auf 4 °C gekühlt. Der Doppelmantel des Permeatgefäßes ist dafür mit einem Durchlaufkühler (Frigomix[®] R, Fa. Sartorius) verbunden. Sowohl im Retentatweg als auch im Permeatweg sind magnetisch-induktive Durchflussmesser installiert.

Wie bereits beschrieben, werden die Filtrationsbedingungen über drei Druckregelkreise eingestellt. Als Messwertgeber der Regelung sind im Feedweg, im Retentatweg und im Permeatweg Druckmessumformer (Typ C-10, Fa. WIKA) installiert. Die Hubbegrenzung des Retentat- und Permeatventils erfolgt über Linear-Weggeber (Typ 4232, Fa. GEMÜ), die über ein Sollwertsignal der digitalen elektropneumatischen Stellungsregler (Typ ePos 1435, Fa. GEMÜ) angesteuert werden.

Der über ein Hydraulikaggregat betriebene Filterhalter Sartocon Slice (Fa. Sartorius) kann bis zu sechs Mikrofiltrationskassetten aufnehmen und ermöglicht entsprechend dem Anlagenkonzept einen vollautomatischen Betrieb bis 400 bar. Der Anpressdruck der Module wird bei der Filtration auf 130 bar sowie bei der Sterilisation auf 80 bar über die DCU3 geregelt.

Die Filterkassetten (Fa. Sartorius) besitzen eine Hydrosart[®] Membran mit einer effektiven Filterfläche von 0.1 m² pro Filtrationsmodul. Die Filtrationen erfolgten in dieser Arbeit mit zwei Crossflow-Kassetten. Die hydrophil regenerierte Hydrosart[®] Cellulosemembran sollte eine Proteinbindung und damit den unerwünschten Fouling-Effekt verhindern. Verwendet wurden zwei verschiedene Sartocon Slice Module, deren Spezifikationen detailliert in Tabelle 5.1 dargestellt sind

Tabelle 5.1: Spezifikationen der eingesetzten Sartocon Slice Kassetten

Membranmaterial	Porengröße	Filterfläche	pH-Stabilität	max. p _{feed} bei 20°C	max. Dauerbetriebstemp.	Dichtung	Vergussmasse
Hydrosart [®]	0.2 µm	0.1 m ²	2-14	4 bar	50 °C	PVDF	Silikon
Hydrosart [®] (open channel)	0.2 µm	0.1 m ²	2-14	4 bar	50 °C	PVDF/PSU	Silikon

5.2 Automatisierung der Sequentiell Integrierten Bioprozessführung

5.2.1 Umsetzung mit S88-Steuerrezepten und DCU-Sequenzen

Die Umsetzung der Automatisierungsaufgaben für die Sequentiell Integrierte Bioprozessführung erfolgte hauptsächlich mit dem für MFCS/win optional erhältlichen Batch Management Modul S88C zur Anfertigung von prozessspezifischen Rezepten. Darüber hinaus wurden die im DCU3-System dediziert für die Crossflow-Filtrationseinheit hinterlegten Sequenzen *XSTERI*, *XWASH* und *CDFIL* genutzt.

Die ca. 45 minütige Sequenz *XSTERI* erlaubt die Sterilisation der Crossflow-Einheit im laufenden Betrieb. Mit der ca. 75 minütigen Sequenz *XWASH* ist es möglich, während einer Kultivierung das Crossflow-System automatisch mit deionisiertem Wasser zu spülen sowie mit NaOH-Lösung zu reinigen. Die Sequenz *CDFIL* dient der Crossflow-Mikrofiltration.



Bild 5.8 zeigt die Struktur des bereits zum Teil in Kapitel 4.4.3 vorgestellten S88-Rezepts, nun ergänzt um die den Abschnitten der gewählten Prozessführungsstrategie entsprechenden Operationen *CAKE FORM.*, *SEQUENCE1* bis *SEQUENCE13* und *TERMINATION*. In MFCS/win 2.1 können die Sequenzen nicht beliebig oft wiederholt werden, sondern sind vielmehr von Beginn an auf 13 Wiederholungen limitiert.

Seit dem Release der MFCS/win Version 3.0 wurde die Erstellung komplexer Prozessstrategien jedoch wesentlich erleichtert. Mit dieser Version können nicht nur DCU-Sequenzen über sog. Equipment-Phasen angesteuert und somit per S88-Rezept in die Automatisierung eingebunden werden. Des Weiteren besteht mit dieser Version auch erstmalig die Möglichkeit, Schleifen in die Rezeptstruktur einzubinden, die eine Wiederholung von einmalig programmierten Vorgängen erlaubt.

Bild 5.8: Operationspfad des S88-Rezepts

Die Umsetzung der Prozessführungsstrategie mit MFCS/win 3.0 Rezepten sowie eine Optimierung der Filtrationsbedingungen erfolgte im Rahmen einer Folgearbeit (Eicke, 2007). Einerseits bestand die Optimierungsaufgabe darin, die Filtrationsbedingungen mit dem Ziel einer erhöhten Produktausbeute sowie einer verlängerten Membranlebensdauer optimal einzustellen. Andererseits sollte ein Permeatfluxregler erprobt werden, mit dem Ziel die Mikrofiltrationen ständig unterhalb eines kritischen Permeatfluxes J_{percrit} durchzuführen, um so einer unerwünschten Verblockung der Membranporen entgegenzuwirken.

5.2.2 Vorbereitung der Filtration durch Aufbau eines Filterkuchens

Zu Beginn der Kuchenbildung wird mit Hilfe der externen Phase *Cake Formation Init* ein Bedienerhinweis erzeugt, der im Operator Service von MFCS/win erscheint. Der Benutzer wird aufgefordert, die DCU3 Sequenz *CDFIL* zu starten und das Bodenablassventil des Bioreaktors zu öffnen. Die DCU3 Sequenz bleibt bis zum Abbruch der Kultivierung aktiv.



Erst nach Bestätigung der Meldung und wahrer Transition werden in der Start-Phase die notwendigen Druckregelungen mit den Sollwerten p_{feedw} von 1.2 bar und p_{retw} von 0.8 bar sowie die Füllstandsregelung mit einem Sollwert von 12 kg initialisiert.

Die Kuchenbildung dauert 15 Minuten. Nach Ablauf dieser Zeit wird in der End-Phase der Timer deaktiviert. Um die Filtration einzuleiten, wird die Schaltervariable sw_caking in der Termination-Phase auf den Zustand [0] und die Variable $sw_filtration$ auf den Wert [1] gesetzt.

Bild 5.9: Phasen des Operationsblocks *CAKE FORMATION*

Nach dem erstmaligen Anfahren der Mikrofiltration befindet sich noch keine Deckschicht auf der Membranoberfläche der Filtermodule. Für den Aufbau des Filterkuchens wurden daher bei $t = 51.28$ h die notwendigen Druckregelungen initialisiert. Bild 5.10 zeigt die Vorbereitung der ersten Mikrofiltration durch Kuchenbildung.

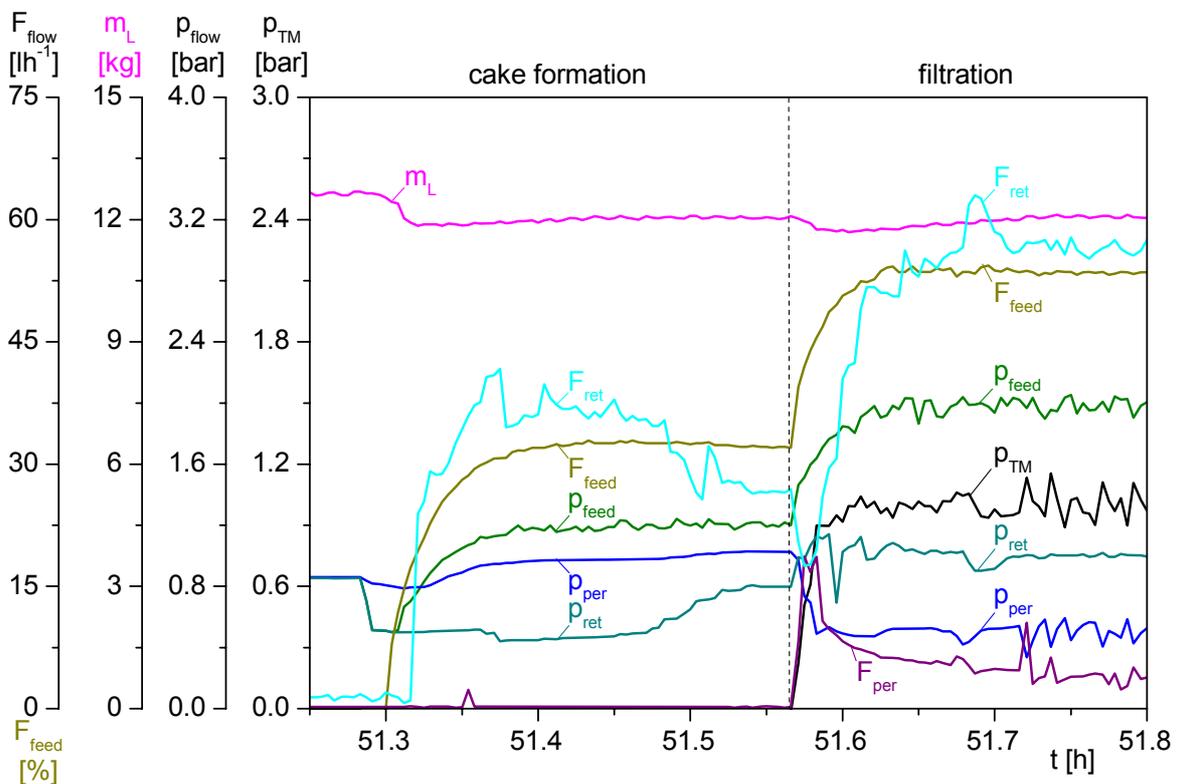


Bild 5.10: Vorbereitung der ersten Mikrofiltration durch Kuchenbildung

Die Crossflow-Filtrationseinheit wurde nach dem Einsetzen der Filtrationsmodule zu Beginn der Kultivierung sterilisiert. Durch diesen Vorgang befand ein Restdruck von 0.86 bar in den Rohrleitungssystemen der Anlage. Das erklärt den Druckverlauf am Ein- und Ausgang der Filtermodule vor Einstellung der eigentlichen Sollwerte und den Verlauf des permeatseitigen Drucks p_{per} vor und während der Kuchenbildung.

Die Masse des Reaktionsvolumens m_L verringerte sich zu Beginn des Vorgangs um 0.75 kg, da die Rohrleitungen der Crossflow-Einheit noch nicht gefüllt waren.

Das Hochlaufen der Zufütterrate F_{feed} folgt einem Verzögerungsverhalten 1. Ordnung. Der Druckverlauf p_{feed} am Eingang des Filtermoduls folgt leicht verzögert dieser Dynamik, bis der Sollwert p_{feedw} von 1.2 bar nach ca. fünf Minuten bei einer Förderleistung der Pumpe F_{feed} von 32.5 % erreicht wurde. Der Sollwert für den Druck am Filterausgang p_{retw} von 0.8 bar wurde erst zum Schluss der Kuchenbildung erlangt. Der Aufbau des Filterkuchens wurde nach 15 Minuten automatisch beendet.

5.2.3 Durchführung der Mikrofiltration

In Bild 5.11 sind die Phasen der identischen Operationsblöcke *SEQUENCE1* bis *SEQUENCE13* zu sehen. Es gibt vier parallel ablaufende Prozesspfade, die jeweils sich zyklisch wiederholende Hauptphasen der SIB codieren.

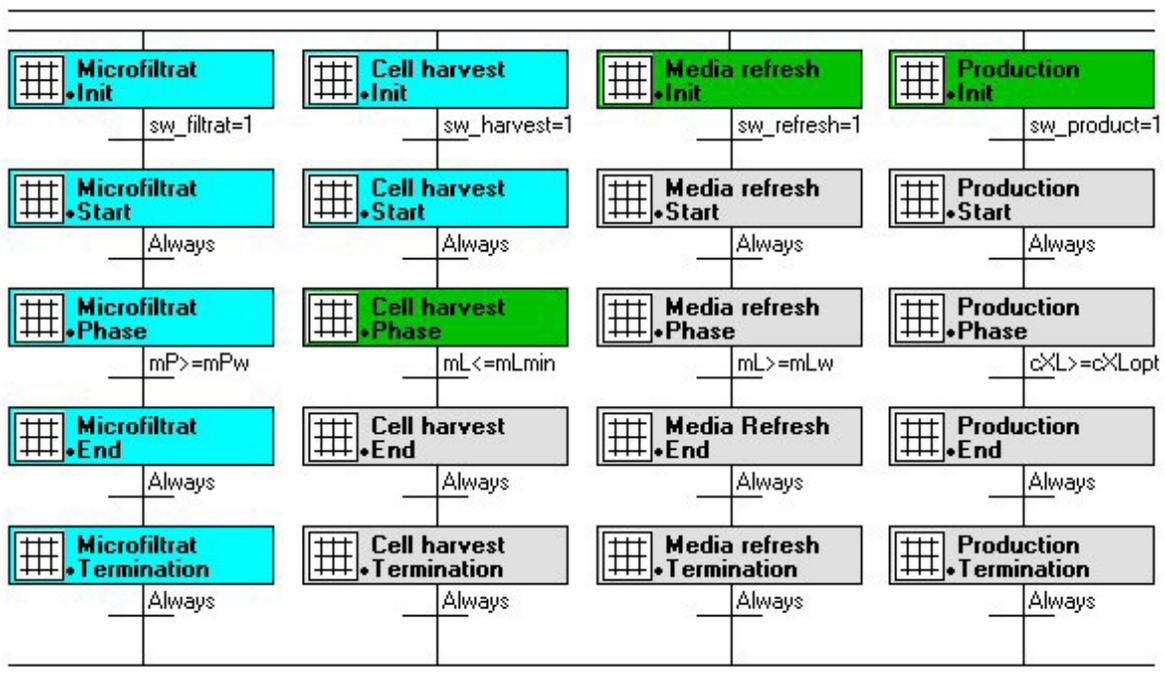


Bild 5.11: Phasen der identischen Operationsblöcke *SEQUENCE1* bis *SEQUENCE13*

Die Init-Phasen der Operationsblöcke enthalten keine Befehle, sondern dienen wie zuvor beschrieben nur der Initialisierung durch Zustandsprüfung der jeweiligen Schalteryvariablen $sw_filtration$, $sw_harvest$, $sw_refresh$ und $sw_production$. Der Zustand dieser Variablen wird in der Termination-Phase des jeweils vorhergehenden Prozessabschnitts gesetzt.

In der Phase *Microfiltration Start* werden die zur Filtration notwendigen Druckregelungen mit $p_{feedw} = 2$ bar, $p_{retw} = 1$ bar und $p_{perw} = 0.5$ bar gestartet. Außerdem wird die Füllstandsregelung mit einem Sollwert von 12 kg initialisiert. Sobald die Masse des Permeattanks den für einen Massenaustausch notwendigen Sollwert m_{pw} von 12 kg erreicht hat, wird die Filtration durch Schließen des Permeatventils in der Phase *Microfiltration End* beendet. Dafür wird der Druckregler auf der Permeatseite in die Stellung MAN geschaltet und die Stellgröße y_{per_SP} mit 0 % direkt vorgegeben. Außerdem werden neue Sollwerte für die Drücke $p_{feedw} = 1.2$ bar und $p_{retw} = 0.8$ bar eingestellt und die Füllstandsregelung durch Begrenzung des Reglerausgangs mit $F_{R3rel_SP} = 0$ % deaktiviert.

Mit einem Sollwert F_{Hrelw} von 50 % wird die Zellernte in der Phase *Cell harvest Start* initialisiert. Sobald die Masse des Flüssigphasenvolumens 9 kg erreicht hat, wird die Übergangsbedingung $m_L \leq m_{Lmin}$ der *Cell harvest Phase* wahr. Die Zellernte wird in der nächsten Phase *Cell harvest End* durch Vorgabe des Sollwerts F_{Hrelw} von 0 % beendet.

Für den Medienrefresh wird die Stellgröße des Füllstandsreglers F_{R3rel_SP} mit 50 % in der Phase *Media refresh Start* direkt vorgegeben – die Pumpe fördert mit halbmaximaler Leistung. Sobald die ursprüngliche Masse des Reaktionsvolumens von 12 kg wieder erreicht ist, wird der Refresh in der End-Phase durch Begrenzung des Reglerausgangs mit $F_{R3rel_SP} = 0$ % beendet.

Die Phasen *Production Start* und *Production End* enthalten keine Befehle, lediglich die Parameter-Phase *Production Phase* enthält eine Transition. Ist die Übergangsbedingung bei einer Zelldichte c_{XLOpt} von 40 gl^{-1} erfüllt, findet eine Weiterschaltung in die nächste Sequenz statt.

Nach Ablauf der Kuchenbildung wurden die Filtrationsbedingungen eingestellt. Der Sollwert für den Eingangsdruck p_{feedw} wurde von 1.2 bar auf 2.0 bar und der Sollwert für den Ausgangsdruck p_{retw} von 0.8 bar auf 1.0 bar erhöht. Der Sollwert für den permeatseitigen Druck p_{perw} wurde mit 0.5 bar vorgegeben. In Bild 5.12 ist die Regelung der drei Drücke dargestellt. Der Transmembrandruck p_{TM} verlief durchschnittlich bei 1.0 bar.

Während der Filtration lag die durchschnittliche Förderrate F_{feed} der Feedpumpe bei 53.2 %. Der Retentatstrom F_{ret} verlief zunächst auf einem annähernd gleichbleibenden Niveau von 56.3 lh^{-1} . Mit zunehmender Filtrationsdauer nahm der Retentatstrom jedoch um 14.5 lh^{-1} bis auf 41.8 lh^{-1} ab. Der Permeatstrom F_{per} sank ebenfalls mit zunehmender Dauer der Filtration. Im unteren Messbereich des Durchflussmessers kam es zu einer Unterbrechung der Datenaufnahme.

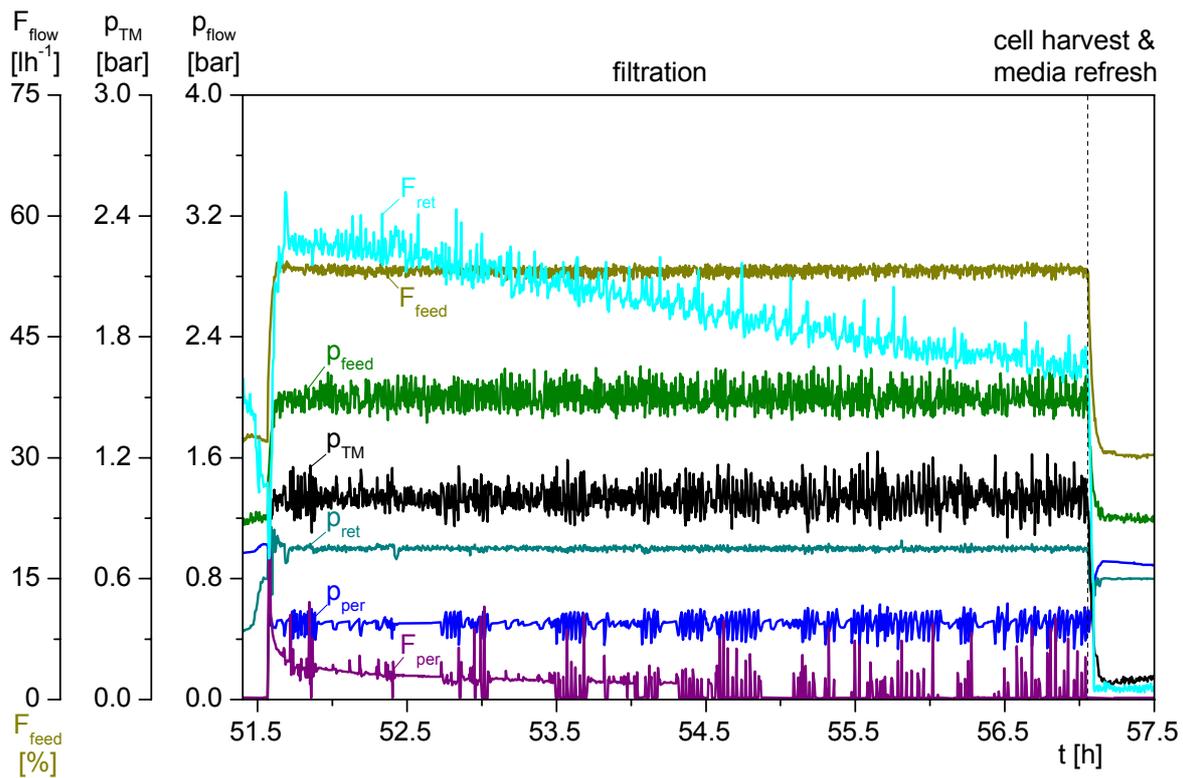


Bild 5.12: Ströme und Druckverhältnisse während einer Mikrofiltration

Wie in Bild 5.13 gezeigt, wurde der Füllstand für die Dauer der Mikrofiltration geregelt und das abgeführte Permeat durch frisches Medium ersetzt.

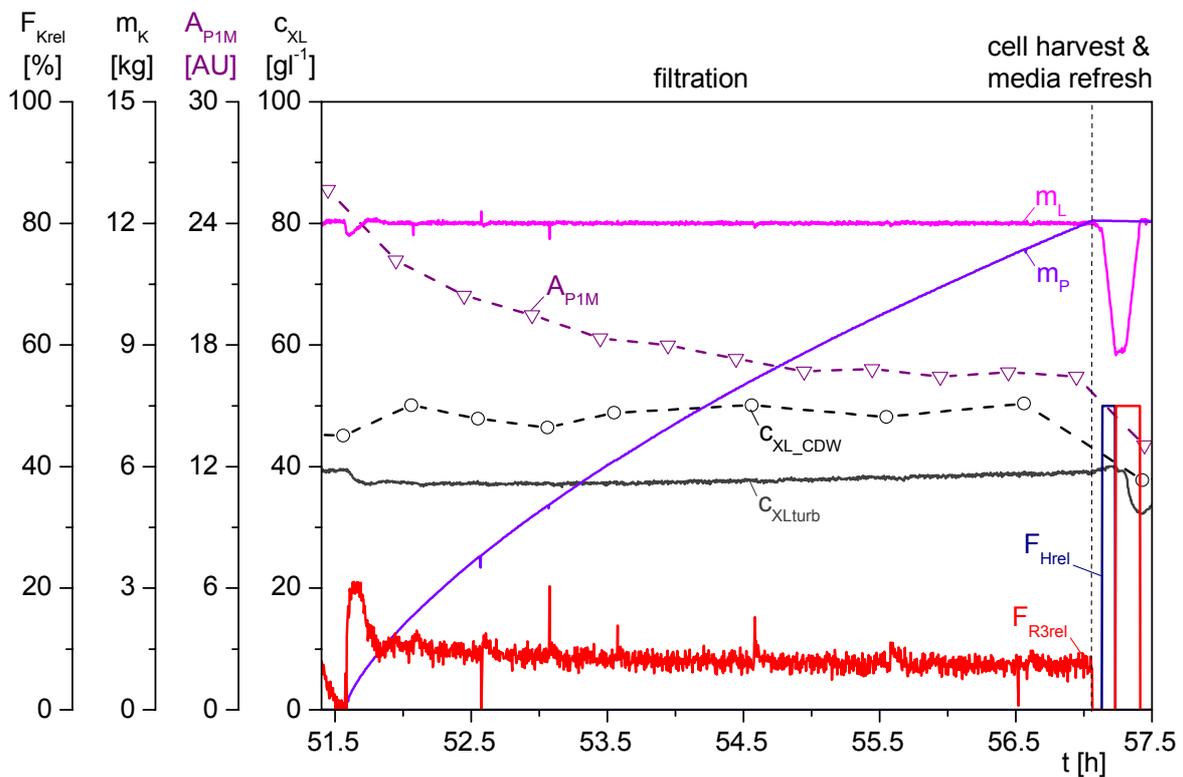


Bild 5.13: Ablauf einer Filtration mit anschließender Konditionierungsphase

Trotz der Verwendung von zwei gereinigten open channel Filterkassetten dauerte die erste Mikrofiltration mit 5.5 h sehr lange. In diesem Zeitraum war nur ein geringes Zellwachstum zu beobachten. Die Produktabsorption im Medium A_{P1M} nahm um 9.4 AU ab. Nach einem Massenaustausch wurde die Filtration automatisch beendet.

5.2.4 Vorbereitung neuer Sequenzen durch Zellernte und Medienrefresh

Wie in Bild 5.14 dargestellt, wurden nach der Mikrofiltration wieder die gleichen Drucksollwerte eingestellt, die schon bei der Kuchenbildung vorlagen. Anschließend fand eine Konditionierung der Kulturbrühe durch Zellernte und Medienrefresh statt.

Dafür wurde zunächst zell- und produktthaltiges Medium mit einer Harvestrate F_{Hrel} von 50 % abgepumpt. Im Anschluss wurde mit einer Zufüterraterate F_{R3rel} von 50 % methanolhaltiges ($C_{S2R3} = C_{S2Mw}$) Frischmedium aufgefüllt und schließlich eine neue Produktionsphase eingeleitet. Die Zelldichte $c_{XL_{turb}}$ sank durch den Verdünnungsschritt von 40.0 gl^{-1} auf 32.3 gl^{-1} , die Zelldichte $c_{XL_{CDW}}$ von 50.4 gl^{-1} theoriekonform auf 37.8 gl^{-1} .

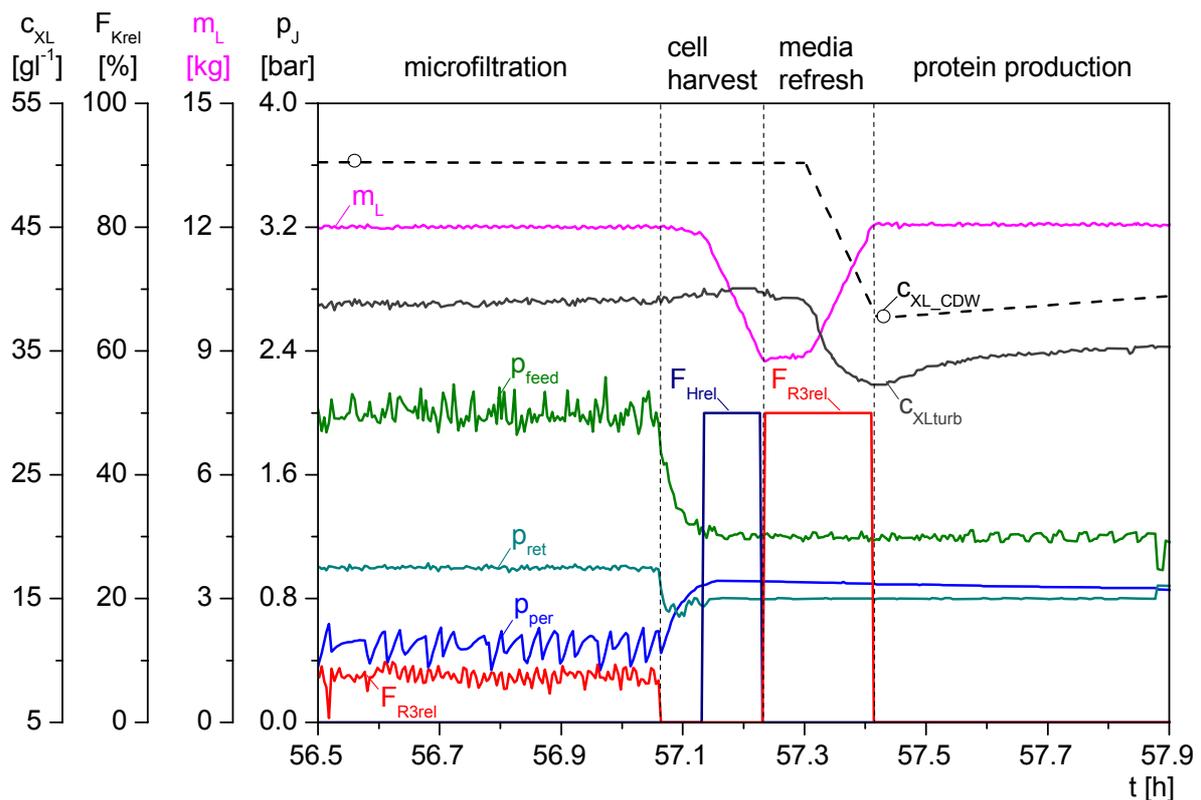


Bild 5.14: Konditionierung der Kulturbrühe nach der 1. Mikrofiltration

Die Zellernte und der Medienrefresh erfolgten nach der in Kapitel 5.1.2 vorgestellten Strategie vollautomatisch. Aufgrund einer hohen Systemauslastung des Prozessleitrechners, kam es

ab $t = 57.08$ h bei beiden Konditionierungsschritten zu einer Verzögerung bei der Ausführung der S88-Befehle. Das Problem konnte bei dieser Kultivierung mehrmals beobachtet werden.

Bild 5.15 zeigt die Konditionierung der Kulturbrühe nach der 6. Mikrofiltration, die fehlerfrei verlief.

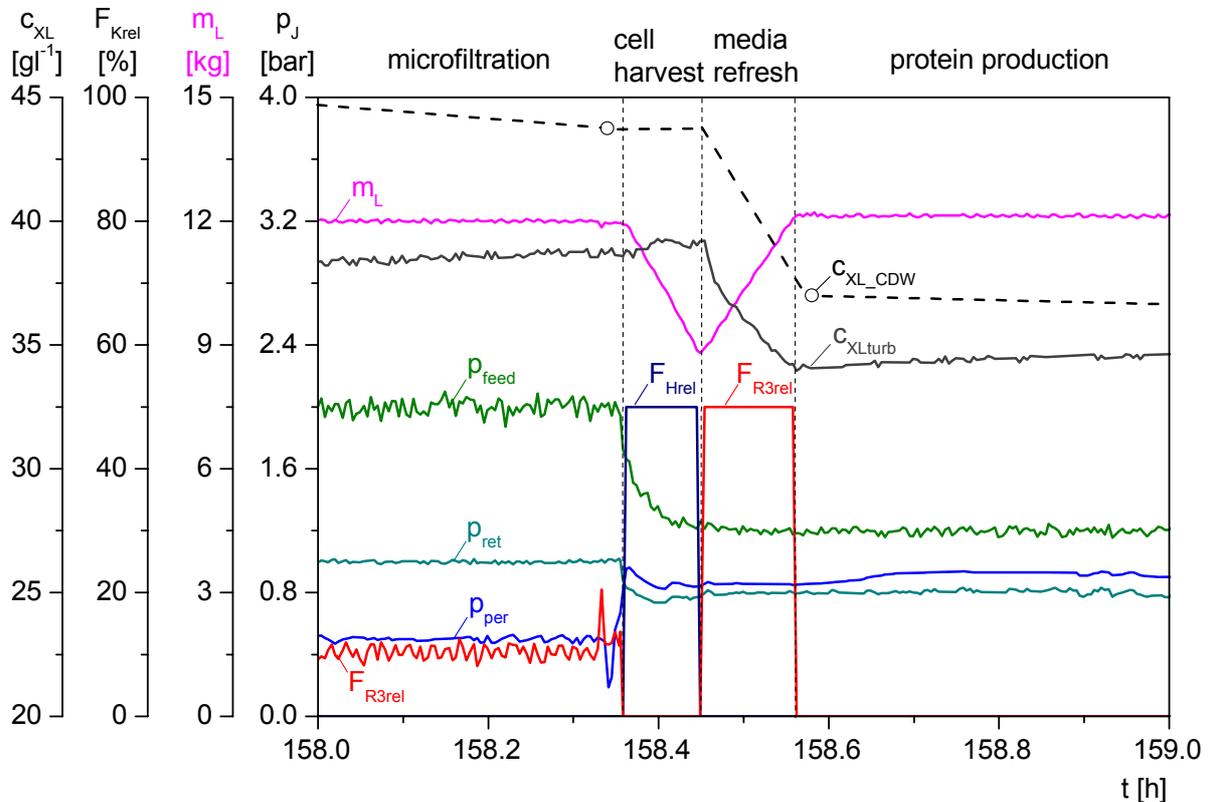


Bild 5.15: Konditionierung der Kulturbrühe nach der 6. Mikrofiltration

5.2.5 Fouling an der Crossflow-Membran

In Bild 5.16 sind vier Sequenzen einer Sequentiell Integrierten Prozessführung dargestellt. Die Berechnung der Zelldichte $c_{XL_{turb}}$ zeigt zu Beginn der Filtrationen MF7, MF8 und MF11 einen ähnlichen Verlauf wie bei den im vorhergehenden Kapitel 5.1.2 beschriebenen Filtrationen MF2 bis MF4. Eine Ausnahme bilden hier MF9 und MF10.

Anhand des Permeatstroms F_{per} ist erkennbar, dass mit zunehmender Kultivierungsdauer die Filtrationsleistung durch Fouling der verwendeten Membranen abnahm. Mögliche Ursachen für Fouling sind Ablagerung sehr feiner Partikel auf der Membran, Adsorption von makromolekularen bzw. kolloidalen Stoffen, Adhäsion und Wachstum von Mikroorganismen (biological fouling), Ausfällungen oder Alterung der Membran (Ripperger, 1993).

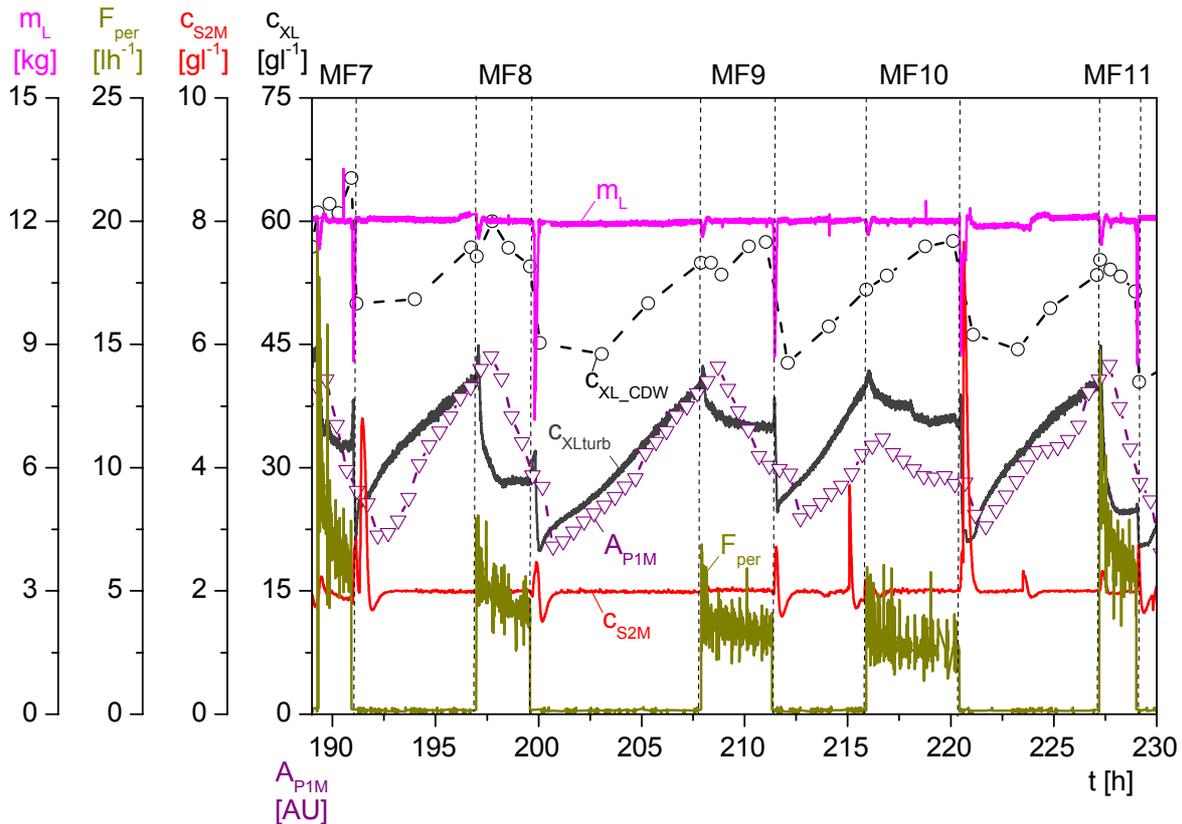


Bild 5.16: Fouling an den Crossflow-Membranen

Tabelle 5.2 zeigt eine Übersicht der Filtrationen MF1 bis MF14 einer Kultivierung. Für jede durchgeführte Mikrofiltration wurde der Sollwert für den Eingangsdruck p_{feedw} von 1.2 bar auf 2.0 bar und der Sollwert für den Ausgangsdruck p_{retw} von 0.8 bar auf 1.0 bar erhöht. Der Sollwert für den permeatseitigen Druck p_{perw} wurde jeweils mit 0.5 bar vorgegeben.

Durch das Fouling der zwei Crossflow-Membranen (open channel, 0.2 m^2 Filterfläche gesamt) verlängerte sich die Dauer der durchgeführten Mikrofiltrationen signifikant. Die in Tabelle 5.2 dargestellte Biotrockenmassedifferenz $\Delta c_{\text{XL_CDW}}$ ist aus der Zelldichte zum Ende abzüglich der Zelldichte zu Beginn einer Filtration berechnet.

Vor der vierten, siebten und der elften Mikrofiltration wurde die Sequenz *CDFIL* an der DCU3 gestoppt. Die Filterkassetten wurden demontiert und mit 1 mol l^{-1} NaOH bei $40 \text{ }^\circ\text{C}$ in einer speziellen Apparatur gereinigt. Währenddessen wurde das Crossflow-System mit Hilfe der DCU3-Sequenz *CRINS* unter Verwendung einer Ersatzmembran mit VE-Wasser gespült. Nach der Reinigung wurden die Filterkassetten wieder in den Filterhalter eingesetzt.

Die Crossflow-Einheit wurde nun mit Hilfe die DCU3-Sequenz *CROSS* sterilisiert. Eine automatische Reinigung ist bisher für die gesamte Anlage nach dem Prozess in der DCU3 integriert. Um das beschriebene membrane fouling zu vermeiden, soll zukünftig ein separates CIP-System für die Crossflow-Einheit etabliert werden.

Tabelle 5.2: Dauer, durchschnittlicher Permeatstrom und Biotrockenmassedifferenz von 14 durchgeführten Mikrofiltrationen MF1 bis MF14 einer Kultivierung

MF	Dauer [h]	\bar{F}_{per} [lh^{-1}]	$\Delta c_{\text{XL_CDW}}$ [gl^{-1}]	Bemerkungen
1	1.53	7.86	-1.95	
2	4.01	3.00	0.40	
3	6.54	1.84	-3.60	
4	2.05	5.85	2.25	vorher Filtrationsmodule gereinigt
5	2.96	4.06	8.25	
6	4.70	2.56	11.75	
7	1.63	7.38	4.20	vorher Filtrationsmodule gereinigt
8	2.62	4.60	-1.20	
9	3.44	3.50	2.50	
10	4.52	2.66	5.95	
11	1.79	6.75	-3.85	vorher Filtrationsmodule gereinigt
12	2.36	5.09	-4.65	
13	3.03	3.97	6.75	
14	6.60	1.82	11.55	doppelter Massenaustausch

6 Grundlagen der Prozessoptimierung mit Statistischer Versuchsplanung

6.1 Der dreiphasige Produktionsablauf und dessen Optimierungspotential

6.1.1 Klassischer Prozessablauf einer HCDC mit *E. coli*

Zusätzlich zur Inclusion Body Problematik führt die Überexpression plasmidcodierter, heterologer Proteine durch *E. coli* zu einer starken Beeinträchtigung der Zellphysiologie. Diese Beeinträchtigung kann soweit führen, dass die Zelle ihren Metabolismus umstellt oder sogar abstirbt (Glick, 1995; Hoffmann and Rinas, 2004).

Die Auswahl einer geeigneten, stressminimierenden Kultivierungsstrategie ist daher ein entscheidender Gesichtspunkt im biotechnologischen Upstream-Prozess. Wie bereits in Kapitel 4 beschrieben, hat sich unter der Vielzahl der entwickelten Techniken die Kultivierung von *E. coli* zu hohen Zelldichten im substratlimitierenden Fed-Batch als robustes und effizientes Verfahren erwiesen (Lee, 1996; Riesenberg and Guthke, 1999).

Bild 6.1 zeigt exemplarisch den klassischen Ablauf einer Kultivierung von *E. coli*. Jede durchgeführte Kultivierung begann mit der Inokulation der Bioreaktoren bei einer vorgelegten Glukosekonzentration von 20 g l^{-1} . Um reproduzierbare Animpfbedingungen zu schaffen, wurde die Vorkultur immer bis zu einer optischen Dichte OD_{600} von 3 bis 4 AU inkubiert. Anschließend wurde die gesamte Vorkultur aliquotiert und jeder der sechs Reaktoren mit 100 ml Zellsuspension angeimpft. Dadurch ergaben sich bei einem Startvolumen V_{L0} von 0.7 l Anfangszelldichten von 0.21 - 0.27 g l^{-1} .

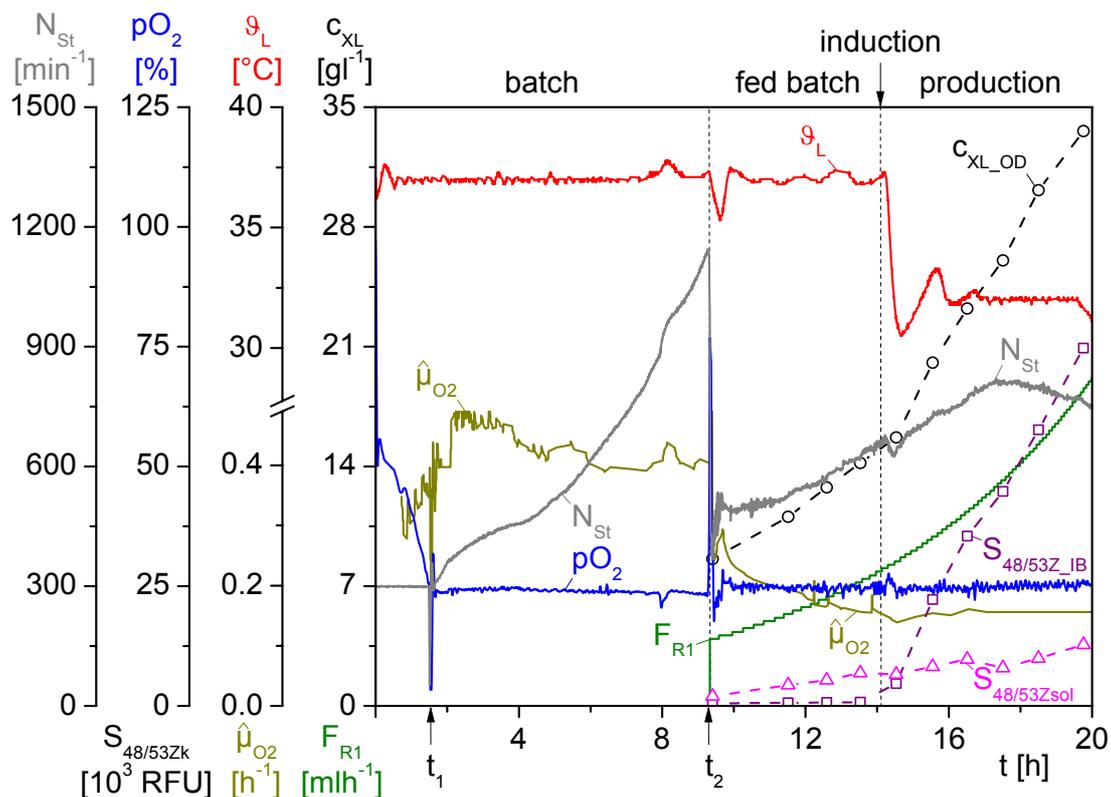


Bild 6.1: Klassischer Verlauf einer Kultivierung von *E. coli*

Für die Vorkultur im Schüttelkolben und die Hauptkultur im Bioreaktor wurde ein modifiziertes M9 Minimalmedium verwendet (Riesenberg *et al.*, 1991). Tetrazyklin wurde dem Medium in einer Konzentration von 50 mg l^{-1} zugesetzt.

Die Zellen wuchsen im Batch unlimitiert mit einer maximalen zellspezifischen Wachstumsrate $\mu_{1\max}$ von ca. 0.43 h^{-1} auf dem Hauptsubstrat Glukose. Im substratlimitierten Fed-Batch wurde ein Sollwert für die zellspezifische Wachstumsrate ($\mu_{1w} < \mu_{1\max}$) von 0.15 h^{-1} über eine exponentielle Glukosezufütterung eingestellt. Diese Fahrweise resultierte in einer Zelldichte $c_{\text{XL_OD}}$ von 32 gl^{-1} , die über eine Messung der optischen Dichte OD bei 600 nm berechnet wurde.

Der Sauerstoffpartialdruck p_{O_2} wurde nach Abfall auf seinen Sollwert bei t_1 im weiteren Verlauf der Kultivierung über eine p_{O_2} /Drehzahlregelung auf 25 % geregelt. Grundsätzlich wurde bei allen durchgeführten Kultivierungen bereits zum Ende des Batches die maximale Rührerdrehzahl $N_{\text{S}_{\max}}$ von 1200 min^{-1} erreicht. Daher wurde zu Beginn der Fed Batch-Phase bei t_2 die anfängliche Luftbegasung durch eine AIR/O₂-Mischbegasung mit einem vorgegebenen Stoffmengenanteil von Sauerstoff am Gaseintritt x_{OGin} von 0.45 ersetzt.

Die Flüssigphasentemperatur ϑ_{L} wurde im Batch und Fed Batch zunächst auf $37 \text{ }^\circ\text{C}$ geregelt. In der Produktionsphase wurde ein anderer, durch den DoE-Versuchsplan definierter, Temperatursollwert per Hand eingestellt. Die Proteinexpression wurde bei einer Zelldichte von 15 gl^{-1} durch die Zugabe von $0.8 \cdot 10^{-3} \text{ moll}^{-1}$ IPTG induziert. Gleichzeitig wurde die Flüssigphasentemperatur ϑ_{L} auf einen Wert von $32 \text{ }^\circ\text{C}$ gesenkt.

Das Fluoreszenzsignal für die lösliche Proteinexpression $S_{48/53Z_{\text{sol}}}$ stieg nach der Induktion auf einen Wert von $13 \cdot 10^3$ RFU. Im Vergleich dazu erreichte das Signal für die unlösliche Proteinexpression $S_{48/53Z_{\text{IB}}}$ mit $75 \cdot 10^3$ RFU einen ca. sechs mal höheren Wert zum Ende der Produktionsphase. Das bereits vor der Induktion vorhandene lösliche Zielprotein weist auf eine nicht vollständige Repression des chromosomalen *lacUV5*-Promotors sowie der plasmidkodierten T7-Promotoren hin, so dass bereits im Batch bzw. Fed-Batch Zielprotein-mRNA durch die T7-RNA-Polymerase transkribiert wurde.

6.1.2 Der modulare Aufbau des Multi-Bioreaktorsystems BIOSTAT® Qplus

Aufgrund der großen Arbeitsvolumina der letztendlichen Produktionsprozesse sowie der personal- und zeitintensiven Umsetzung sequentieller Kultivierungen ist eine Rückskalierung zur Prozessoptimierung unabdingbar. Der technische Ansatz dieses Problem zu lösen, ist das parallele Setup von mehreren Laborreaktoren unter Aufrechterhaltung der Konfiguration eines einzelnen Produktionsreaktors (Schmidt, 2005; Li *et al.*, 2006).

Bild 6.2 zeigt das für die Kultivierung von *E. coli* eingesetzte Multi-Bioreaktorsystem BIOSTAT® Qplus-6 der Fa. Sartorius. Das Bioreaktorsystem besteht aus sechs autoklavierbaren 1-L Rührkesselreaktoren, zwei Versorgungseinheiten mit Vorlagenstationen, der zugehö-

rigen Vorortautomatisierung DCU4 und dem übergeordneten SCADA-System MFCS/win. Neben der werkseitigen Standardinstrumentierung mit Sonden zur Messung von pH, Gelöstsauerstoffpartialdruck pO_2 , Temperatur und Schaum ist jeder Bioreaktor zusätzlich mit einem O_2 - (Typ BCP-O2) und CO_2 -Sensor (Typ BCP-CO2) der Fa. BlueSens, für die Abgasanalyse ausgestattet. Die Bereitstellung der Prozessgase Luft, Sauerstoff und Stickstoff erfolgt über zwei externe Gasmischstationen mit eingebauten Mass-Flow-Controllern (Fa. Bürkert bzw. Fa. Bronkhorst High-Tech, Niederlande) durch Begasungsringe am Boden der Bioreaktoren.

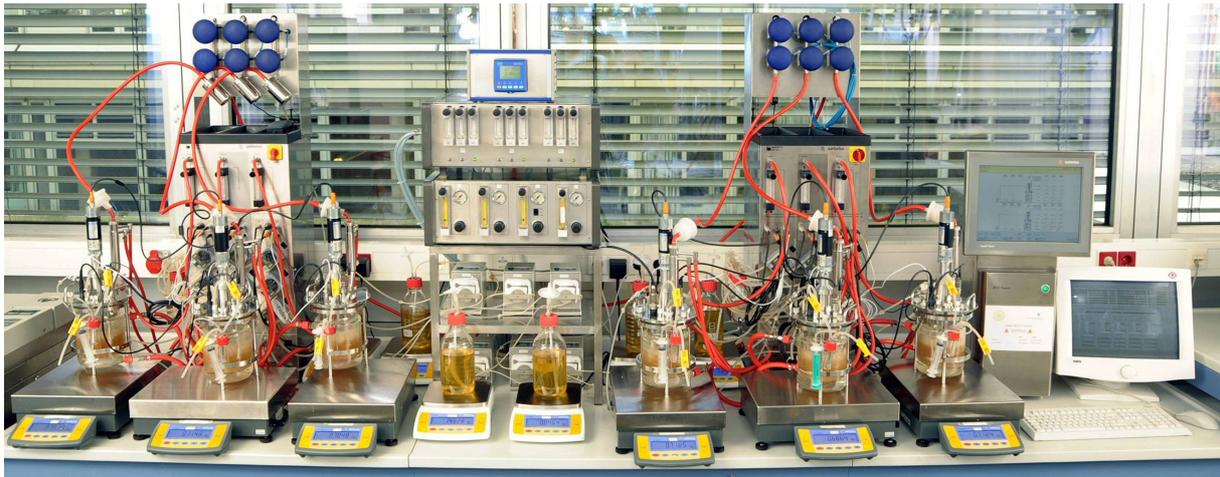


Bild 6.2: Das Multi-Bioreaktorsystem BIostat[®] Qplus mit erweiterter Reaktorperipherie

Ein substratlimitierter Fed Batch-Betrieb wird mit der Integration von stetig ansteuerbaren Schlauchpumpen (Typ Reglo Digital MS-2/12) der Fa. ISMATEC, Schweiz, mit einer maximalen Zufürrate von 0.33 lh^{-1} erreicht. Die Anbindung von insgesamt 12 Waagen (verschiedene CP-Modelle) der Fa. Sartorius ermöglicht eine Erfassung der Reservoir- und Reaktortolumina.

Das DCU4-System erlaubt die unabhängige Grundsteuerung der 6 Bioreaktoren. Im Zentrum der DCU4 steht die eingebaute SIMATIC Microbox (Typ PC 420) der Fa. Siemens, die das System entsprechend der Konfiguration steuert. Die Firmware und die Konfiguration sind gemeinsam auf einer CF-Speicherkarte hinterlegt, die über einen Steckplatz auf der Microbox installiert ist.

Die Microbox ist mit einer seriellen RS-232-, vier USB- und zwei RJ-45 Ethernet-Schnittstellen ausgestattet. Über diese Schnittstellen erfolgt eine Abfrage zum Hauptschalter (power failure signal), die Kommunikation mit dem Touch-Display und den Versorgungseinheiten sowie die Datenübertragung zu dem in das Automatisierungssystem eingebundenen MFCS/win Server.

6.1.3 Anbindung der Analysen- und Automatisierungskomponenten an MFCS/win

Die Integration der verschiedenen Analysen- und Automatisierungskomponenten in die Datenerfassung mit MFCS/win erfordert unterschiedliche Lösungen. Bild 6.3 zeigt die Verknüpfung der eingebunden Systeme PC-intern über OPC, sowie PC-extern über serielle Protokolle bzw. Ethernet.

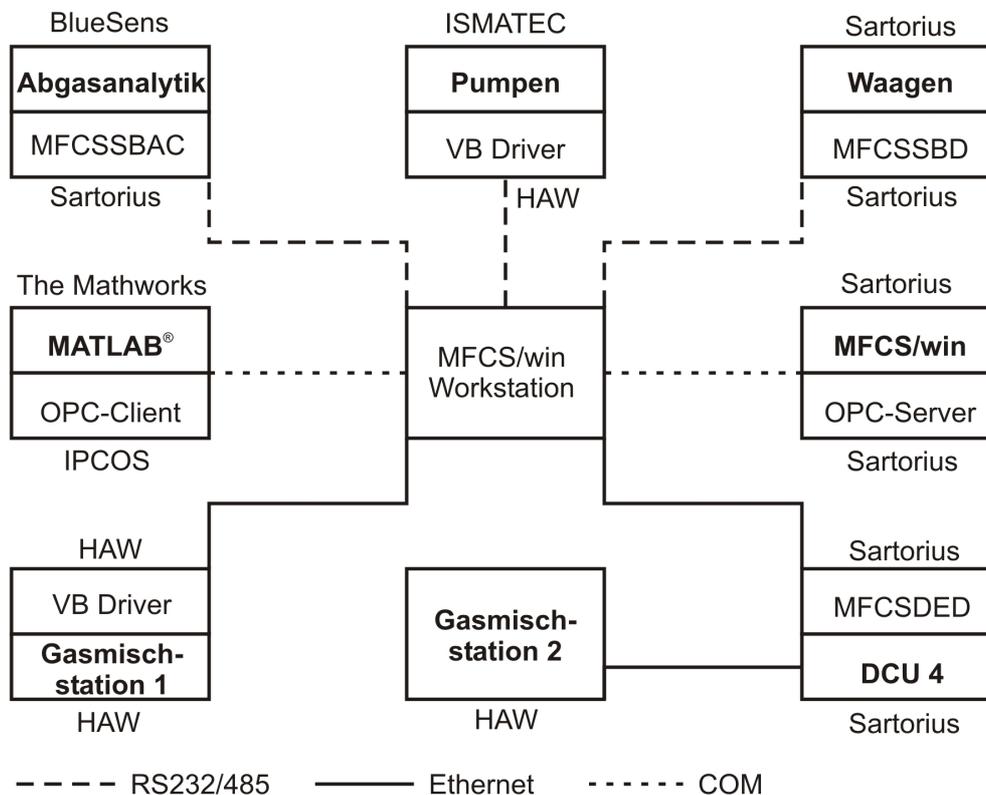


Bild 6.3: Anbindung der Analysen- und Automatisierungskomponenten an MFCS/win

Wie in Kapitel 3.2.2 beschrieben, erfolgt die Kommunikation der für die Abgasanalytik eingesetzten O₂- und CO₂-Sensoren mit MFCS/win über den PLS-internen BACCom-Treiber MFCSSBAC. Die Anbindung des MATLAB®-Tools MAC4MFCS über eine OPC Server/Client-Schnittstelle ist in Kapitel 3.3.2 beschreiben.

Aufgrund der begrenzten seriellen Schnittstellen sind die für die Wägung der Reaktoren und der Feedvorlagen eingesetzten Waagen jeweils über ein RS-232-Kabel mit nachfolgendem RS-232/USB-Konverter über nur einen USB-Port an den Prozessleitreechner angeschlossen. Für die Kommunikation mit MFCS/win wird der optional erhältliche Serial Balance Driver (MFCSSBD) eingesetzt.

Die Pumpen für den Fed-Batch Betrieb sind jeweils mit einer seriellen RS-232-Schnittstelle ausgestattet. In Reihe geschaltet, können sie mit einem einzigen COM-Port des Prozessleitreechners verbunden werden. Die Treiber für die Kommunikation der Pumpen aber auch der Gasmischstation 1 mit MFCS/win wurden an der HAW Hamburg im Forschungs- und Transferzentrum Bioprozess- und Analysetechnik mit VisualBasic® entwickelt.

Die in der Gasmischstation 1 verbauten Mass-Flow-Controller der Fa. Bürkert sind über ihren proprietären RS-485 Bus an einen Seriell-Ethernet Device Server (Typ IOLAN SDS4) der Fa. Perle Systems, USA, angebunden. Der Device Server bietet insgesamt vier konfigurierbare Anschlüsse, von denen einer für die Datenübertragung zum Prozessleitrechner per RJ-45 Ethernetkabel genutzt wird. Die drei anderen Anschlüsse bieten zusätzliche RS-232 Schnittstellen auf der Rückseite der Gasmischstation. Das Betriebssystem des Prozessleitrechners registriert die Schnittstellen des Device Servers als zusätzliche COM-Ports, womit sie für die Kommunikation über den VisualBasic[®] Treiber und somit die Bedienung über MFCS/win zugänglich werden.

Die Signale der analog anzusteuernenden Mass-Flow-Controller der Fa. Bronkhorst High-Tech, Niederlande, in der Gasmischstation 2 werden über einen Ethernet Feldbus-Koppler (Typ I/O System 750) der Fa. WAGO Kontakttechnik in ein digitales Signal umgewandelt. Die Datenübertragung erfolgt im Gegensatz zur Gasmischstation 1 über ein RJ-45 Ethernetkabel direkt an die SIMATIC Microbox der DCU4. Damit lassen sich die Mass-Flow-Controller lokal von der DCU4 oder im Remote-Betrieb über MFCS/win bedienen. Die DCU4 kommuniziert mit MFCS/win über den DCU-Ethernet-Driver (MFCSDDED).

6.1.4 Einfluss- und Zielgrößen für eine optimierte Proteinproduktion

Die Optimierung biotechnologischer Herstellungsprozesse geht einher mit der Suche nach optimalen Expressionsbedingungen, die in einer hohen Raum-Zeit-Ausbeute (RZA) sowie einer guten Produktqualität resultieren. Im biopharmazeutischen Forschungsbereich wird die Erhöhung der Gesamtausbeute physiologisch aktiver Proteine in der löslichen Fraktion angestrebt, um das aufgereinigte Protein nach dem Herstellungsprozess weiteren Analysen, die aktive Proteinproben erfordern, zeitnah bereitstellen zu können.

Dazu zählen z. B. ein sekundäres Screening mit biochemischen Aktivitätsassays oder Interaktionsanalysen, eine Strukturanalyse durch röntgenkristallografische Methoden oder Kernspinresonanzspektroskopie (NMR¹) oder bereits eine DMPK²-Studie, bei der sämtliche Eigenschaften zu Absorption, Verteilung, Metabolismus und Ausscheidung des potentiellen Biopharmazeutikums untersucht werden.

Traditionell wird versucht, die zellspezifische Produktbildungsrate $q_{p1/X}$ herabzusetzen, um eine korrekte Faltung des Zielproteins durch rechtzeitige Interaktion mit einer Klasse von Hilfsproteinen (Chaperonen) zu ermöglichen (Baneyx and Mujacic, 2004). Molekulare Chaperone unterstützen die Proteinfaltung und den Zusammenbau von Proteinkomplexen, indem sie instabile Faltungsintermediate anhand exponierter hydrophober Bereiche erkennen und deren unkontrollierte Aggregation unterbinden (Polier and Bracher, 2009).

¹ NMR - Nuclear Magnetic Resonance Spectroscopy

² DMPK - Drug Metabolism / Pharmacokinetics

Eine Verringerung der unkontrollierten Aggregation sowie eine Herabsetzung des proteolytischen Abbaus rekombinanter Proteine kann durch eine Vielzahl von Strategien erreicht werden (Übersicht in Sørensen and Mortensen, 2005 sowie Sahdev *et al.*, 2008). Ausschlaggebender Faktor ist bei diesen Strategien immer die Physiologie des Expressionswirtes, die einerseits durch molekularbiologische Methoden im Vorfeld, oder andererseits durch verfahrenstechnische Eingriffe während des Prozesses manipuliert werden kann.

Höhere Wachstumsraten können z. B. während der Produktion oder selbst vor der Induktion die rekombinante Proteinproduktion erhöhen (wahrscheinlich durch einen höheren Ribosomengehalt), andererseits den zellulären Plasmidgehalt senken. Daher hängt es vom untersuchten System ab, bei welcher Wachstumsrate die Produktbildungsrate ein Maximum erreicht (Seo and Bailey, 1986; Hoffmann *et al.*, 2004).

Durch Kultivierung bei niedrigen Temperaturen werden einerseits die Transkriptions- und Translationsrate sowie andererseits die Stärke hydrophober Interaktionen, die zur Proteinfaltung beitragen, reduziert (Schein, 1989). Des Weiteren wurden bei niedrigen Temperaturen eine erhöhte Expression von Chaperonen (Mogk *et al.*, 2002; Ferrer *et al.*, 2003) sowie eine geringere Proteaseaktivität festgestellt (Hunke and Betton, 2003; Spiess *et al.*, 1999).

Hohe IPTG-Konzentrationen führen zu einer vollständigen Induktion des Lac-Promotors, jedoch aufgrund der zusätzlichen „metabolischen Last“ (engl.: metabolic burden) für die Erhaltung und Expression der Fremd-DNA nicht notwendigerweise zu einer maximalen Expression des Zielproteins (Glick, 1995). Durch eine partielle Titration der vorhandenen Repressormoleküle, d. h. durch die Anwendung niedriger IPTG-Konzentrationen, kann die Transkription der T7-RNA-Polymerase und somit auch die des plasmidkodierte Zielgens verringert werden.

In dieser Arbeit wurden die drei zuvor genannten gängigsten verfahrenstechnischen Strategien angewendet, die einen Einfluss auf die Syntheserate des Zielproteins nehmen können. Bild 6.4 zeigt einen Vergleich zweier Produktionsphasen, in denen unterschiedliche Sollwerte 1 und 2 einerseits für die zellspezifische Wachstumsrate μ_w , sowie andererseits für die Flüssigphasentemperatur ϑ_{Lw} , bei gleicher Induktorkonzentration C_{IPTGw} eingestellt wurden. Die Zeitachse ist auf den Induktionszeitpunkt normiert.

Die Temperatursprünge resultieren in einer transienten Fluktuation der zellspezifischen Wachstumsraten, die sich im weiteren Verlauf jedoch gut an den vorgegeben Sollwerten orientieren. Die Proteinexpressionen sind für die ersten zwei Stunden auf einem annähernd gleichen Level, weichen dann jedoch deutlich voneinander ab, so dass letztlich eine höhere Ausbeute sowohl an löslichen als auch unlöslichen Protein für die Faktorkombination 2 vorliegt.

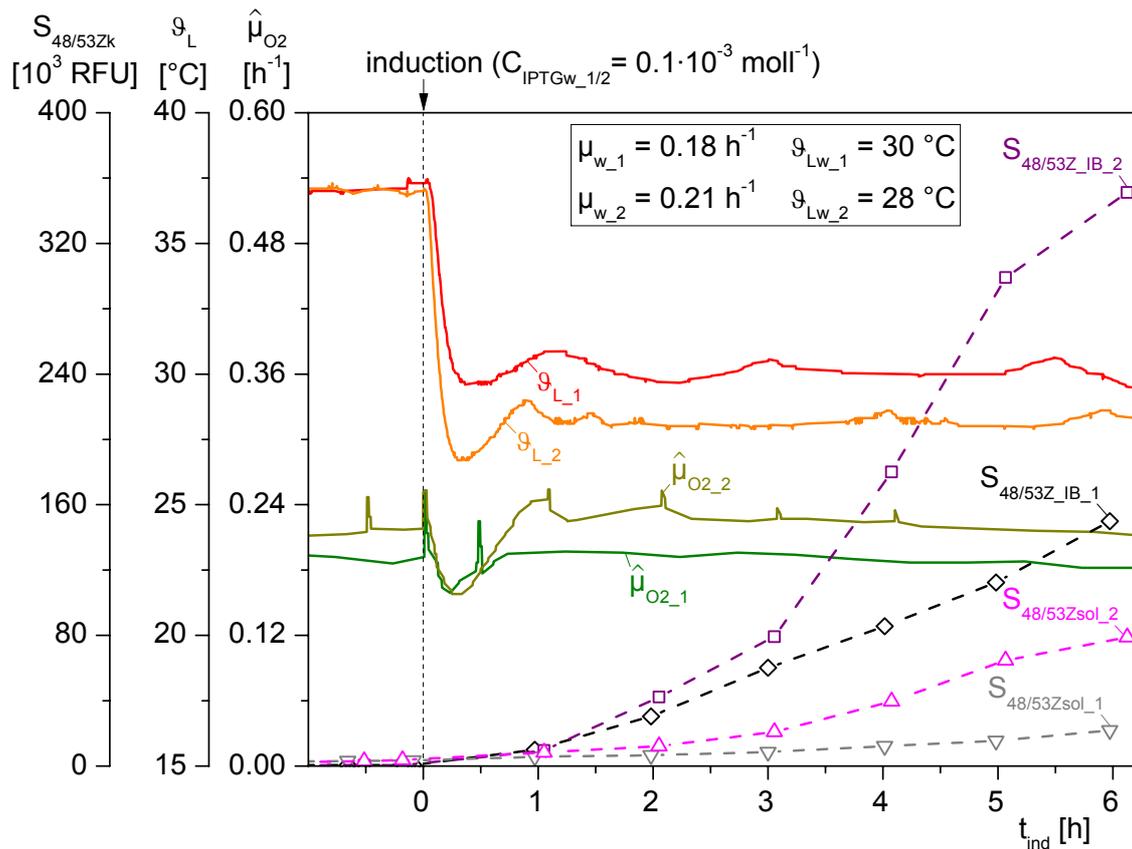


Bild 6.4: Vergleich der Proteinexpression mit unterschiedlich gewählten Faktorkombinationen für die zellspezifische Wachstumsrate und die Flüssigphasentemperatur

Für die Darstellung der Zusammenhänge ist das in Bild 6.5 gezeigte vereinfachte Modell des *E. coli* Produktionsprozesses hilfreich. Die auf das System wirkenden variablen Faktoren sind der Sollwert der zellspezifischen Wachstumsrate μ_w im Fed Batch (X_1), die Flüssigphasentemperatur ϑ_{Lw} ab Induktion (X_2) und die molare IPTG-Konzentration C_{IPTGW} (X_3). Die zur Beurteilung des Prozesses geeigneten Gütekriterien sind die RZA³ von zellinternem löslichem Protein Y_1 sowie die RZA von Inclusion Bodies Y_2 , wobei es gilt Y_1 zu maximieren.

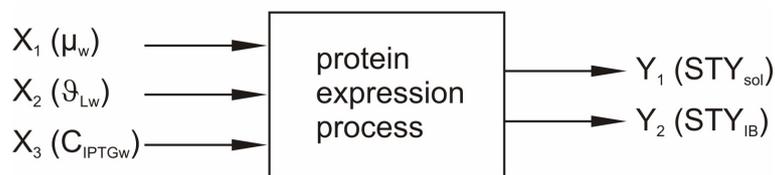


Bild 6.5: Analytierte Einfluss- und Zielgrößen der Proteinproduktion mit *E. coli*

³ RZA - engl.: space-time-yield, STY

6.2 Prozessoptimierung mit Statistischer Versuchsplanung

6.2.1 Genereller Ablauf und Modelle der Statistischen Versuchsplanung

Es wird deutlich, dass die Wahl geeigneter Expressionsbedingungen im Fokus einer optimierten Proteinproduktion steht. Eine systematische Herangehensweise zur Beschreibung multifaktorieller Beziehungen bzw. zum Auffinden einer geeigneten Faktorkombination ist die Statistische Versuchsplanung (SVP⁴).

Die Etablierung der SVP ist dem englischen Statistiker und Wissenschaftler R. A. Fisher zu verdanken. Er hatte Anfang der Zwanzigerjahre die Idee, Versuche zur Gewinnung von analytischen Daten systematisch unter bestimmten aus der Statistik abgeleiteten Bedingungen zu planen und auszuführen, um mit möglichst wenig Experimenten beschreibbare Zusammenhänge zwischen Einflussgrößen (Faktoren) X_j und Zielgrößen (Antworten) Y_k zu schaffen. Als Auswerteverfahren entwickelte er die Varianzanalyse (Fisher, 1935).

Fisher erreichte eine sprunghafte Effektivitätssteigerung durch das gleichzeitige Variieren aller Faktoren im Gegensatz zur Einfaktormethode, die in der Literatur als COST⁵, OFAT⁶ oder weniger gebräuchlich auch Ceteris-Paribus-Variation bekannt ist (Scheffler, 1997; Rinne, 2008; Eriksson *et al.*, 2008; Montgomery, 2009).

Bild 6.6 zeigt einen Vergleich der herkömmlichen Versuchsplanung mit der Einfaktormethode zu einem faktoriellen Design der SVP. Eingezeichnet sind die dem Experimentator unbekanntes Isoquanten einer Zielgröße in Abhängigkeit von zwei Faktoren X_1 und X_2 . Die schwarzen Kreise illustrieren die jeweiligen Versuchspunkte, der weiße Stern die momentan verwendete Faktorkombination – den sog. Zentralpunkt (CP⁷).

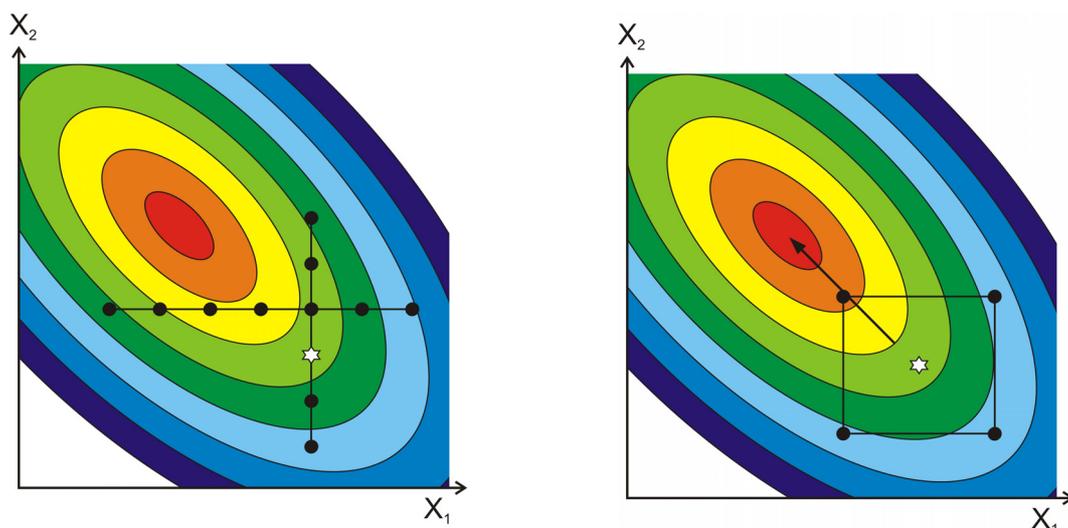


Bild 6.6: Vergleich der Einfaktormethode (links) mit einem Faktordesign nach der Statistischen Versuchsplanung (rechts)

⁴ SVP - engl.: Design of Experiments, DoE

⁵ COST - Changing One Single factor at a Time

⁶ OFAT - One Factor At a Time

⁷ CP - Center Point

Mit der Einfaktormethode werden Versuchsergebnisse generiert, die eine Abhängigkeit der Zielgröße von jeweils einem Faktor bei Konstanz aller anderen Faktoren erkennen lassen. Die erforderliche Anzahl der Versuche ist groß, und im Optimalbereich liegen nur wenige oder gar keine Versuche. Die SVP hingegen gestattet alle Faktoren innerhalb des Versuchsplanes simultan zu verändern. Mit wenigen Versuchen gelingt es dadurch, signifikante Faktoren von unwesentlichen zu unterscheiden, und die Richtung zu erkennen, in der das Optimum zu erwarten ist. Wechselwirkungen zwischen den Einflussgrößen werden erkannt und auch quantitativ erfasst (Scheffler, 1997).

Mit Hilfe der SVP soll der Effekt der analysierten Einflussgrößen auf die RZA untersucht werden. Das Ergebnis ist die mathematische Beschreibung der Ursache-Wirkung-Beziehung in Form eines empirischen Modells, das durch Regressionsanalyse aufgestellt wird und als eine Näherung nur wesentliche Glieder enthalten soll.

Je nach Umfang und Art der Daten kann dies ein multiples lineares Regressionsmodell (multiple linear regression, MLR), ein Hauptkomponenten-Regressionsmodell (principal component regression, PCR) oder ein Partial Least Square-Regressionsmodell (partial least square regression, PLS) sein, wobei die DoE-Software MODDE die Modellbildung über MLR und PLS unterstützt.

Hauptsächlich wird die MLR zur Auswertung in der Versuchsplanung herangezogen, da in den meisten Fällen eine lineare Versuchsanordnung und somit wenig Kollinearität in den unabhängigen Variablen x_j vorliegt. PLS findet Anwendung in komplizierten Designs, bei denen mehrere korrelierte Antworten vorhanden sind oder Kollinearität in den unabhängigen Variablen x_j vorliegt (Kessler, 2007; Eriksson *et al.*, 2008).

Neben einfachen linearen Regressionsmodellen,

$$y_k = \beta_0 + \sum_{i=1}^3 \beta_i \cdot x_i + \varepsilon_k \quad k = 1 \text{ (sol)}, 2 \text{ (IB)}, \quad (6.1)$$

werden Interaktionsmodelle 1. Ordnung betrachtet, bei denen der Regressand y_k ,

$$y_k = \beta_0 + \sum_{i=1}^3 \beta_i \cdot x_i + \sum_{1 \leq i < j}^3 \beta_{ij} \cdot x_i \cdot x_j + \beta_{123} \cdot x_1 \cdot x_2 \cdot x_3 + \varepsilon_k \quad k = 1 \text{ (sol)}, 2 \text{ (IB)}, \quad (6.2)$$

im Fall von $j = 3$ drei Regressoren durch 2^j Terme beschrieben wird.

β_0 ist der konstante Term, β_i sind die Regressionskoeffizienten der linearen Abhängigkeiten, β_{ij} wichtet die nicht-lineare Interaktion von x_i und x_j , β_{123} beschreibt die gemeinsame Interaktion aller drei Faktoren und ε ist der Modellfehler.

In der Praxis sind Wechselwirkungseffekte höherer Ordnung häufig so klein, dass sie in den Bereich der Streuung, d. h. in den Bereich der zufälligen Fehler fallen und daher oft vernachlässigt werden können (Eriksson *et al.*, 2008). Das erste Ziel einer DoE-Untersuchung wird

daher sein, statistisch nicht signifikante Terme auszuschließen bzw. die einflussreichsten Faktoren in einem Screening zu identifizieren.

Das im Screening erhaltene lineare oder interagierende Modell beschreibt bei zwei oder mehr Faktoren eine flache oder gekrümmte Ebene im Euklidischen Raum und kann für Vorhersagen durch Intrapolation oder Extrapolation eingesetzt werden. Eine Extrapolation führt zwar aus dem untersuchten Gebiet und damit auch aus dem Gültigkeitsbereich des Modells heraus, sie kann jedoch bei der Methode des steilsten Anstiegs als Hilfsmittel benutzt werden, um einen neuen Designraum für weitergehende Versuche festzulegen (Box and Wilson, 1951)

In vielen Fällen werden optimale Bedingungen, also ein Minimum, Maximum oder einen Sattelpunkt von Versuchsbedingungen gesucht. Das sog. Antwortflächenverfahren (RSM⁸) bietet für Optimierungsaufgaben eine Sammlung von mathematischen und statistischen Techniken, die eine Modellierung einer gekrümmten Kurve (ein Faktor) oder Wirkungsfläche (zwei oder mehr Faktoren) zulassen (Box and Wilson, 1951). Für diese Modellierung sind quadratische oder sogar kubische Modelle notwendig. Kubische Modelle finden insbesondere bei der Modellierung lebender Systeme Anwendung. In der industriellen Praxis sind sie jedoch selten relevant, da sie zu komplex sind und zu viele Experimente erfordern (Eriksson *et al.*, 2008).

Im Fall von drei Faktoren kann aus den Experimentalergebnissen ein Polynom 2. Grades mit der Antwort y_k ,

$$y_k = \beta_0 + \sum_{i=1}^3 \beta_i \cdot x_i + \sum_{1 \leq i < j}^3 \beta_{ij} \cdot x_i \cdot x_j + \sum_{i=1}^3 \beta_{ii} \cdot x_i^2 + \varepsilon_k \quad k = 1 \text{ (sol)}, 2 \text{ (IB)}, \quad (6.3)$$

abgeleitet werden, wobei β_{ii} die Koeffizienten der quadratischen Parameter repräsentiert.

Mit diesem Modell ist es möglich, eine Voraussage der Antwortwerte für jede mögliche Faktorkombination innerhalb der experimentellen Region zu treffen, und einen Punkt der in optimalen Betriebsbedingungen resultiert, den sog. Set-Point, zu identifizieren.

Gerade biotechnologische Herstellungsprozesse sind häufig durch Schwankungen in Produkt- und Nebenproduktgehalt in der Fermentationsbrühe gekennzeichnet. Daher soll mit einem Robustheitstest, dem letzten Schritt einer DoE-Untersuchung, die Sensitivität der Antwort auf kleine Faktoränderungen bestimmt werden.

Das Design für den Robustheitstest wird um den Set-Point aus der Optimierung zentriert und meistens mit einem einfachen linearen Modell evaluiert, da nur geringe Faktorvariationen und daher keine Abweichungen von der Linearität erwünscht sind. Die Methode ist dabei robust, wenn gewährleistet ist, dass sich einerseits der Mittelwert der Antwortgrößen auf einem gewünschten Level befindet und andererseits die Variabilität um dieses Level so gering wie möglich ist (Montgomery, 2009).

⁸ RSM - Response-Surface-Methodolgy

Tabelle 6.1 fasst die strategischen Schritte einer vollständigen Versuchsplanung zusammen, wobei jeder einzelne Schritt auch separat durchgeführt werden kann.

Tabelle 6.1: Schritte einer Prozessoptimierung mit Hilfe der Statistischen Versuchsplanung

Ziel	Modell und Modellgeometrie	Ergebnis
Screening	linear oder interagierend Gerade (ein Faktor) bzw. unverzerrt flache oder gekrümmte Hyperebene im 3-D-Raum (zwei oder mehr Faktoren)	Identifikation signifikanter Faktoren / Definition eines neuen Designraums
Optimierung	quadratisch oder kubisch parabolisch gekrümmte Kurve (ein Faktor) im 2-D-Raum oder gekrümmte Hyperfläche (zwei oder mehr Faktoren) im 3-D-Raum	Identifikation eines optimalen Punkts oder Bereichs auf Basis signifikanter Faktoren
Robustheit	linear oder interagierend Gerade (ein Faktor) bzw. unverzerrt flache oder gekrümmte Hyperebene im 3-D-Raum (zwei oder mehr Faktoren)	Feststellung der Robustheit gegenüber kleinen Fluktuationen in den Faktorlevels

6.2.2 Die Versuchsplanungs- und Auswertungssoftware MODDE

Die Statistische Versuchsplanung ist nicht unbedingt an die Verwendung spezieller Software gebunden. Für einfache Anwendungen genügt oftmals ein Taschenrechner oder ein PC mit der üblichen Software zur Tabellenkalkulation. Für statistisch nicht versierte Benutzer bzw. kompliziertere Anwendungen empfiehlt sich jedoch der Einsatz von Spezialsoftware, die eine modellgestützte Durchführung der Versuchsplanung und -auswertung erleichtert.

In dieser Arbeit wurde die Software MODDE (MODELing and DEsign) in der Version 8 und 9 der Fa. UMETRICS, Schweden, verwendet. Bild 6.7 zeigt die übersichtlich gestaltete und nach kurzer Einarbeitung leicht bedienbare Benutzeroberfläche des Programms, das der Richtlinie 21 CFR⁹ Part 11 für elektronische Aufzeichnungen und Unterschriften der U.S. Lebens- und Arzneimittelbehörde FDA entspricht.

Das Programm bietet eine Vielzahl von Hilfsmitteln, um den vollständigen Ablauf einer DoE-Untersuchung zu planen und auszulegen. Hierbei ebnet grafische und tabellarische Tools den Weg zu einer schnellen und objektiven Analyse bzw. Bearbeitung der experimentell gewonnenen Daten. Dies führt schließlich zu einer vereinfachten und anwendungsorientierten statistischen Auswertung sowie Modellierung der Faktor-Antwort-Beziehung.

⁹ CFR - Code of Federal Regulations

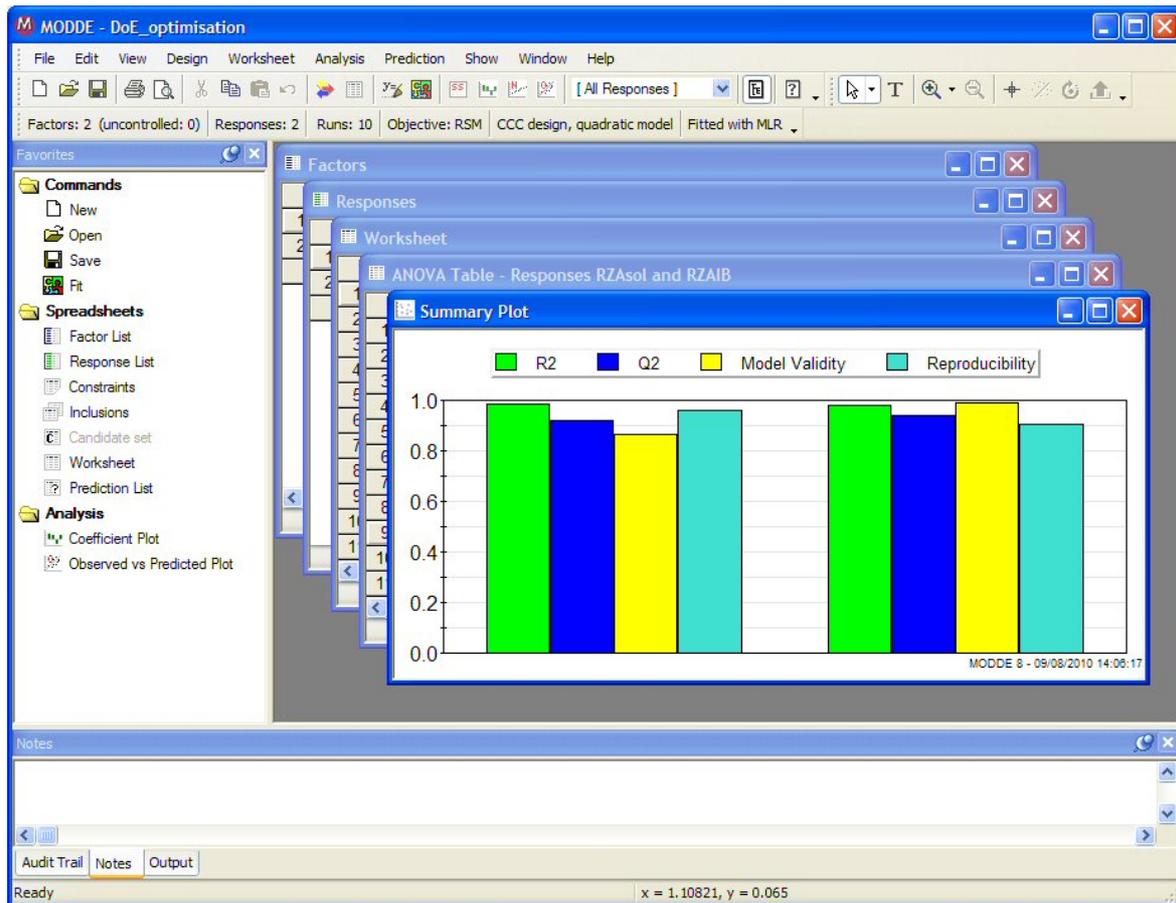


Bild 6.7: Benutzeroberfläche der DoE-Software MODDE

6.3 Festlegung des Versuchsbereich durch Spezifikation irregulärer Faktorbereiche

6.3.1 Zusammenhang zwischen Kultivierungstemperatur und Wachstumsrate

Das mesophile Bakterium *E. coli* wächst über einen weiten Temperaturbereich von ca. 40 °C, maximales Wachstum ist bei einer Temperatur von 37 bis 39 °C zu beobachten. Im so genannten normalen Arrhenius-Temperaturbereich, der bei *E. coli* von 21 °C bis 37 °C reicht, ist die Beziehung zwischen Wachstumsrate und Kultivierungstemperatur linear. Bei höheren oder tieferen Temperaturen fällt sie dagegen progressiv und nähert sich einer vertikalen Asymptote (Ingraham, 1987).

Unter unlimitierten Bedingungen nimmt *E. coli* nach einem Temperatursprung im Normalbereich sofort die neue Wachstumsrate an. Ein Sprung zu Temperaturen oberhalb bzw. unterhalb des Arrhenius-Bereichs ist mit erheblichen Veränderungen in der Proteinsynthese verbunden und ruft eine Hitze- bzw. Kälteschockantwort hervor (Lemaux, 1978; Jones *et al.*, 1987). Zur Ermittlung des funktionalen Zusammenhangs der Faktoren wurde eine Repeated Batch-Kultivierung in einem 5 L Laborbioreaktor BIOSTAT[®] Bplus der Fa. Sartorius Stedim durchgeführt. Bild 6.8 zeigt den Verlauf der Kultivierung ab Batchende.

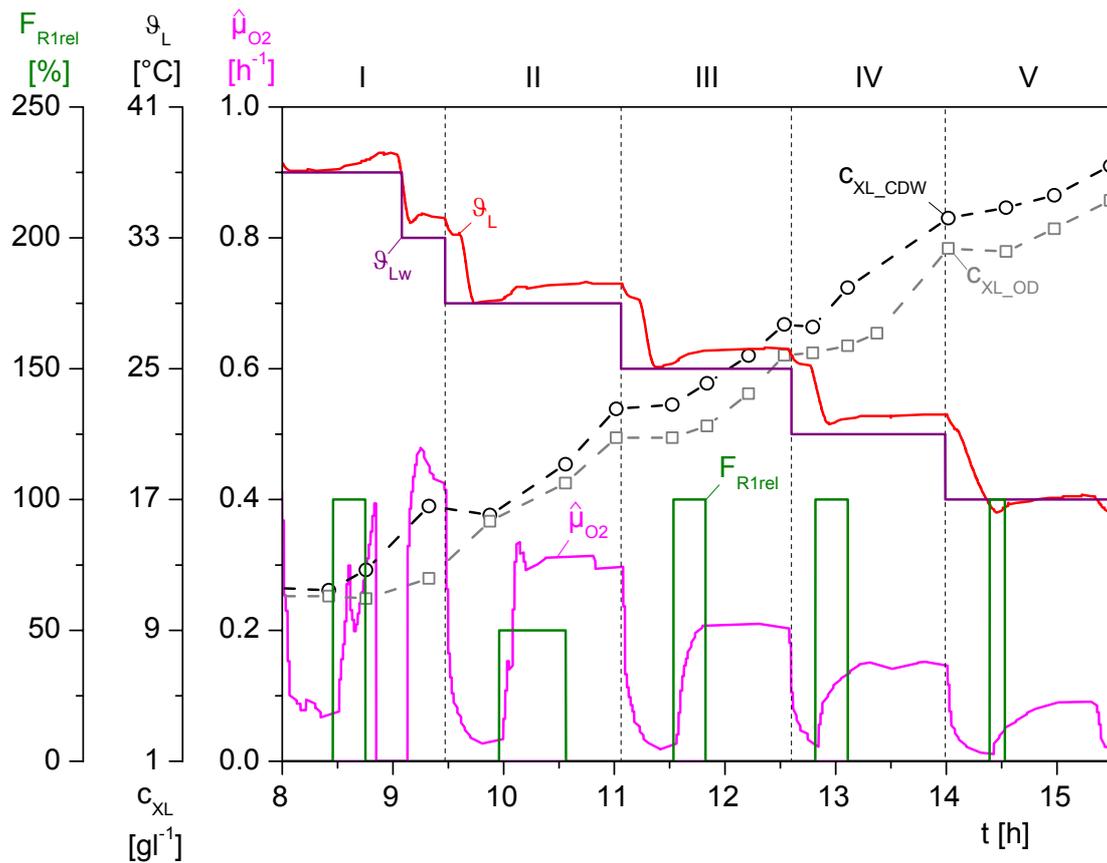


Bild 6.8: Repeated Batch-Kultivierung von *E. coli* MSD 5248 mit äquidistanten Temperatursprüngen von 4 °C in den Phasen I bis V

In den Phasen I bis V wurde ausgehend von 37 °C in äquidistanten Schritten von 4 °C jeweils ein neuer Sollwert für die Flüssigphasentemperatur ϑ_{LW} eingestellt. Der normale Arrhenius-Bereich wurde mit der Vorgabe eines suboptimalen Sollwerts von 17 °C in Phase V verlassen. Es ist zu erkennen, dass in jeder Phase die Temperatur mit einer bleibenden positiven oder negativen Regeldifferenz $\Delta\vartheta_L$ geregelt wird. In Phase I wurde der neue Temperatursollwert zu spät eingestellt, so dass dieses Intervall nicht zur Auswertung herangezogen werden konnte.

Im Gegensatz zum klassischen Repeated Batch-Prozess fand zum Ende der Phasen keine Konditionierung der Kulturbrühe durch Zellernte und Mediumauffrischung statt. Vielmehr wurde nach Verstoffwechslung des in jeder Phase neu gebildeten Acetats spurenelementhaltiges Feed-Medium (Riesenberg *et al.*, 1991) mit einer Glukosevorlagenkonzentration c_{S1R1} von 660 gl^{-1} als Substratpuls nachgeführt. Über eine maximale bzw. halbmaximale Ansteuerung der Zufütterpumpe F_{R1} stellte sich so eine vorgegebene Endkonzentration von 10 gl^{-1} in den Phasen I bis IV sowie 5 gl^{-1} in der Phase V ein.

Die vollständige Verstoffwechslung von Glukose sowie der Verbrauch des Metaboliten Acetat konnte wie in Bild 6.9 gezeigt sehr gut anhand der verfahrenstechnischen Prozessgrößen pO_2 und pH-Wert verfolgt werden. Nach dem ersten Anstieg des pO_2 -Wertes, der den

Verbrauch der vorgelegten Glukose anzeigt, wurde die pH-Regelung deaktiviert. Der pH-Wert verschiebt sich daraufhin in den basischen Bereich und nähert sich einem Plateau. Der zweite pO_2 -Peak deutet auf die vollständige Metabolisierung von Acetat hin, so dass die pH-Regelung wieder aktiviert und gleichzeitig die Zufütterung von Glukose gestartet wurde.

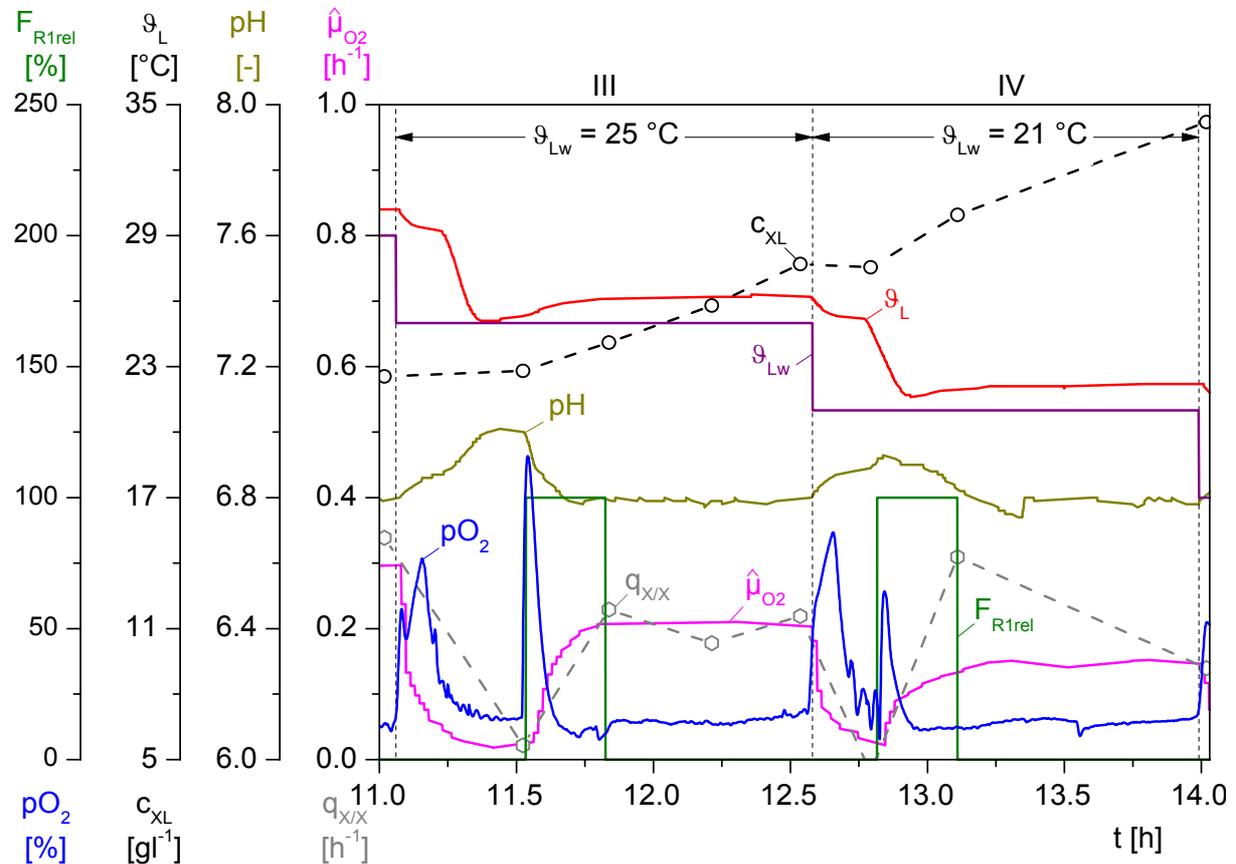


Bild 6.9: Indirekte Anzeige des Glukose- und Acetatverbrauchs über pO_2 und pH-Wert in den Repeated Batch-Phasen III und IV

Tabelle 6.2 zeigt eine Übersicht der Ergebnisse für die Batch-Phase sowie für die nachfolgenden fünf Repeated Batch-Phasen. Die Grundlage zur Berechnung der zellspezifischen Wachstumsrate bildet die nach $q_{X/X}$ umgestellte Zelldichtebilanz, wobei die Zelldichte aus der Offline-Bestimmung der Biotrockenmasse (CDW) bzw. der optischen Dichte (OD) resultiert. Die mittleren Reaktionsraten wurden postexperimentell mit Hilfe von MATLAB[®] nach dem Least-Square-Verfahren ermittelt.

Die Werte für die online beobachtbare Wachstumsrate $\hat{\mu}_{O_2}$ und die Regeldifferenz $\Delta\theta_L$ wurden über eine gewichtete Mittelwertbildung berechnet. Eine Gewichtung war notwendig, da MFCS/win Daten nicht äquidistant sondern ereignisabhängig vom konfigurierten Totband speichert. Prozesswerte, die über einen längeren Zeitraum konstant, d. h. innerhalb des Speichertotbandes waren, erhielten so eine höhere Gewichtung.

Tabelle 6.2: Auswertung der Batch-Phase und der Repeated Batch-Phasen I bis V

Phase	ϑ_{Lw} [°C]	$\bar{\vartheta}_L$ [°C]	$\Delta\bar{\vartheta}_L$ [°C]	\bar{q}_{X/X_CDW} [h ⁻¹]	\bar{q}_{X/X_OD} [h ⁻¹]	$\bar{q}_{X/X}^1$ [h ⁻¹]	$\bar{\mu}_{O_2}$ [h ⁻¹]
Batch	37	37.52	-0.52	-	0.442	0.442	0.442
I	33	34.27	-1.27	-	-	-	-
II	29	29.94	-0.94	0.309	0.267	0.288	0.305
III	25	26.08	-1.08	0.208	0.234	0.221	0.208
IV	21	22.11	-1.11	0.140	0.193	0.167	0.147
V	17	16.97	+0.03	0.073	0.096	0.085	0.087

¹ Einfaches arithmetisches Mittel aus \bar{q}_{X/X_CDW} und \bar{q}_{X/X_OD}

Bild 6.10 zeigt den funktionalen Zusammenhang zwischen Kultivierungstemperatur und Wachstumsrate im untersuchten Temperaturbereich [ϑ_{Lmin} , ϑ_{Lmax}]. Durch lineare Regression der identifizierten maximalen Wachstumsraten $\bar{q}_{X/X}$ (○) auf die mittleren Temperaturen $\bar{\vartheta}_L$, konnte die dargestellte Kennlinie mit einem Bestimmtheitsmaß R^2 von 0.993 erhalten werden.

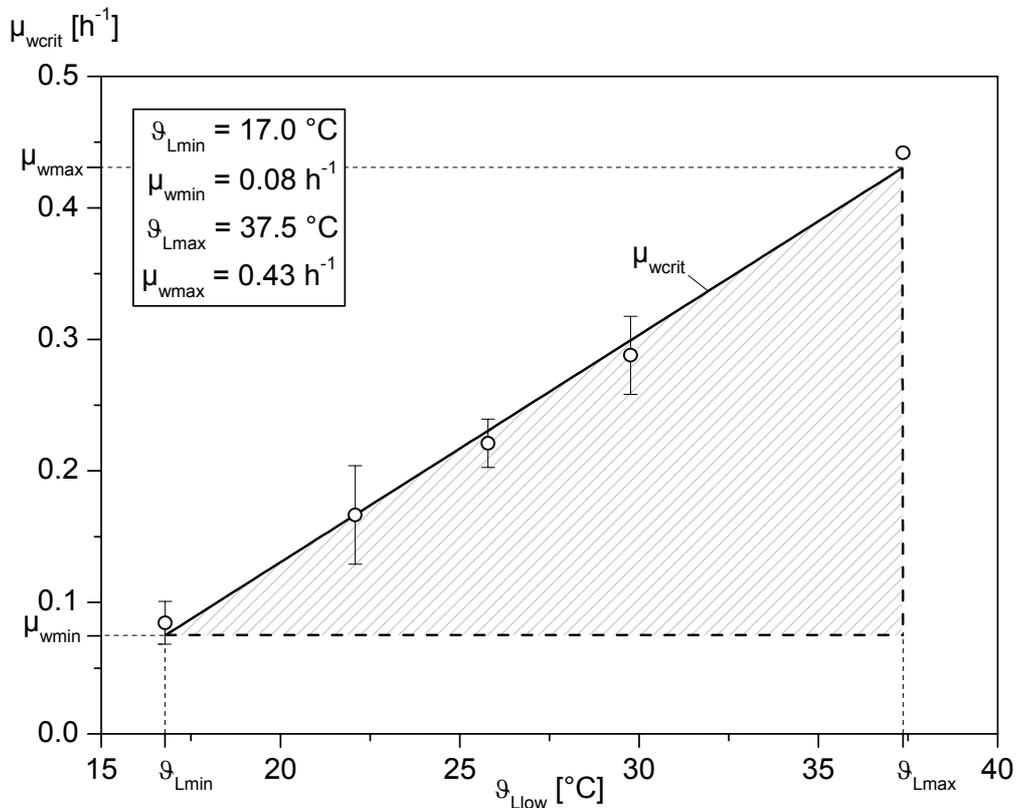


Bild 6.10: Linearer Zusammenhang zwischen Kultivierungstemperatur ϑ_L und steuerbarer zellspezifischer Wachstumsrate μ_w

Der reguläre Versuchsbereich ist durch die schraffierte, dreieckige Fläche unter der Regressionsgeraden dargestellt. Innerhalb dieses Versuchsbereichs können die Faktoreinstellungen unabhängig voneinander variiert werden. Der kritische Wert der steuerbaren zellspezifischen Wachstumsrate μ_{wcrit} ,

$$\mu_{\text{wcrit}} = \mu_{\text{wmax}} - \frac{\mu_{\text{wmax}} - \mu_{\text{wmin}}}{\vartheta_{\text{Lmax}} - \vartheta_{\text{Lmin}}} \cdot (\vartheta_{\text{Lmax}} - \vartheta_{\text{Llow}}), \quad (6.4)$$

wird dabei durch den unteren Wert der Kultivierungstemperatur ϑ_{Llow} bestimmt.

6.3.2 Zusammenhang zwischen Acetatbildung und Wachstumsrate

Beim aeroben Wachstum von *E. coli* auf acetogenen C-Quellen wie z. B. Glukose kann das unerwünschte Nebenprodukt Acetat gebildet und exkretiert werden. Dieses Phänomen wird als Overflow-Metabolismus oder bakterieller Crabtree-Effekt bezeichnet und resultiert aus einem Ungleichgewicht zwischen Aufnahme der primären C-Quelle und ihrer Umwandlung zu Biomasse und Produkten (Crabtree, 1929; Doelle *et al.*, 1982; Rinas *et al.*, 1989; Luli and Strohl, 1990; Xu *et al.*, 1999a and 1999b).

Die Akkumulation von Acetat wirkt sich negativ auf das Prozessverhalten aus. Insbesondere wird von einer Abnahme der Biomasseausbeute, einer Wachstumsinhibierung bei hohen Konzentrationen sowie einer Verminderung der Produktion rekombinanter Proteine berichtet (Konstantinov *et al.*, 1990; Luli and Strohl, 1990; Riesenberg *et al.*, 1990; Rothen *et al.*, 1998; Contiero *et al.*, 2000).

Im Allgemeinen ist die Menge an Acetat, die während einer aeroben Kultivierung gebildet wird, vom *E. coli* Stamm, den Wachstumsbedingungen, der Glukosekonzentration im Medium und der allgemeinen Zusammensetzung des Kulturmediums abhängig. Strategien zur Minimierung der Acetatakkumulation sind daher entweder bioprozesstechnischer oder genetischer Natur und reichen von der Entwicklung komplexer Algorithmen zur Kontrolle der Glukosekonzentration im Medium bis hin zu gezielten Veränderungen des zentralen Stoffwechsels von *E. coli* (Han *et al.*, 1992; Holms, 1996; De Mey *et al.*, 2007).

Insbesondere hervorzuheben sind Strategien, bei denen das Wachstum durch limitierte Zufütterraten der C-Quelle beschränkt wurde. Hier verringerte sich die Acetatexkretion und wurde in Chemostatkulturen unter einer kritischen Wachstumsrate μ_{crit} von 0.3 h^{-1} sogar gestoppt (Reiling *et al.*, 1985; El-Mansi and Holms, 1989).

Um die Wachstumsrate zu bestimmen, bei der die aerobe Acetatproduktion einsetzt, wurde in einem 5 L Laborbioreaktor BIOSTAT[®] Bplus der Fa. Sartorius Stedim eine Fed-Batch Kultivierung nach der direkten μ -stat Methode durchgeführt (Luttmann *et al.*, 1994).

Bild 6.11 zeigt den Verlauf der Fed-Batch Phase, in der die beobachtbare zellspezifische Wachstumsrate $\hat{\mu}_{O_2}$ über die Substratzugabe und der Sauerstoffpartialdruck p_{O_2} über die Rührerdrehzahl konstant gehalten wurde. Die p_{O_2} /agitation-Regelung ist aus Übersichtsgründen nicht dargestellt.

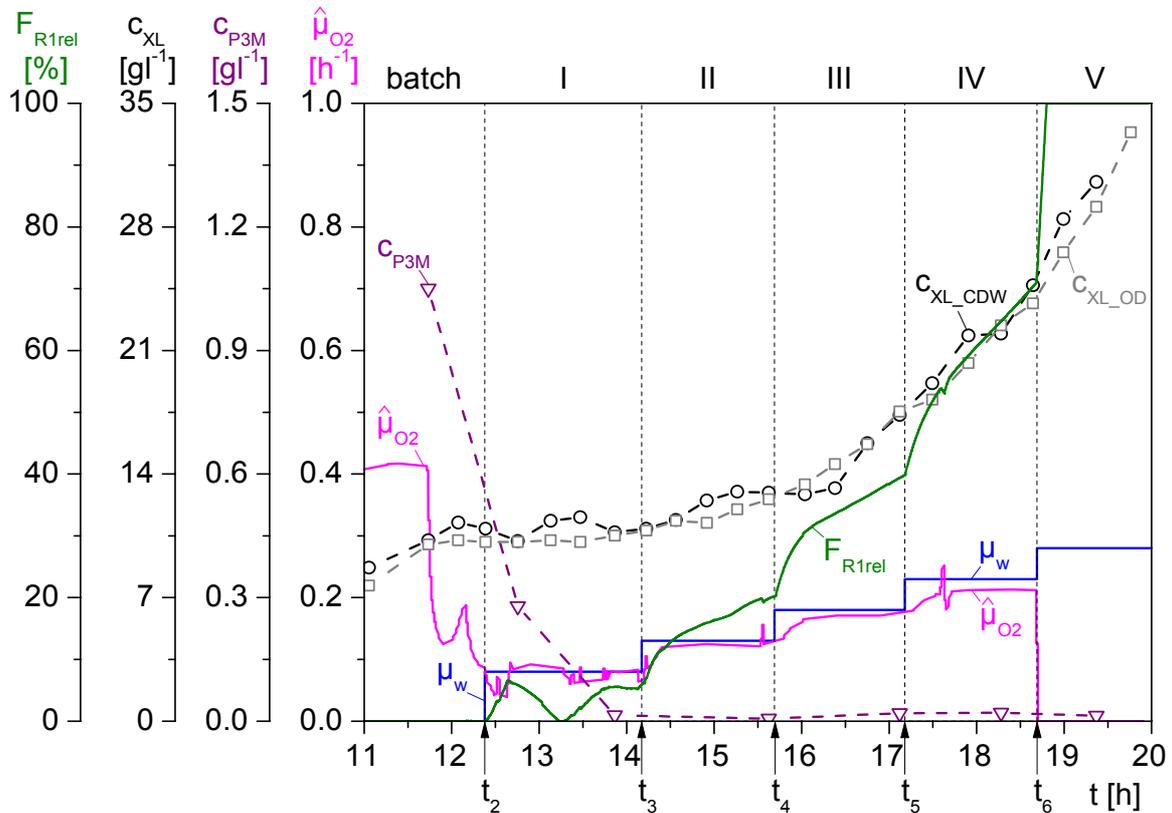


Bild 6.11: Fed-Batch Phase einer Kultivierung von *E. coli* MSD 5248 mit geregelter zellspezifischer Wachstumsrate nach dem direkten μ -stat Prinzip

Nach dem Glukose-Batch sowie einer vierzigminütigen Übergangsphase zur Wiederaufnahme des bis zum Batchende gebildeten Acetats, wurde bei t_2 ein erster Sollwert $\mu_{w2} = 0.08 \text{ h}^{-1}$ vorgegeben. Im weiteren Verlauf des Experimentes erfolgten vier weitere äquidistante Erhöhungen des Sollwertes in Schritten von 0.05 h^{-1} . Es ist zu erkennen, dass der PID-Regler die Regelgröße $\hat{\mu}_{O_2}$ gut an die Sollwerte heranführte, dennoch in jeder Phase eine positive Regeldifferenz bleibt, die mit zunehmender Prozessdauer und ansteigendem Sollwert größer wird.

Bei t_6 wurde der Sauerstoffeingangsmolenbruch $x_{O_{Gin}}$ erhöht, da die Rührerdrehzahl ihren Maximalwert erreichte. Dieser Stelleingriff führte zu einer Fehlberechnung der beobachtbaren Wachstumsrate. Bedingt durch den Nullwert der Regelgröße folgte der μ /feed-Regler der vermeintlich höheren Substratanforderung der Zellen mit einem Anheben der Steuergröße auf ihren Maximalwert von 100 %.

Tabelle 6.3 gibt eine Übersicht der eingestellten Sollwerte sowie postexperimentell berechneten mittleren Wachstumsraten und den gemessenen Acetatkonzentrationen zum Ende eines Sollwertintervalls. Ein Vergleich zwischen den offline und online bestimmten Werten zeigt besonders mit steigender Wachstumsrate große Unterschiede. Dennoch kam es im gesamten Verlauf der Fed-Batch Phase zu keiner Acetatbildung, was dafür spricht, dass die limitiert zugefütterte Glukose vollständig oxidativ verwertet werden konnte.

Tabelle 6.3: Übersicht der abschnittsweise ermittelten Wachstumsraten sowie der zum Ende eines Sollwertintervalls gemessenen Acetatkonzentrationen

Phase	μ_w [h ⁻¹]	$\hat{\mu}_{O_2}$ [h ⁻¹]	$\Delta\hat{\mu}_{O_2}$ [h ⁻¹]	\bar{q}_{X/X_CDW} [h ⁻¹]	\bar{q}_{X/X_OD} [h ⁻¹]	$\bar{q}_{X/X}^1$ [h ⁻¹]	C_{P3Mend} [gl ⁻¹]
I	0.08	0.079	+0.001	0.054	0.084	0.069	0.015
II	0.13	0.123	+0.007	0.127	0.113	0.120	0.006
III	0.18	0.171	+0.009	0.297	0.265	0.281	0.019
IV	0.23	0.211	+0.019	0.253	0.250	0.252	0.021
V	0.28	-	-	0.245	0.296	0.271	0.013

¹ Einfaches arithmetisches Mittel aus \bar{q}_{X/X_CDW} und \bar{q}_{X/X_OD}

6.3.3 Einfluss der IPTG-Konzentration auf die Proteinexpression

Die Verwendung von IPTG in biopharmazeutischen Prozessen ist nicht nur teuer, sondern aufgrund eines möglichen toxischen Effekts auf den Wirtsorganismus sowie einer gesundheitsschädlichen Wirkung auf den Menschen auch kritisch (Kosinski *et al.*, 1992; Donovan *et al.*, 1996). Aus diesen Gründen ist es vorteilhaft weniger Induktor zu verwenden, und dennoch das gleiche Expressionslevel zu erreichen.

IPTG wird im Gegensatz zu Laktose nicht von *E. coli* metabolisiert. Das Enzym β -Galactosidase (Genprodukt von *lacZ*) ist zwar in der Lage die O-glykosidische Bindung im Laktose-Molekül zu hydrolysieren, so dass die beiden Monosaccharide Galaktose und Glukose entstehen, nicht jedoch die S-glykosidische Bindung im IPTG-Molekül (Herzenberg, 1959; Knippers, 2006).

Durch das Enzym β -Galactosid-Transacetylase (Genprodukt von *lacA*) kann IPTG in der Zelle allerdings zu Acetyl-IPTG umgewandelt werden. Die physiologische Bedeutung dieses Enzyms für den Bakterienstoffwechsel ist noch nicht abschließend erforscht, es wird jedoch vermutet, dass es eine entgiftende Wirkung auf Laktoseanalogons hat, die eine Zelle schädigen könnten (Zabin *et al.*, 1962; Roderick, 2005).

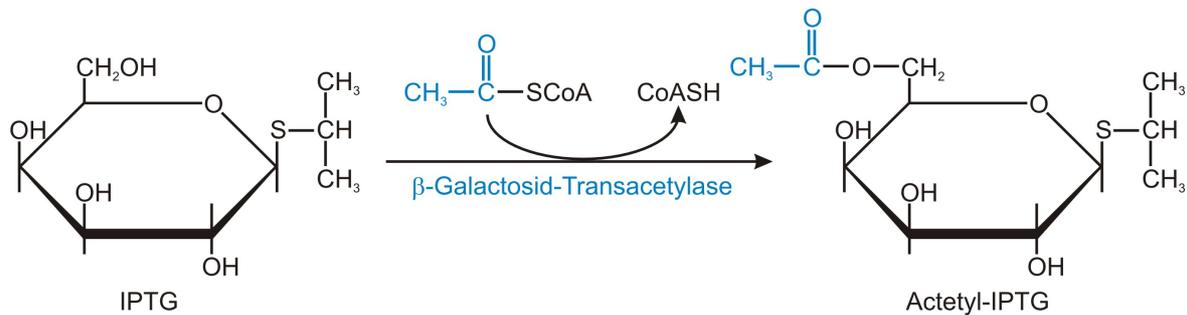


Bild 6.12: Enzymatische Umwandlung von IPTG zu Acteyl-IPTG durch die β -Galactosid-Transacetylase (Tropp, 2008)

Im Gegensatz zu Laktose ist der Transport von IPTG in die Zelle nicht genreguliert. Zwar nimmt die β -Galactosid-Permease (Genprodukt von *lacY*) am Transport von IPTG über die Zellmembran teil, dennoch ist dieser Mechanismus im Vergleich zu anderen Transportmechanismen oder Diffusion in *E. coli* weniger signifikant (Jensen *et al.*, 1993; Khlebnikov and Keasling, 2002; Noel *et al.*, 2009).

Unabhängig vom Transportmechanismus hat die in der Zelle vorliegende IPTG-Konzentration einen Einfluss auf die Höhe der rekombinanten Proteinexpression. Eine teilweise oder vollständige Titration der zellinternen *lac*-Repressormoleküle führt zu unterschiedlich hohen Expressionsleveln sowie parallel zu einer Beeinflussung des Wirtsmetabolismus. Unter ungünstigen Umständen kann eine zu hoch gewählte Induktorkonzentration mit Wachstumsinhibierung oder sogar Zellyse einhergehen (Kosinski *et al.*, 1992; Striedner *et al.*, 2003).

In der Literatur gibt es nur wenige Studien, die sich mit der Optimierung der IPTG-Konzentration für die rekombinante Proteinexpression beschäftigen und in den seltensten Fällen wird ein Grund für die Verwendung einer bestimmten IPTG-Konzentration angegeben. Häufig wird mit 1.0 mmol^{-1} induziert, es werden jedoch Konzentrationen in einem Bereich von 0.005 bis 5.0 mmol^{-1} eingesetzt (Donovan *et al.*, 1996).

6.3.4 Grenzen des regulären Versuchsbereichs

Der reguläre Bereich für die Durchführung der DoE-Versuche wird durch minimale und maximale Faktorstufen der quantitativen Einflussgrößen zellspezifische Wachstumsrate μ_w im Fed Batch, Flüssigphasentemperatur ϑ_{LW} ab Induktion und molare IPTG-Konzentration C_{IPTGw} definiert.

Wie in den vorangegangenen Kapiteln verdeutlicht, ist jedoch infolge verfahrenstechnischer, biologischer oder anderweitiger Einschränkungen bzw. aufgrund empirischer oder theoretischer Vorkenntnisse über das zu untersuchende System nicht jede beliebige Faktorstufenkombination realisierbar. Die Terminologie der Statistischen Versuchsplanung bezeichnet eine solche Begrenzung als „Constraint“ (Kessler, 2007; Eriksson *et al.*, 2008).

Tabelle 6.4 gibt eine Übersicht der gewählten Faktoren mit ihrem jeweiligen Grenzwertintervall, in dem alle Einflussgrößen unabhängig voneinander variierbar sind. Dabei entspricht der minimale oder maximale Grenzwert eines Faktors $X_{j\min}$ bzw. $X_{j\max}$ nicht notwendigerweise der später für das Design tatsächlich festgelegten unteren oder oberen Stufenhöhe $X_{j\text{low}}$ bzw. $X_{j\text{high}}$.

Tabelle 6.4: Gesamtverzeichnis der Einflussgrößen mit identifiziertem Grenzwertintervall

Faktor	Variable	Bezeichnung	Dimension	$X_{j\min}$	$X_{j\max}$
X_1	μ_w	zellspez. Wachstumsrate im Fed-Batch	h^{-1}	0.08	0.27
X_2	ϑ_{Lw}	Flüssigphasentemperatur ab Induktion	$^{\circ}\text{C}$	17.0	37.5
X_3	C_{IPTGw}	molare IPTG-Konzentration	mmol^{-1}	0.005	5.0

6.3.5 Koordinatentransformation der Faktorstufen

Für eine Einflussgröße ist praktisch nie ein beliebig kleiner oder ein beliebig großer Wert sinnvoll. Die Vorkenntnisse über das zu untersuchende System, die Möglichkeiten der Realisierung, die Aufgabenstellung usw. legen die untere und die obere sinnvolle Stufenhöhe für die jeweilige Einflussgröße fest.

Die Vorgehensweise in der Versuchsplanung und -auswertung erfordert eine Normierung der Stufenhöhe durch lineare Transformation der physikalischen Werte der Faktorstufen X_j in dimensionslose kodierte Werte x_j ,

$$x_j = \frac{2 \cdot (X_j - X_{j\text{cp}})}{X_{j\text{high}} - X_{j\text{low}}}, \quad j = 1, 2, 3, \quad (6.5)$$

wobei $X_{j\text{cp}}$ den natürlichen Wert im Zentralpunkt, $X_{j\text{high}}$ und $X_{j\text{low}}$ die natürlichen Werte der oberen bzw. unteren Stufe eines Faktors wiedergeben.

Durch die Koordinatentransformation ergibt sich eine Darstellung, bei der der Abstand vom Zentralpunkt bis hin zur unteren und oberen Stufe eine Einheit beträgt. Die Normierung (Kodierung) ist wichtig, da sie die Untersuchung von Variablen mit unterschiedlicher Größenordnung ermöglicht (Bezerra *et al.*, 2008).

6.4 Ermittlung der Raum-Zeit-Ausbeute als relevante Zielgröße

6.4.1 Fluoreszenzmessung mit dem GENios™ Reader

Während der Fed Batch- und Produktionsphase einer Kultivierung wurden Proben für die Zielproteinanalytik bei einer Temperatur von -20 °C zurückgestellt. Nach einem Ultraschallaufschluss der zurückgestellten Proben wurde die lösliche Fraktion des *E. coli* Lysats von der unlöslichen Fraktion durch Zentrifugation getrennt. Die Inclusion Bodies wurden anschließend über ein entwickeltes Solubilisierungsverfahren in ihre native Proteinkonformation überführt. Die genaue Vorgehensweise ist im Anhang dieser Arbeit dokumentiert.

Durch die Verwendung von ZsGreen als Fusionspartner konnte eine Zielproteinanalytik über den Einsatz eines Fluoreszenzreaders etabliert werden. Die Messung der aufbereiteten Proben wurde mit dem GENios™ Reader der Fa. Tecan durchgeführt. Die Steuerung des in Bild 6.13 dargestellten Geräts sowie die Datenerfassung erfolgt über einen externen PC mit Hilfe der Magellan® Software.



Bild 6.13: GENios™ Fluoreszenzreader, rechts mit schwarzer 96-Well Mikrotiterplatte

Das optische Signal der Fluoreszenzmessung beider Zielproteinfraktionen $S_{48/53Z_{sol}}$ bzw. $S_{48/53Z_{IB}}$ wurde unter Verwendung von 96-Well Platten der Fa. Nunc mit einer Wellenlängenkombination $\lambda_{ex} = 480 \pm 20\text{ nm}$ und $\lambda_{em} = 535 \pm 20\text{ nm}$ gemessen. Damit wurde die Anregungswellenlänge von ZsGreen von 496 nm und die Emissionswellenlänge von 506 nm abgedeckt. Es wurden schwarze Mikrotiterplatten verwendet, die minimalen Hintergrund und minimales Streulicht bei Fluoreszenzuntersuchungen bieten.

6.4.2 Berechnung der Raum-Zeit-Ausbeute

Es ist bekannt, dass die RNA-Polymerase bei *E. coli* BL21 Stämmen auch ohne IPTG-Induktion stromaufwärts des chromosomalen *lacUV5*-Promotors bindet und geringe Mengen an T7-RNA-Polymerase synthetisieren kann. Verbunden mit einer nicht vollständigen Repression aller plasmidkodierten T7-Promotoren, ist so trotz der patentierten Tandem-ppop-Operatorsequenz eine basale Expression des Zielprodukts möglich.

Da den verwendeten *E. coli*-Stämmen ein Sekretionsmechanismus fehlt, kann während des Prozesses Zielprodukt nur aus toten, lysierten Zellen als lösliche zellexterne Proteinfraction $m_{P1M_{sol}}$ bzw. unlösliche zellexterne Proteinfraction $m_{P1M_{IB}}$ in die Umgebung abgegeben werden.

Da im Kultivierungsmedium nicht-physiologische Bedingungen herrschen, sowie durch enzymatische Abbaureaktionen Produktartefakte entstehen können, wird der Massentransfer löslichen Proteins von der Zell- in die Medienphase,

$$\dot{m}_{P1M_{lys}}(t) = q_{X/X_{lys}}(t) \cdot m_{P1Z_{sol}}(t), \quad (6.6)$$

im Weiteren bilanztechnisch nicht berücksichtigt.

Die unlösliche zellexterne Proteinfraction $m_{P1M_{IB}}$ wird dagegen bei der Trennung von nativem Protein und Inclusion Bodies bedingt durch den Dichteunterschied durch die Zentrifugation wieder in die feuchte Zellmasse zurückgeführt.

Aufgrund der Vorgehensweise bei der Probenaufarbeitung war eine analytische Trennung der zellinternen und zellexternen Inclusion Body Fraction nicht möglich, so dass die beiden Anteile zusammengeführt und als Gesamtfraction $m_{P1Z_{IB}}$ in der Zellphase Z notiert werden.

Das abzentrifugierte Pellet der Proben für die Offline-Proteinanalytik besteht einerseits aus intakten, andererseits aus Anteilen der während der Kultivierung lysierten Zellen. Da das Verhältnis von lebenden zu toten Zellen nicht bekannt ist, wird für die weitere Bilanzierung angenommen, dass die gemessene Zellmasse m_{XL} zu 100 % produktbildend ist.

Die Grundlage zur Auswertung der DoE-Versuche bildet die Massenbilanz für den löslichen ($k = sol$) und unlöslichen ($k = IB$) zellinternen Proteinanteil,

$$\dot{m}_{P1Zk}(t) = -\frac{F_S(t)}{V_L(t)} \cdot m_{P1Zk}(t) + q_{P1/Xk}(t) \cdot m_{XL}(t), \quad (6.7)$$

die neben dem produktbehafteten Verlust mit dem abfließenden Probenahmestrom F_S , auch die zellmassenbedingte Zunahme an Produkt mit den zellspezifischen Produktbildungsraten $q_{P1/X_{sol}}$ bzw. $q_{P1/X_{IB}}$ betrachtet.

Durch Umstellung der Massenbilanz aus Gl. (6.7) wird die auf den löslichen bzw. unlöslichen Anteil bezogene Produktivität,

$$PRD_k(t) = P_1ER_k(t) \cdot V_L(t) = q_{P1/Xk}(t) \cdot m_{XL}(t) = \dot{m}_{P1Zk}(t) + \frac{F_S(t)}{V_L(t)} \cdot m_{P1Zk}(t), \quad (6.8)$$

hergeleitet, die eine differentielle Änderung der Produktmasse mit der Zeit beschreibt.

Ein verfahrenstechnisch sinnvolles Gütekriterium zur Bewertung der Versuche ist jedoch die auf die gesamte Reaktionszeit und das Endvolumen bezogene integrale Produktivität. Dieser Wert wird hier als mittlere Raum-Zeit-Ausbeute (RZA) bezeichnet.

Wie Bild 6.14 zu entnehmen, wird die Reaktionszeit durch das gewählte Intervall vom Beginn der Fed Batch Phase t_2 bis zum Ende des Prozesses t_{end} begrenzt. Da die Dauer der Batchphase durch die Vorgabe gleicher Anfangsbedingungen bei allen durchgeführten Kultivierungen annähernd gleich lang war, wird diese Zeit nicht in das Intervall mit einbezogen.

Die zur Berechnung der Raum-Zeit-Ausbeuten erforderlichen Prozessgrößen können sowohl vor der Probenahme mit $t_{\text{Si}(-)}$ als auch nach der Probenahme mit $t_{\text{Si}(+)}$ betrachtet werden. Der Probenahmestrom F_S wird dabei über die nahezu impulsförmige Probenahmerate ΔV_S bei $t = t_{\text{Si}}$ approximiert.

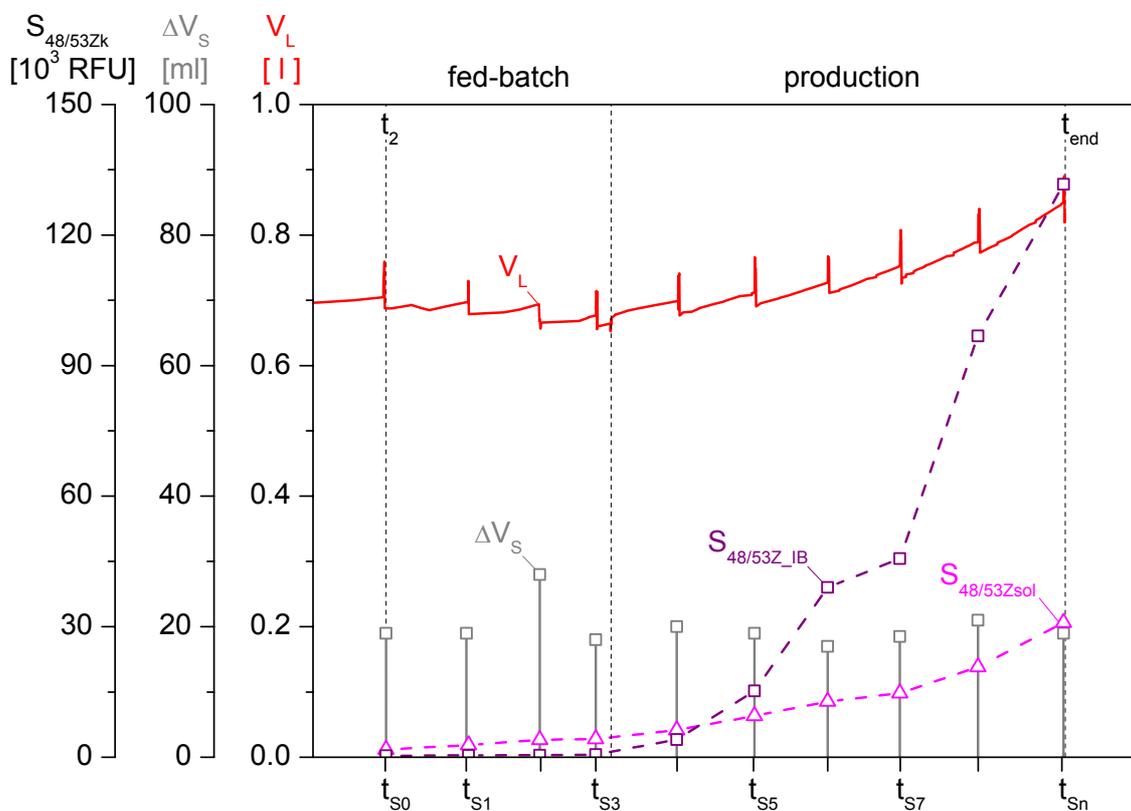


Bild 6.14: Bioreaktionsvolumen V_L , Fluoreszenzintensität für lösliches und unlösliches Zielprotein $S_{48/53Zsol}$ bzw. $S_{48/53Z_IB}$ sowie Probenahmenvolumen ΔV_S zu unterschiedlichen Probenahmezeitpunkten t_{Si}

Über eine Mittelwertbildung der Produktivitäten PRD_{sol} und PRD_{IB} im gewählten Integrationsintervall $[t_{S0}, t_{Sn}]$ kann die mittlere volumetrische Raum-Zeit-Ausbeute für das lösliche Zielprotein sowie für den unlöslichen Inclusion Body Anteil,

$$\begin{aligned} RZA_k &= \frac{1}{V_L(t_{Sn(+)})(t_{Sn} - t_{S0})} \cdot \int_{t_{S0(+)}^{t_{Sn(+)}} PRD_k(t) dt \\ &= \frac{1}{V_L(t_{Sn(+)})(t_{Sn} - t_{S0})} \cdot \left[\int_{t_{S0(+)}^{t_{Sn(+)}} \dot{m}_{P1Zk}(t) dt + \int_{t_{S0(+)}^{t_{Sn(+)}} F_S(t) \cdot c_{P1Zk}(t) dt \right], \end{aligned} \quad (6.9)$$

berechnet und durch anschließende Integration in eine numerisch auswertbare Form,

$$\begin{aligned} RZA_k &= \frac{1}{V_L(t_{Sn(+)})(t_{Sn} - t_{S0})} \cdot \left[V_L(t_{Sn(+)})(c_{P1Zk}(t_{Sn}) - c_{P1Zk}(t_{S0})) \right. \\ &\quad \left. + \sum_{i=1}^n \Delta V_S(t_{Si}) \cdot \delta(t - t_{Si}) \cdot c_{P1Zk}(t_{Si}) \right], \end{aligned} \quad (6.10)$$

überführt werden.

Eine Proteinquantifizierung über elektrophoretische Methoden wurde in dieser Arbeit nicht durchgeführt. Die in Gl. (6.10) aufgeführten Konzentrationen c_{P1Zsol} und $c_{P1Z_{IB}}$ werden daher durch die jeweilige gemessene Fluoreszenzintensität $S_{48/53Zsol}$ und $S_{48/53Z_{IB}}$ ersetzt, so dass nun die mittlere volumetrische Raum-Zeit-Ausbeute RZA^* [RFUh⁻¹] für den löslichen und unlöslichen Anteil,

$$\begin{aligned} RZA_k^* &= \frac{1}{V_L(t_{end(+)})(t_{end} - t_2)} \cdot \left[V_L(t_{end(+)})(S_{48/53Zk}(t_{end}) - S_{48/53Zk}(t_2)) \right. \\ &\quad \left. + \sum_{i=1}^n \Delta V_S(t_{Si}) \cdot S_{48/53Zk}(t_{Si}) \right], \end{aligned} \quad (6.11)$$

zur Bewertung der durchgeführten Kultivierungen herangezogen werden kann.

7 Ermittlung optimaler Betriebsparameter mit Statistischer Versuchsplanung

7.1 Die Durchführung der Screeningversuche

7.1.1 Festlegung der Faktorstufen für das vollfaktorielle Screening-Design

Das experimentelle Ziel der Screening-Versuche ist die Identifikation der Faktoren, die einen signifikanten Einfluss auf die Raum-Zeit-Ausbeute haben, sowie die Bestimmung von neuen Faktoreinstellungen. Außerdem soll der Einfluss der in die Aminosäuresequenz des Zielproteins eingefügten Punktmutation [L767P] auf die lösliche Raum-Zeit-Ausbeute untersucht werden und als Entscheidungsgrundlage für die Auswahl eines *Escherichia coli* Stamms für die nachfolgenden Optimierungsexperimente dienen.

Für die Durchführung der Versuche wurde unter Berücksichtigung der gerätetechnischen Ausstattung mit sechs Bioreaktoren ein vollfaktorieller Versuchsplan mit zwei Stufen für die drei Faktoren gewählt, so dass mit Hilfe des Interaktionsmodells aus Gleichung 6.2 neben den drei Haupteffekten auch alle möglichen Wechselwirkungen (drei 2-Faktor-Wechselwirkungen und eine 3-Faktor-Wechselwirkung bzw. ein Blockeffekt) getrennt erfasst werden können. Da die Faktoren gleichzeitig und in zwei Stufen variiert werden, liegen die Versuchspunkte in den Ecken eines Würfels.

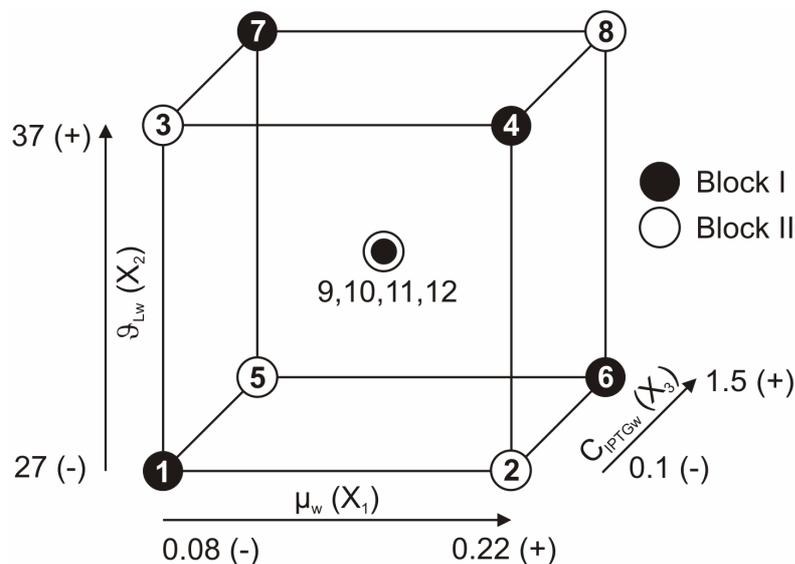


Bild 7.1 zeigt die Geometrie des vollfaktoriellen Designs mit den unteren und oberen Faktorstufen $X_{j\text{low}}$ bzw. $X_{j\text{high}}$ sowie vier zusätzlichen Zentralpunkten in der Mitte des Würfels. Jede Einflussgröße ist dabei durch eine Dimension repräsentiert, so dass insgesamt zwölf Experimente für jeden *E. coli* Stamm resultieren.

Bild 7.1: Geometrie des 2^3 Screening-Designs

Mit Hilfe der Zentralpunkte ist es möglich eine Abschätzung für die Reproduzierbarkeit der Ergebnisse bzw. für die Versuchsstreuung zu erhalten, sowie zu prüfen, ob der lineare Modellansatz die Realität ausreichend gut beschreibt (Scheffler, 1997).

Die untere Faktorstufe für die Flüssigphasentemperatur ab Induktion $X_{2\text{low}}$ wurde so gewählt, dass mit 27 °C eine Temperaturdifferenz von 10 °C zur optimalen Wachstumstemperatur von *E. coli* und gleichzeitig oberen Faktorstufe $X_{2\text{high}}$ von 37 °C herrscht. Nach Gleichung 6.7

kann nun der kritische Wert der steuerbaren zellspezifischen Wachstumsrate μ_{wcrit} bei der niedrigen Kultivierungstemperatur von 27 °C mit 0.25 h^{-1} berechnet werden. Die Einstellung für die obere Faktorstufe der zellspezifischen Wachstumsrate im Fed-Batch $X_{1\text{high}}$ wurde unterhalb dieses kritischen Wertes mit 0.22 h^{-1} festgelegt. Die untere Faktorstufe $X_{1\text{low}}$ entspricht dem minimalen Grenzwert μ_{wmin} von 0.08 h^{-1} . Bild 7.2 verdeutlicht die Lage der oberen und unteren Faktorstufen für Temperatur und Wachstumsrate im regulären Versuchsbereich.

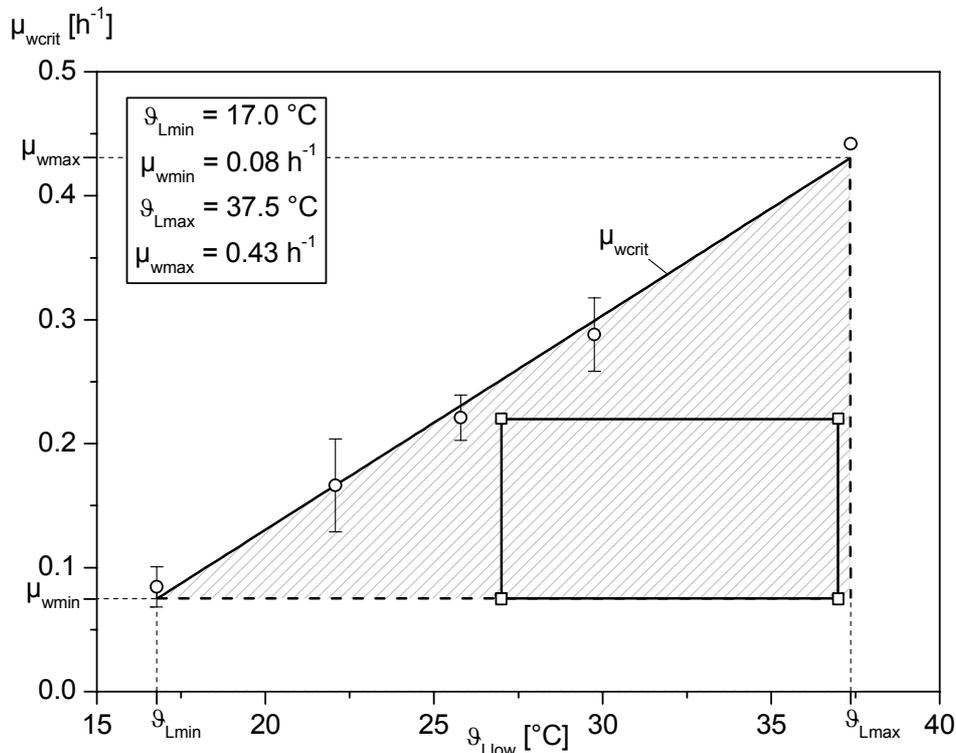


Bild 7.2: Faktoreinstellungen für Temperatur θ_{LW} und Wachstumsrate μ_w im Screening

Bei der Wahl der Faktorstufen für die molare IPTG-Konzentration C_{IPTGw} war die Nähe des Zentralpunkts $X_{3\text{cp}}$ zu der in der Literatur häufig gewählten Konzentration von 1.0 mmol^{-1} wichtig (Donovan *et al.*, 1996). Die untere Stufe $X_{3\text{low}}$ spiegelt die bisher im FTZ Bioprozess- und Analysetechnik zur Induktion der Proteinexpression verwendete IPTG-Konzentration bei *E. coli* Kultivierungen wider. Damit liegen die Stufenparameter für jeden Faktor der Planmatrix des Versuchsplans fest. Tabelle 7.1 gibt eine Übersicht über alle Rahmenbedingungen für die Durchführung der Screening-Versuche.

Tabelle 7.1: Physikalische Grenzwerte und festgelegte Stufenhöhen der Einflussgrößen

Faktor	Variable	Dimension	$X_{j\text{min}}$	$X_{j\text{max}}$	$X_{j\text{low}}$	$X_{j\text{cp}}$	$X_{j\text{high}}$	ΔX_j
X_1	μ_w	h^{-1}	0.08	0.27	0.08	0.15	0.22	0.07
X_2	θ_{LW}	°C	17.0	37.5	27.0	32.0	37.0	5.0
X_3	C_{IPTGw}	mmol^{-1}	0.005	5.0	0.1	0.8	1.5	0.7

7.1.2 Aufstellung der Designmatrix und Bestimmung der Konditionszahl

Die Anwendung von Gleichung 6.5 erlaubt die Transformation der physikalischen Werte der festgelegten Faktorstufen X_j in dimensionslose kodierte Werte x_j . Die untere Stufe der Faktoren wird durch (-1) oder nur (-), die obere Stufe durch (+1) oder nur (+) und der Zentralpunkt durch (0) gekennzeichnet.

Damit kann die Designmatrix \underline{X} ,

$$\underline{X} = \begin{matrix} x_0 & \begin{matrix} x_1 & x_2 & x_3 \\ -1 & -1 & -1 \\ +1 & -1 & -1 \\ +1 & +1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & +1 \\ +1 & -1 & +1 \\ +1 & -1 & +1 \\ +1 & +1 & +1 \\ +1 & 0 & 0 \\ +1 & 0 & 0 \\ +1 & 0 & 0 \\ +1 & 0 & 0 \end{matrix} & \begin{matrix} x_{12} & x_{13} & x_{23} & x_{123} \\ +1 & +1 & +1 & -1 \\ -1 & -1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ +1 & -1 & -1 & -1 \\ +1 & -1 & -1 & +1 \\ -1 & +1 & -1 & -1 \\ -1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} \end{matrix} \quad (7.1)$$

aufgestellt werden, die einerseits die unabhängigen Variablen x_j in normierter Form, andererseits schwarz umrandet die Spaltenvektoren der Planmatrix mit den durchzuführenden Versuchen repräsentiert.

Mit der Konditionszahl der Designmatrix \underline{X} kann die Güte eines Designs vor und nach der experimentellen Durchführung bewertet werden. Formal wird die Konditionszahl κ ,

$$\kappa = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}, \quad (7.2)$$

aus dem Verhältnis von größtem Eigenwert λ_{\max} und kleinstem Eigenwert λ_{\min} der Matrix $\underline{X}^T \cdot \underline{X}$ definiert. Je näher κ an 1 liegt, desto besser konditioniert ist das Design.

Informativ ist dieser Wert das Verhältnis aus längster und kürzester Designdiagonale und ein Maß für die Orthogonalität eines Designs. Wegen der Orthogonalität sind die Glieder des Regresionsmodells unabhängig additiv, d. h. nicht signifikante Glieder dürfen ohne Änderung in den anderen Gliedern gestrichen werden (Scheffler, 1997). Falls bei der Durchführung Experimente fehlschlagen oder Messwerte der Zielgrößen als Ausreißer erkannt werden, ist eine Änderung der Konditionszahl ± 0.5 zulässig (Eriksson *et al.*, 2008). Andernfalls sollte das Design überprüft bzw. sollten Versuche wiederholt werden.

7.1.3 Aufstellung der Planmatrix und Auswertung der Antwortmatrix

Tabelle 7.2 zeigt die randomisierte Planmatrix mit dem Verzeichnis der Einflussgrößen in natürlicher und kodierter Darstellung sowie die Antwortmatrix mit den berechneten Zielgrößen für beide *E. coli* Stämme.

Die Einzelversuche wurden softwareseitig so eingeteilt, dass einerseits in beiden Kultivierungsläufen pro *E. coli* Stamm jede Faktorstufenkombination gleich häufig auftrat sowie andererseits die Plankonstruktion eine Abschätzung des Blockeffekts an Stelle der 3-Faktor-Wechselwirkung ermöglichte. Mit dem Blockeffekt kann die Größe des systematischen Fehlers oder anders ausgedrückt die Konsistenz der Versuchsdurchführung erfasst werden. Des Weiteren wurde die Abfolge der Einzelversuche automatisch auch innerhalb der Blöcke randomisiert, um eine Verfälschung der Effekte durch verbliebene unbekannte, nicht kontrollierbare Veränderungen zu vermeiden (Kleppmann, 2009; Scheffler, 1997).

Tabelle 7.2: Übersicht der randomisierten Planmatrix mit Einflussgrößen in natürlicher und kodierter Darstellung sowie der Antwortmatrix mit den Zielgrößen

Versuch Abfolge (5247/5248)		Faktoren			Antworten MSD5247		Antworten MSD5248	
		μ_w [h ⁻¹]	ϑ_{Lw} [°C]	C_{IPTGw} [mmol l ⁻¹]	RZA_{sol} [RFU h ⁻¹]	RZA_{IB} [RFU h ⁻¹]	RZA_{sol} [RFU h ⁻¹]	RZA_{IB} [RFU h ⁻¹]
1	12/2	0.08 (-)	27 (-)	0.1 (-)	741	8618	1051	8961
2	6/12	0.22 (+)	27 (-)	0.1 (-)	1386	13998	3492	15229
3	1/7	0.08 (-)	37 (+)	0.1 (-)	488	1676	545	1020
4	9/5	0.22 (+)	37 (+)	0.1 (-)	582	2217	1148	4312
5	2/8	0.08 (-)	27 (-)	1.5 (+)	694	5999	1072	7675
6	10/6	0.22 (+)	27 (-)	1.5 (+)	1168	13409	-	-
7	11/1	0.08 (-)	37 (+)	1.5 (+)	587	1752	530	1668
8	5/11	0.22 (+)	37 (+)	1.5 (+)	678	2112	791	3059
9	3/3	0.15 (0)	32 (0)	0.8 (0)	734	5358	777*⁺	8876
10	4/4	0.15 (0)	32 (0)	0.8 (0)	908	6091	1089	8217
11	7/9	0.15 (0)	32 (0)	0.8 (0)	297*⁺	3295*⁺	1154	8280
12	8/10	0.15 (0)	32 (0)	0.8 (0)	845	6117	1227	7776

Fehlende Werte in der Antwortmatrix resultieren aus einem fehlgeschlagenen Experiment. Die Werte in Fettdruck wurden entweder über den Grubbs-Test (*) mit einer Irrtumswahrscheinlichkeit von 0.10, oder über die Auftragung der studentisierten Residuen (†) im Wahrscheinlichkeitsnetz der Normalverteilung als Ausreißer erkannt und nicht für die weitere Auswertung berücksichtigt. Die Konditionszahl der Designmatrix änderte sich nach Entfernen der Ausreißer von 1.225 auf 1.173 für *E. coli* MSD5247 bzw. auf 1.538 für *E. coli* MSD5248. Da die maximal zulässige Änderung von ± 0.5 in beiden Fällen nicht überschritten wird, kann weiterhin von einer Orthogonalität des Designs ausgegangen werden.

7.1.4 Prüfung der Antwortgrößen auf Normalverteilung mit dem W-Test

Zahlreiche statistische Tests und Verfahren setzen eine Normalverteilung der Antwortgrößen voraus. Dazu zählt z. B. auch der in dieser Arbeit verwendete, und im nächsten Kapitel 7.1.5 vorgestellte Ausreißer-Test von Grubbs (Grubbs, 1969). Für eine statistische Bewertung der erhaltenen Ergebnisse aus der linearen Regressionsanalyse wird dagegen eine Normalverteilung der Residuen nach erfolgreicher Modellbildung vorausgesetzt (Hartung, 2005; Sachs and Hedderich, 2009).

Die DoE-Software MODDE bietet keine numerische, sondern nur eine visuelle Prüfung auf Normalverteilung. Antwortgrößen können mit einem Anwothistogramm bzw. Box-Whisker-Plot, Residuen über die Auftragung im Wahrscheinlichkeitsnetz (engl.: probability plot) geprüft werden. Im Anwothistogramm wird die Festlegung der unteren bzw. oberen Klassengrenze sowie die Wahl der Klassenzahl und -breite durch die DoE-Software MODDE definiert und ist nicht vom Benutzer beeinflussbar. Das kann insbesondere bei einer kleinen Stichprobenzahl zu subjektiven Fehlinterpretationen führen. Daher wurde in dieser Arbeit aus der Vielzahl der möglichen statistischen Prüfverfahren der sog. W-Test zur numerischen Prüfung der Antwortgrößen auf Normalverteilung ausgewählt (Shapiro and Wilk, 1965).

Der W-Test basiert auf einer Varianzanalyse der Antwortgrößen. Er weist im Gegensatz zum χ^2 -Test bzw. Kolmogorow-Smirnow-Test eine höhere Teststärke auf und kann bereits bei der Überprüfung von kleineren Stichproben mit $n > 2$ besser jedoch $n > 5$ angewendet werden (Scheffler, 1997; Duller, 2008). Die Grundlage für den W-Test bildet die Berechnung der Testgröße W_{kcal} für die lösliche ($k = 1$) und unlösliche ($k = 2$) Raum-Zeit-Ausbeute,

$$W_{kcal} = \frac{\left(\sum_{i=1}^m a_i \cdot (y_{k(n-i+1)} - y_{k(i)}) \right)^2}{\sum_{i=1}^n (y_{k(i)} - \bar{y}_k)^2}, \quad (7.3)$$

wobei a_i die tabellierten Koeffizienten repräsentiert und $m = n \cdot 2^{-1}$ für eine gerade Anzahl n an Versuchen bzw. $m = (n-1) \cdot 2^{-1}$ für eine ungerade Anzahl an Versuchen gilt.

Die Nullhypothese H_0 geht von einer Normalverteilung der Antwortgrößen aus und wird abgelehnt, wenn die Testgröße W_{kcal} einen kritischen tabellierten Wert $W_{tab}(n, \alpha)$ unterschreitet bzw. diesem entspricht. Die Irrtumswahrscheinlichkeit α , d. h. die größte Wahrscheinlichkeit für eine irrtümliche Ablehnung der H_0 -Hypothese, wurde mit 0.05 festgelegt. Die Umsetzung des Testverfahrens erfolgte mit einem programmierten MATLAB[®]-Skript, in dem die Koeffizienten a_i und die Quantile für den W-Test als Parameter hinterlegt wurden.

Die Prüfung auf Normalverteilung der Antwortgrößen nach Shapiro und Wilk wurde mit und ohne Berücksichtigung der vermutlichen Ausreißer durchgeführt, um eine eventuelle Verfälschung der Teststatistik zu vermeiden. Die Nullhypothese H_0 , dass die Daten aus einer normalverteilten Grundgesamtheit stammen, konnte nach Entfernung der Ausreißer mit einer Ausnahme (*) für alle Antwortgrößen bei $n = 11$ (MSD5247) bzw. $n = 10$ (MSD5248) berechtigt angenommen werden. Tabelle 7.3 zeigt die Ergebnisse für den W-Test.

Mit einer Testgröße W_{1cal} von 0.639 für die RZA des löslichen Proteinanteils bei *E. coli* MSD5248 wurde die Alternativhypothese H_A , wahrscheinlich bedingt durch den außergewöhnlich hohen Messwert $RZA_{sol} = 3492 \text{ RFUh}^{-1}$ bei Versuch 2, abgelehnt, folglich H_0 für die weitere Analyse der Daten unberechtigt angenommen.

Tabelle 7.3: Ergebnisse für die Prüfung auf Normalverteilung der Antwortgrößen mit dem W-Test nach Shapiro und Wilk

<i>E. coli</i> Stamm	n	α	$W_{tab}(n, \alpha)$	W_{1cal}	W_{2cal}
MSD5247	12	0.05	0.859	0.944	0.853
	11		0.850	0.886	0.859
MSD5248	11	0.05	0.850	0.628	0.911
	10		0.842	0.639*	0.916

7.1.5 Prüfung der Antwortgrößen auf Ausreißer mit dem Grubbs-Test

Bevor der untersuchte Prozess durch ein semi-empirisches Modell dargestellt werden kann (induktive Statistik), ist eine sorgfältige Analyse der Rohdaten von großer Bedeutung (deskriptive Statistik). Der Grubbs-Test wurde zur Identifikation von Ausreißern verwendet. Voraussetzung für die Anwendung des Grubbs-Tests ist, dass das Verteilungsgesetz der Zielgröße y_k in der Grundgesamtheit einer Normalverteilung entspricht. Der Test kann bereits bei einer kleinen Anzahl an Stichproben $n \geq 3$ angewendet werden (Scheffler, 1997; Danzer *et al.*, 2001). Der Grubbs-Test prüft die Hypothese H_0 , es gibt keine Ausreißer in den Daten, bzw. die Alternativhypothese H_A , es gibt mindestens einen Ausreißer.

Für die Prüfung der Nullhypothese H_0 wird die Testgröße G_{kcal} für die lösliche ($k = 1$) und unlösliche ($k = 2$) Raum-Zeit-Ausbeute,

$$G_{\text{kcal}(i)} = \frac{|y_{k(i)} - \bar{y}_k|}{SD_k}, \quad (7.4)$$

berechnet, die eine absolute Abweichung eines vermutlichen Ausreißers vom Mittelwert der Stichprobe im Verhältnis zur Standardabweichung SD_k ,

$$SD_k = \sqrt{\frac{\sum_{i=1}^m (y_{k(i)} - \bar{y}_k)^2}{m-1}}, \quad (7.5)$$

mit $m = n - n_{\text{cp}}$ für alle regulären Versuchspunkte bzw. $m = n_{\text{cp}}$ für alle Versuche am Zentralpunkt beschreibt. Die Nullhypothese H_0 kann mit einer Irrtumswahrscheinlichkeit α von 0.10 abgelehnt werden, sobald der Wert der Teststatistik G_{kcal} größer ist als der entsprechende kritische Wert $G_{\text{tab}}(n, \alpha)$.

Tabelle 7.4: Ergebnisse für die Prüfung der Antwortgrößen auf Ausreißer

Versuch	MSD5247			MSD5248		
	$G_{\text{tab}}(n, \alpha)$	$G_{1\text{cal}}$	$G_{2\text{cal}}$	$G_{\text{tab}}(n, \alpha)$	$G_{1\text{cal}}$	$G_{2\text{cal}}$
1	1.909	0.157	0.458	1.828	0.177	0.591
2		1.886	1.487		2.199*	1.837*
3		0.958	0.869		0.669	0.988
4		0.660	0.766		0.082	0.333
5		0.306	0.043		0.156	0.335
6		1.196	1.374		-	-
7		0.645	0.855		0.684	0.859
8		0.356	0.786		0.430	0.583
9	1.425	0.138	0.108	1.425	1.438*	1.303
10		0.769	0.660		0.138	0.155
11		1.448*	1.446*		0.466	0.016
12		0.541	0.679		0.834	1.131

Tabelle 7.4 zeigt die Ergebnisse für die Prüfung der Antwortgrößen auf Ausreißer. Demnach wird die Nullhypothese für alle regulären Versuche bei MSD5247 angenommen, für Versuch 11 am Zentralpunkt jedoch abgelehnt, da die Werte der Teststatistik größer sind als der kritische Wert $G_{\text{tab}}(4, 0.10) = 1.425$.

Bei MSD5248 wurde H_0 für Versuch 2 unberechtigt angenommen. Die Ergebnisse wurden nicht als Fehlmessung, sondern vielmehr als sachlogisch interessant interpretiert, da, wie in Kapitel 6.1.4 beschrieben, erwartet wird, dass die Kombination aus hoher Wachstumsrate, sowie niedriger Temperatur und niedriger Induktorkonzentration in einer erhöhten löslichen RZA resultiert. Versuch 9 am Zentralpunkt wurde nicht nur für die RZA des löslichen Proteinanteils als Ausreißer erklärt, sondern aufgrund der Vorgehensweise bei der Offline-Proteinanalytik sowie aufgrund des nach Entfernung positiven Einflusses auf die Güteparameter des finalen Modells auch für die RZA des unlöslichen Proteinanteils.

Die Umsetzung des Testverfahrens erfolgte wie bei dem W-Test mit einem programmierten MATLAB[®]-Skript, in dem die Quantile als Parameter hinterlegt wurden. Nach der Detektion eines Ausreißers, wurde dieser aus dem Datensatz gelöscht und die verbleibenden Einzelwerte erneut mit dem W-Test nach Shapiro und Wilk auf Normalverteilung geprüft (Wilrich and Henning, 1998).

7.1.6 Statistische Modellbildung über Multiple Lineare Regression (MLR)

Der funktionale Zusammenhang zwischen den analysierten Einflussgrößen und der jeweiligen Raum-Zeit-Ausbeute soll in Form eines empirischen Modells beschrieben werden, das über Regressionsanalyse erhalten wird. Dafür wurden in Kapitel 6.2.1 verschiedene multiple Regressionsmodelle vorgestellt, deren Regressionskoeffizienten über die Lösung des linearen Gleichungssystems,

$$\underline{y}_k = \underline{X} \cdot \underline{\beta}_k + \underline{\varepsilon}_k, \quad (7.6)$$

mit der Methode der kleinsten Quadrate MKQ (engl.: ordinary least squares method, OLS) nach Carl Friedrich Gauß geschätzt werden.

Gleichung (7.6) repräsentiert die Modelle der Gleichungen (6.1) bis (6.3) in Matrixschreibweise. \underline{y}_k ist hierbei der n-dimensionale Zufallsvektor, der die Werte der jeweiligen RZA enthält. Die Designmatrix \underline{X} enthält in jeder Zeile die kodierten Stufen der unabhängigen Variablen, in der ersten Spalte ergänzt durch den Wert 1. Der Spaltenvektor $\underline{\beta}_k$ enthält die zu schätzenden Modellparameter und der n-dimensionale Spaltenvektor $\underline{\varepsilon}_k$ die unkorrelierten Fehlerzufallsvariablen.

Aus Übersichtsgründen wird für die weitere mathematisch-statistische Betrachtung der Index für den löslichen $k = 1$ (sol) und unlöslichen $k = 2$ (IB) zellinternen Proteinanteil vernachlässigt.

Die Methode der kleinsten Quadrate fordert, dass der Parametervektor so zu bestimmen ist, dass die Summe der quadrierten Residuen, auch Fehlerquadratsumme genannt,

$$\sum_{i=1}^n \varepsilon_i^2 = \underline{\varepsilon}^T \cdot \underline{\varepsilon} = (\underline{y} - \underline{X} \cdot \underline{\beta})^T \cdot (\underline{y} - \underline{X} \cdot \underline{\beta}) \rightarrow \min, \quad (7.7)$$

minimiert wird.

Die Lösung dieses Minimierungsproblems durch partielle Differentiation von Gleichung (7.7) nach den Regressionskoeffizienten β ,

$$\left. \frac{\partial \underline{\varepsilon}^T \cdot \underline{\varepsilon}}{\partial \underline{\beta}} \right|_{\beta} = -2 \cdot \underline{X}^T \cdot \underline{y} + 2 \cdot \underline{X}^T \cdot \underline{X} \cdot \underline{\hat{\beta}} = 0, \quad (7.8)$$

führt auf die sogenannten Least-Square Normalgleichungen,

$$\underline{X}^T \cdot \underline{X} \cdot \underline{\hat{\beta}} = \underline{X}^T \cdot \underline{y}, \quad (7.9)$$

aus denen der erwartungstreue Schätzvektor für die Modellparameter,

$$\underline{\hat{\beta}} = (\underline{X}^T \cdot \underline{X})^{-1} \cdot \underline{X}^T \cdot \underline{y}, \quad (7.10)$$

berechnet werden kann.

Es wird deutlich, dass die geschätzte Modellfunktion,

$$\underline{\hat{y}} = \underline{X} \cdot \underline{\hat{\beta}} = \underline{H} \cdot \underline{y}, \quad (7.11)$$

linear von den beobachteten Werten abhängt, wobei H ,

$$\underline{H} = \underline{X} \cdot (\underline{X}^T \cdot \underline{X})^{-1} \cdot \underline{X}^T, \quad (7.12)$$

die sog. „hat“-Matrix¹ bezeichnet.

Bei der linearen Regression haben die Antwortwerte an den äußeren Rändern einen stärkeren Einfluss auf die Steigung der Regressionsgeraden als die Antwortwerte in der Mitte. Man spricht von einer stärkeren „Hebelwirkung“² dieser Antwortwerte. Die Hauptdiagonale der hat-Matrix besteht aus den sog. Leverage-Punkten (Verstärkungspunkten, Hebelpunkten), die den Einfluss einer jeden Beobachtung auf den Erwartungswert für diese Beobachtung messen (Hoaglin and Welsch, 1978; Cleff, 2008).

¹ hat-Matrix - dt.: Dach- oder Hut-Matrix

² Hebelwirkung - engl.: leverage

7.1.7 Prüfung der Residuen auf Normalverteilung und Ausreißer

Beim Least-Square Verfahren bleibt der wahre Fehlerterm ε unbekannt und kann nur geschätzt werden. An Stelle des Spaltenvektors $\underline{\varepsilon}$ mit den nicht beobachtbaren Fehlerzufallsvariablen aus Gleichung (7.6), wird dafür zunächst der Residuenvektor,

$$\underline{e} = \underline{y} - \underline{\hat{y}}, \quad (7.13)$$

aus der Abweichung der beobachteten Antwort von der geschätzten Antwort berechnet. Mit Hilfe der Varianzanalyse³ kann dann zur Fehlerabschätzung der Regression die Varianz bzw. die Standardabweichung der Residuen herangezogen werden. Die Residuen-Standardabweichung ist als Gütemaß der Regression auch als Wurzel des mittleren quadratischen Fehlers RMSE⁴ bekannt.

Nach erfolgter Modellbildung über MLR können die Residuen aus Gleichung (7.13) in ihrer natürlichen Dimension im Wahrscheinlichkeitsnetz der Normalverteilung aufgetragen werden. Allerdings wird zur Prüfung auf Normalverteilung bzw. Ausreißererkenung entweder eine dimensionslose „standardisierte“ Darstellung der Residuen,

$$e_{\text{sta}(i)} = \frac{y_i - \hat{y}_i}{SD_{\text{RES}}}, \quad (7.14)$$

bzw. eine dimensionslose „studentisierte“ Darstellung der Residuen,

$$e_{\text{stu}(i)} = \frac{y_i - \hat{y}_i}{\left[SD_{\text{RES}} \cdot \sqrt{1 - h_{ii}} \right]}, \quad (7.15)$$

bevorzugt, d. h. die Werte wurden durch die Residuen-Standardabweichung SD_{RES} ,

$$SD_{\text{RES}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - r}}, \quad (7.16)$$

geteilt und in Gleichung (7.15) zusätzlich um das „leverage“ korrigiert. Die Elemente h_{ii} entstammen der Hauptdiagonalen der hat-Matrix aus Gleichung (7.12).

Falls der Fehlervektor $\underline{\varepsilon}$ normalverteilt ist, sollten die standardisierten bzw. studentisierten Restfehler zufällig um den erwarteten Mittelwert Null verteilt sein, d. h. ca. 68.3 % der Residuen sollten im Limit ± 1 , ca. 95.5 % im Limit ± 2 und virtuell alle im Limit ± 3 liegen. Die Grenze für die zuverlässige Detektion eines Ausreißers wird einerseits mit $-3 \leq e_{\text{sta/stu}} \leq +3$ (Montgomery, 2009) andererseits mit $-4 \leq e_{\text{sta/stu}} \leq +4$ (Eriksson *et al.*, 2008) vorgegeben.

³ Varianzanalyse - engl.: analysis of variance, ANOVA

⁴ RMSE - engl.: root mean square error

Bild 7.3 zeigt beispielhaft die studentisierten Residuen für den löslichen Proteinanteil $e_{1\text{stu}}$ als Teilergebnis der mit dem Originaldatensatz durchgeführten Multiplen Linearen Regression der Antwort für *E. coli* MSD5247 nach Löschung der nicht-signifikanten Terme. Es ist klar zu erkennen, dass der studentisierte Residualwert für den löslichen Proteinanteil in Experiment 11 außerhalb des Grenzbereichs liegt. Wird dieses Experiment gelöscht, ergibt sich für den Rest der normalverteilten Residuen eine Gerade.

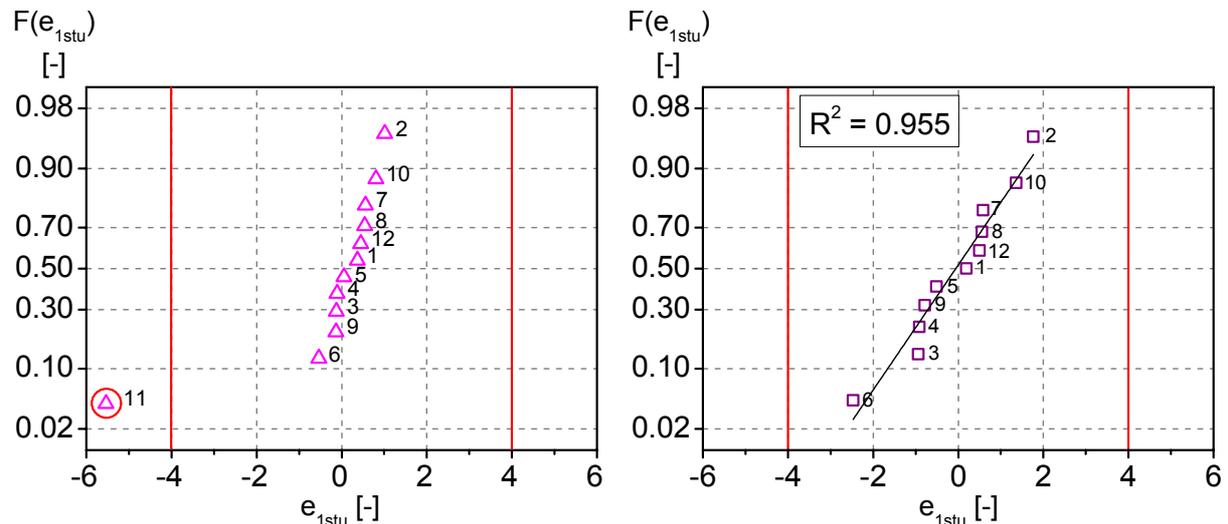


Bild 7.3: Darstellung der studentisierten Residuen für den löslichen Proteinanteil $e_{1\text{stu}}$ bei *E. coli* MSD5247 im Wahrscheinlichkeitsnetz der Normalverteilung

7.1.8 Identifikation statistisch signifikanter Faktoren

Das Ergebnis der Regressionsanalyse ist die mathematische Beschreibung der Ursache-Wirkung-Beziehung in Form eines empirischen Modells, das als eine Näherung nur statistisch signifikante Glieder enthalten soll. Die Signifikanz der Modellterme kann über das Konfidenzintervall der jeweiligen Regressionskoeffizienten β_i oder über den t-Test bzw. der darauf basierenden p-Wert Methode identifiziert werden (Eriksson *et al.*, 2008; Montgomery, 2009; Scheffler, 1997).

Bei den Methoden wird die Alternativhypothese H_A geprüft, der Regressionskoeffizient und damit der Effekt ist signifikant ($\beta_i \neq 0$), bzw. die Nullhypothese H_0 , der Einfluss des Regressionskoeffizienten auf die Zielgröße ist nur zufällig und damit vernachlässigbar ($\beta_i = 0$).

Für ein zweiseitiges Konfidenzintervall werden die Konfidenzgrenzen,

$$\hat{\beta}_i - t_{\alpha/2, n-r} \cdot \text{SE}_{\text{COE}} \leq \beta_i \leq \hat{\beta}_i + t_{\alpha/2, n-r} \cdot \text{SE}_{\text{COE}}, \quad (7.17)$$

mit Hilfe der t-Verteilung ($\alpha = 0.05$) berechnet.

Der Standardfehler der Koeffizienten (engl.: standard error of regression coefficients) SE_{COE} ,

$$SE_{COE} = \sqrt{MS_{RES} \cdot c_{ii}} = SD_{RES} \cdot \sqrt{c_{ii}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - r}} \cdot c_{ii}, \quad (7.18)$$

wird aus der Varianz MS_{RES} bzw. Standardabweichung SD_{RES} der Residuen und dem jeweiligen Element c_{ii} der Hauptdiagonalen der Matrix $(\underline{X}^T \cdot \underline{X})^{-1}$ definiert.

Mit der Wahrscheinlichkeit $1 - \alpha$, dem Konfidenzniveau, liegt der wahre Wert innerhalb bzw. mit der Irrtumswahrscheinlichkeit α , dem Signifikanzniveau, außerhalb des Konfidenzintervalls. Schließt das Konfidenzintervall den Wert Null ein, kann der wahre Effekt auch Null, d. h. nicht vorhanden sein (Annahme der H_0). Daher ist die Grenze des Konfidenzintervalls auch gleichzeitig die Signifikanzgrenze.

Neben dem Konfidenzintervall kann der t-Test für jeden Modellparameter Auskunft über die Signifikanz geben. Für die Prüfung der Alternativhypothese H_A wird die Testgröße t_{cal} ,

$$t_{cal(i)} = \left| \frac{\hat{\beta}_i}{SE_{COE}} \right|, \quad (7.19)$$

berechnet, die einzelnen Regressionskoeffizienten mit ihrem Standardfehler ins Verhältnis setzt. Die Nullhypothese H_0 kann mit einer Irrtumswahrscheinlichkeit α von 0.05 abgelehnt werden, sobald der Wert der Teststatistik t_{kcal} größer ist als der entsprechende kritische Wert $t_{\alpha/2, n - r - 1}$.

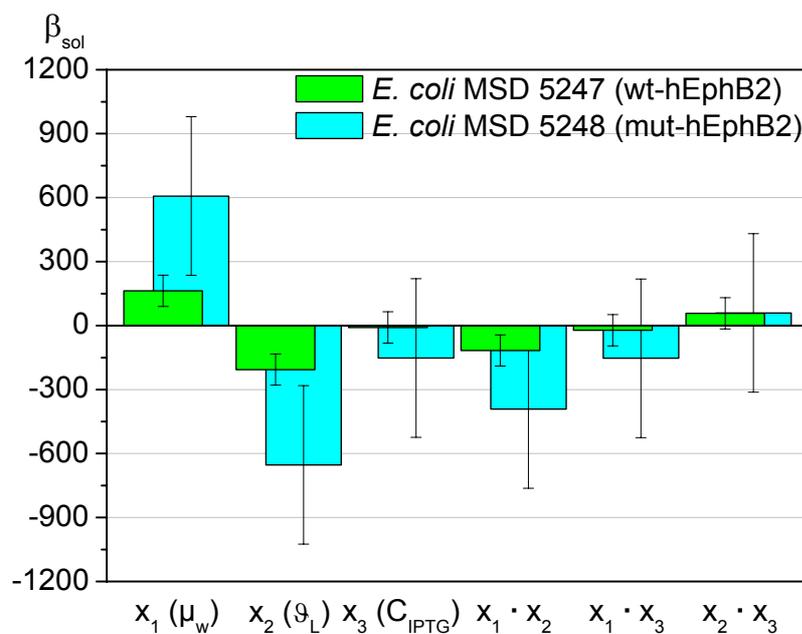
Sowohl die Beurteilung der Signifikanz über das Konfidenzintervall der jeweiligen Regressionskoeffizienten β_i als auch über den t-Test kann bei unterschiedlich festgelegten Irrtumswahrscheinlichkeiten zu Zweideutigkeiten führen. Daher ist es üblich, anstatt der berechneten Prüfgröße t_{cal} den darauf basierenden p-Wert anzugeben, der als das kleinste Signifikanzniveau betrachtet werden kann, das zu einer Ablehnung der Nullhypothese führen würde. Ist der p-Wert kleiner oder gleich der allgemein üblich vorgegebenen Irrtumswahrscheinlichkeit α von 0.05, so wird H_0 verworfen, ansonsten behält man H_0 bei. Eine Übersicht über die Interpretationsmöglichkeiten verschiedener p-Werte gibt Tabelle 7.5.

Bei der Aufstellung der Regressionsmodelle wurde iterativ vorgegangen, d. h. nicht signifikante Modellterme wurden schrittweise, beginnend mit dem jeweils kleinsten nicht signifikanten Modellterm, eliminiert. Nach jedem Schritt wurde die Auswirkung auf das korrigierte Bestimmtheitsmaß R_{adj}^2 bzw. insbesondere auf die Prognosegüte Q^2 und die Modellvalidität MV bzw. den Lack of Fit beurteilt. Das iterative Vorgehen hat gezeigt, dass weder der Blockeffekt, noch die damit vermengte 3-Faktor-Wechselwirkung signifikant sind – der p-Wert war stets > 0.10 .

Tabelle 7.5: Interpretation von p-Werten (verändert nach Mendenhall *et al.*, 2009)

p-Wert	Alternativhypothese	Bedeutung
≤ 0.001	Annahme	höchst signifikant
≤ 0.01	Annahme	sehr signifikant
≤ 0.05	Annahme	signifikant
≤ 0.10	i. Allg. Ablehnung	nur Tendenz zur Signifikanz
> 0.10	Ablehnung	nicht signifikant

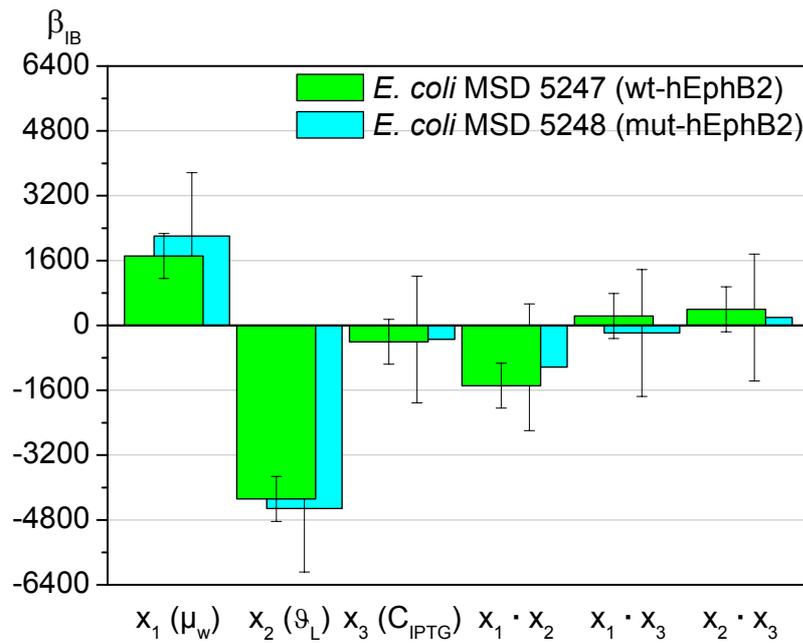
Bild 7.4 sowie Bild 7.5 zeigt einen Vergleich der Regressionskoeffizienten für das Modell der löslichen und unlöslichen Proteinexpression bei *E. coli* MSD5247 und MSD5248. Die Koeffizienten sind skaliert und zentriert mit einem Konfidenzniveau von 95 % aufgetragen.



Es ist klar zu erkennen, dass die linearen Terme für die zellspezifische Wachstumsrate μ_w (β_1) und die Flüssigphasentemperatur (β_2) sowie ihre Zwei-Faktor-Interaktion (β_{12}) über den linearen Term für die Induktorkonzentration C_{IPTG} (β_3) und die anderen Terme höherer Ordnung dominieren.

Bild 7.4: Vergleich der Regressionskoeffizienten für die lösliche Proteinexpression

Daher wurden beide Zwei-Faktor-Interaktionen, die mit der molaren IPTG-Konzentration korrelieren (β_{13} und β_{23}) sowie der zugehörige lineare C_{IPTG} -Term aus dem Modell gelöscht. Sie wurden als statistisch nicht-signifikant betrachtet, da ihr zugehöriger p-Wert größer als 0.05 ist und ihr Konfidenzintervall den Wert 0 einschließt.



Des Weiteren ist zu erkennen, dass die Terme für die unlösliche Proteinexpression beider *E. coli* Stämme eine annähernd gleiche Größenordnung aufweisen, während die Terme für die lösliche Proteinexpression für den Stamm MSD5248 ca. drei mal höher sind.

Bild 7.5: Vergleich der Regressionskoeffizienten für die unlösliche Proteinexpression

Dies weist darauf hin, dass die in die Aminosäuresequenz des Zielproteins eingefügte Punktmutation [L767P] einen positiven Effekt auf die lösliche Proteinexpression hat. Um hier objektiv einen statistisch signifikanten Unterschied feststellen zu können, wird in Kapitel 7.1.11 ein t-Test verwendet.

Damit liegen die MLR-Modelle für die lösliche und unlösliche Raum-Zeit-Ausbeute für *E. coli* MSD5247,

$$\hat{y}_1 = 801 + 163 \cdot x_1 - 206.8 \cdot x_2 - 116.8 \cdot x_1 \cdot x_2, \quad (7.20)$$

$$\hat{y}_2 = 6122 + 1711 \cdot x_1 - 4283 \cdot x_2 - 1486 \cdot x_1 \cdot x_2, \quad (7.21)$$

bzw. *E. coli* MSD5248,

$$\hat{y}_1 = 1401 + 692.7 \cdot x_1 - 738.7 \cdot x_2 - 476.7 \cdot x_1 \cdot x_2, \quad (7.22)$$

$$\hat{y}_2 = 7746 + 2374 \cdot x_1 - 4690 \cdot x_2 - 1203 \cdot x_1 \cdot x_2, \quad (7.23)$$

fest.

Die gesamten Ergebnisse der Regressionsanalyse sind in Tabelle 7.6 wiedergegeben, wobei die linke Spalte den Wert vor, die rechte Spalte den Wert nach der Entfernung nicht-signifikanter Modellterme (*) darstellt. Die Koeffizienten sind skaliert und zentriert dargestellt.

Tabelle 7.6: Ergebnisse der Regressionsanalyse für das Screening

Term	lösliche Proteinexpression <i>E. coli</i> MSD 5247						unlösliche Proteinexpression <i>E. coli</i> MSD 5247					
	Koeffizient	Standardfehler	p-Wert	Konf.intervall	Konf.intervall		Koeffizient	Standardfehler	p-Wert	Konf.intervall	Konf.intervall	
β_0	801.0	22.50	<0.001	62.48	62.12		6122	170.3	<0.001	473.0	557.3	
β_1	163.0	26.39	0.003	73.27	72.84		1711	199.7	0.001	554.6	653.5	
β_2	-206.8	26.39	0.001	73.27	72.84		-4283	199.7	<0.001	554.6	653.5	
β_3	-8.750	26.39	0.757*	73.27*	-		-404.6	199.7	0.113*	554.6*	-	
β_{12}	-116.8	26.39	0.011	73.27	72.84		-1486	199.7	0.002	554.6	653.5	
β_{13}	-21.75	26.39	0.456*	73.27*	-		231.1	199.7	0.312*	554.6*	-	
β_{23}	57.50	26.39	0.095*	73.27*	-		397.4	199.7	0.118*	554.6*	-	
Term	lösliche Proteinexpression <i>E. coli</i> MSD 5248						unlösliche Proteinexpression <i>E. coli</i> MSD 5248					
	Koeffizient	Standardfehler	p-Wert	Konf.intervall	Konf.intervall		Koeffizient	Standardfehler	p-Wert	Konf.intervall	Konf.intervall	
β_0	1339	93.41	<0.001	297.3	188.9		7322	393.6	<0.001	1253	635.5	
β_1	607.3	116.8	0.014	371.6	227.4		2203	492.0	0.021	1566	765.2	
β_2	-653.3	116.8	0.011	371.6	227.4		-4519	492.0	0.003	1566	765.2	
β_3	-152.2	116.8	0.283*	371.6*	-		-346.5	492.0	0.532*	1566*	-	
β_{12}	-391.3	116.8	0.044	371.6	227.4		-1032	492.0	0.127	1566	765.2	
β_{13}	-153.8	116.8	0.280*	371.6*	-		-187.0	492.0	0.729*	1566*	-	
β_{23}	59.21	116.8	0.647*	371.6*	-		195.2	492.0	0.718*	1566*	-	

7.1.9 Mehrfaktorielle Varianzanalyse und Signifikanztest der Regressionsmodelle

Als wichtigstes Instrument zur Auswertung von Experimenten dient die Varianzanalyse, da viele der Prüfverfahren bzw. Gütekriterien für die Regression ebenfalls das Grundprinzip der ANOVA, die sog. Streuungszerlegung, nutzen (Backhaus *et al.*, 2003).

Bild 7.6 zeigt eine Übersicht der nachfolgend erläuterten Zerlegung der um den Mittelwert korrigierten Gesamtabweichung in die beiden Komponenten erklärte und nicht erklärte Abweichung bei insgesamt n durchgeführten Versuchen, bzw. bei f durchgeführten Versuchen (Anzahl der unabhängigen Faktorstufen) mit einfacher Zählung von c Mehrfachrealisierungen an einem Versuchspunkt und r zu bestimmenden Koeffizienten.

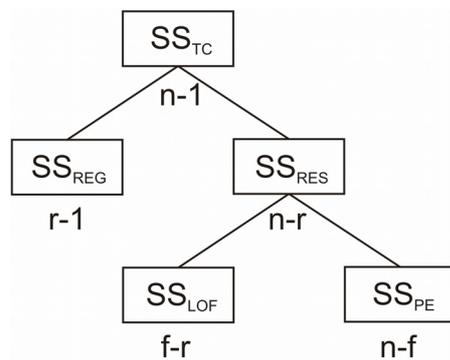


Bild 7.6: Zusammenhang zwischen den Quadratsummen SS und ihren zugehörigen Freiheitsgraden DF

Als Maß für die Abweichungen werden die jeweiligen Quadratsummen (engl.: sum of squares, SS) herangezogen. Jeder Quadratsumme kann eine Anzahl an Freiheitsgraden (engl.: degrees of freedom, DF), d. h. eine Anzahl an frei bestimmbar Summanden, zugeordnet werden.

Aus den mittleren quadratischen Abweichungen (engl.: mean sum of squares, MS), d. h. der Division der Quadratsummen durch die Freiheitsgrade werden schließlich erwartungstreue Schätzungen der Varianzen σ^2 erhalten, aus deren Wurzelbildung dann die Standardabweichungen (engl.: standard deviation, SD) als verständlichere Interpretation hervorgehen.

Die Summe der quadrierten Gesamtabweichungen der Antwortwerte vom Mittelwert (engl.: total sum of squares corrected for the mean) SS_{TC} ,

$$SS_{TC} = \sum_{i=1}^n (y_i - \bar{y})^2 = (\underline{y} - \bar{y})^T \cdot (\underline{y} - \bar{y}), \quad (7.24)$$

mit $n-1$ Freiheitsgraden wird in eine Quadratsumme für die erklärten Abweichungen der Modell- bzw. Regressionswerte vom Mittelwert (engl.: regression sum of squares) SS_{REG} ,

$$SS_{REG} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = (\underline{\hat{y}} - \bar{y})^T \cdot (\underline{\hat{y}} - \bar{y}), \quad (7.25)$$

mit $r-1$ Freiheitsgraden und in eine Quadratsumme für die nicht erklärten Abweichungen der Antwortwerte von den Modellwerten (engl.: residual/error sum of squares) SS_{RES} ,

$$SS_{\text{RES}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = (\underline{y} - \underline{\hat{y}})^T \cdot (\underline{y} - \underline{\hat{y}}) = \underline{e}^T \cdot \underline{e}, \quad (7.26)$$

mit $n-r$ Freiheitsgraden zerlegt.

Mit Hilfe der verschiedenen Freiheitsgrade kann nun aus den Gleichungen (7.24) bis (7.26) die Gesamtvarianz (engl.: total sum of squares) MS_{TC} ,

$$MS_{\text{TC}} = \frac{SS_{\text{TC}}}{DF_{\text{TC}}} = \frac{SS_{\text{TC}}}{n-1}, \quad (7.27)$$

die auf die Regression bezogene Varianz (engl.: regression sum of squares) MS_{REG} ,

$$MS_{\text{REG}} = \frac{SS_{\text{REG}}}{DF_{\text{REG}}} = \frac{SS_{\text{REG}}}{r-1}, \quad (7.28)$$

und die Restvarianz (engl.: residual/error sum of squares) MS_{RES} ,

$$MS_{\text{RES}} = \frac{SS_{\text{RES}}}{DF_{\text{RES}}} = \frac{SS_{\text{RES}}}{n-r}, \quad (7.29)$$

berechnet werden.

Je größer die erklärte bzw. je kleiner die nicht erklärte Abweichung, desto besser ist das Regressionsmodell. Mit Hilfe des F-Tests, der die beiden Varianzen MS_{REG} und MS_{RES} miteinander vergleicht,

$$F_{\text{REG}} = \frac{MS_{\text{REG}}}{MS_{\text{RES}}}, \quad (7.30)$$

kann eine erste Aussage über die Gültigkeit bzw. Qualität der Modelle getroffen werden. Sofern der kritische F-Wert $F_{\text{tab}}(DF_{\text{REG}}, DF_{\text{RES}}, \alpha)$ größer als die Prüfgröße F_{REG} ist, kann die Nullhypothese H_0 abgelehnt werden, da sich die beiden Varianzen signifikant unterscheiden. Ist die zugehörige Wahrscheinlichkeit p , dass die beiden Varianzen aus der gleichen Verteilung stammen kleiner oder gleich der allgemein üblich vorgegebenen Irrtumswahrscheinlichkeit α von 0.05, so wird H_0 verworfen, ansonsten behält man H_0 bei.

Das erhaltene Regressionsmodell soll adäquat oder anders ausgedrückt mit den Versuchsergebnissen konsistent sein, d. h. es soll keine signifikanten Abweichungen geben. In Gleichung (7.26) wurde die nicht erklärte Reststreuung aus den Abweichungen zwischen Experimental- und Modellwerten berechnet. Beide Werte können fehlerbehaftet sein, die Experimentalwerte durch die Versuchsstreuung und die Modellwerte durch den Modellfehler oder den sog. Anpassungsdefekt (engl.: lack of fit), d. h. durch eine gewisse Inadäquatheit des Modells (Scheffler, 1997).

Sofern Mehrfachrealisierungen an einem Versuchspunkt, meistens dem Zentralpunkt (engl.: center point, cp) eines Designs vorliegen, kann die Residualquadratsumme der nicht erklärten Abweichung aus Gleichung (7.26) SS_{RES} in eine reine Zufallsstreuung (engl.: pure error sum of squares) SS_{PE} ,

$$SS_{PE} = \sum_{i=1}^{n_{cp}} (y_{cp(i)} - \bar{y}_{cp})^2 = (\underline{y}_{cp} - \bar{y}_{cp})^T \cdot (\underline{y}_{cp} - \bar{y}_{cp}), \quad (7.31)$$

mit $n-f$ Freiheitsgraden und in einen Lack of Fit (engl.: sum of squares for lack of fit) SS_{LOF} ,

$$SS_{LOF} = \sum_{i=1}^{n-n_{cp}} (\bar{y}_i - \hat{y}_i)^2 + \sum_{i=n-n_{cp}+1}^{n_{cp}} (\bar{y}_{cp} - \hat{y}_i)^2, \quad (7.32)$$

mit $f-p$ Freiheitsgraden zerlegt werden.

Da die Zufallsstreuung aus den mehrfachen Realisierungen am Zentralpunkt des faktoriellen Designs resultiert, ist sie eine modellunabhängige Größe und wird auch Wiederholungsfehler (engl.: replicate error) genannt. Ist die Versuchsstreuung klein, reichen etwa drei bis fünf Realisierungen an einem Versuchspunkt aus, um diese sicher genug zu ermitteln und den Genauigkeitsanforderungen nachkommen zu können. Aus den Abweichungen der Mittelwerte von der Regressionsgeraden entsteht der Anpassungs- oder Modellfehler (engl.: model error) (Eriksson *et al.*, 2008).

Über einen Vergleich der beiden Varianzen MS_{PE} ,

$$MS_{PE} = \frac{SS_{PE}}{DF_{PE}} = \frac{SS_{PE}}{n-f}, \quad (7.33)$$

und MS_{LOF} ,

$$MS_{LOF} = \frac{SS_{LOF}}{DF_{LOF}} = \frac{SS_{LOF}}{f-r}, \quad (7.34)$$

können die erhaltenen Modelle während bzw. nach der iterativen Eliminierung von einzelnen Gliedern auf Adäquatheit mit dem Lack of Fit-Test geprüft werden.

Für die Prüfung der Nullhypothese H_0 , der Anpassungsdefekt ist signifikant kleiner als die Versuchsstreuung, wird die Testgröße F_{LOF} ,

$$F_{LOF} = \frac{MS_{LOF}}{MS_{PE}}, \quad (7.35)$$

als Quotient der entsprechenden Varianzen berechnet und gegen einen kritischen F-Wert $F_{tab}(DF_{LOF}, DF_{PE}, \alpha)$ verglichen.

Ist $F_{\text{tab}} < F_{\text{LOF}}$, so sind offensichtlich Modellterme zu wenig enthalten und das Regressionspolynom spiegelt die Realität nicht adäquat wider. Ist $F_{\text{tab}} > F_{\text{LOF}}$, ist die systematische Abweichung signifikant kleiner als die Streuung und die Nullhypothese somit erfüllt. Mit dem Risiko $\alpha = 0.05$ kann die H_0 irrtümlich abgelehnt und das Modell irrtümlich als nicht adäquat angesehen werden. Sofern der zugehörige p-Wert p_{LOF} größer als die vorgegebene Irrtumswahrscheinlichkeit α ist, liegt im Vergleich nur ein kleiner Modellfehler und somit kein lack-of-fit vor.

Die Ergebnisse der Varianzanalyse sind in Tabelle 7.7 wiedergegeben. In Klammern sind die kritischen F-Werte für ein Signifikanzniveau von 5 % mit den entsprechenden Freiheitsgraden nach der F-Tabellierung angegeben. Da die oberen F-Werte stets größer als die tabellierten kritischen F-Werte sind, wird die Erklärungsleistung der Regressionsmodelle als ein Resultat akzeptiert, das nicht zufallsbestimmt ist. Die Nullhypothese, dass Regressionsvarianz und Restvarianz aus der gleichen Verteilung stammen, wird für alle vier Modelle abgelehnt, da die Wahrscheinlichkeit p stets kleiner als die vorgegebene Irrtumswahrscheinlichkeit α von 0.05 ist.

Die unteren F-Werte sind stets kleiner als die tabellierten kritischen F-Werte, so dass die verbliebenen signifikanten linearen bzw. Interaktionsterme auf adäquate Regressionsmodelle hinweisen. Für *E. coli* MSD5247 liegt der p-Wert weit über dem festgelegten Signifikanzniveau von 5 %, der Modellfehler ist somit signifikant kleiner als die Versuchsstreuung. Für *E. coli* MSD5248 liegt der p-Wert jedoch sowohl für die lösliche als auch die unlösliche Proteinexpression nah an der vorgegebenen Irrtumswahrscheinlichkeit α . Der lack-of-fit Test ist in beiden Fällen zwar bestanden, dennoch ist die Tendenz zu einer gewissen Inadäquatheit der Modelle gegeben.

Tabelle 7.7: ANOVA für das Screening

		lösliche Proteinexpression <i>E. coli</i> MSD 5247					unlösliche Proteinexpression <i>E. coli</i> MSD 5247					
	DF	SS·10 ³ [RFUh ⁻¹]	MS·10 ³ [RFUh ⁻¹]	SD·10 ³ [RFUh ⁻¹]	F-Wert (F _{tab}) [-]	p-Wert	DF	SS·10 ³ [RFUh ⁻¹]	MS·10 ³ [RFUh ⁻¹]	SD·10 ³ [RFUh ⁻¹]	F-Wert (F _{tab}) [-]	p-Wert
Regress.	3	663.6	221.2	0.470	29.14 (4.35)	< 0.001	3	187877	62626	7.914	102.5 (4.35)	< 0.001
Residual	7	53.13	7.590	0.087	-	-	7	4277	611.0	0.782	-	-
Lack of Fit	5	37.61	7.522	0.087	0.969 (19.30)	0.578	5	3906	781.2	0.884	4.207 (19.30)	0.203
Pure Error	2	15.52	7.761	0.088	-	-	2	371.4	185.7	0.431	-	-
Total Corrected	10	716.7	71.67	0.268	-	-	10	192154	19215	4.384	-	-
		lösliche Proteinexpression <i>E. coli</i> MSD 5248					unlösliche Proteinexpression <i>E. coli</i> MSD 5248					
	DF	SS·10 ³ [RFUh ⁻¹]	MS·10 ³ [RFUh ⁻¹]	SD·10 ³ [RFUh ⁻¹]	F-Wert (F _{tab}) [-]	p-Wert	DF	SS·10 ³ [RFUh ⁻¹]	MS·10 ³ [RFUh ⁻¹]	SD·10 ³ [RFUh ⁻¹]	F-Wert (F _{tab}) [-]	p-Wert
Regress.	3	6018	2006	1.416	35.82 (4.76)	< 0.001	3	157436	52479	7.244	82.78 (4.76)	< 0.001
Residual	6	336.0	56.01	0.237	-	-	6	3804	634.0	0.796	-	-
Lack of Fit	4	326.5	81.63	0.286	17.13 (19.25)	0.056	4	3653	913.3	0.956	12.11 (19.25)	0.078
Pure Error	2	9.533	4.766	0.069	-	-	2	150.8	75.41	0.275	-	-
Total Corrected	9	6354	706.0	0.840	-	-	9	161240	17916	4.233	-	-

7.1.10 Gütemaße der Multiplen Linearen Regression

Nachdem die Regressionsfunktion geschätzt wurde, ist deren Güte zu überprüfen, d. h. es ist zu klären, wie gut sie als Modell der Realität geeignet ist. Die Varianzanalyse bildet die Basis zur Konstruktion von Gütemaßen.

Rechnerisch wird die Stärke einer linearen Abhängigkeit zweier Größen durch den Korrelationskoeffizienten R von Pearson beschrieben (Pearson, 1909). Das Quadrat des Korrelationskoeffizienten wird auch als Bestimmtheitsmaß R^2 bezeichnet und misst die Güte der Anpassung der Regressionsfunktion an die empirischen Daten (engl.: goodness of fit).

Aus dem Verhältnis von erklärter Streuung aus Gleichung (7.25) zur Gesamtstreuung aus Gleichung (7.24) ergibt sich das Bestimmtheitsmaß R^2 ,

$$R^2 = \frac{SS_{\text{REG}}}{SS_{\text{TC}}}, \quad (7.36)$$

mit einem normierten Wertebereich zwischen minimal Null und im anderen Extremfall, wenn die gesamte Streuung erklärt wird, maximal Eins.

Das Bestimmtheitsmaß kann auch durch Subtraktion des Verhältnisses der nicht erklärten Streuung aus Gleichung (7.26) zur Gesamtstreuung aus Gleichung (7.24) vom Maximalwert,

$$R^2 = 1 - \frac{SS_{\text{RES}}}{SS_{\text{TC}}}, \quad (7.37)$$

ermittelt werden.

Aus Gleichung (7.37) wird deutlich, dass das Minimierungskriterium zur Schätzung der Regressionsbeziehung aus der Methode der kleinsten Quadrate gleichbedeutend mit der Maximierung des Bestimmtheitsmaßes ist.

Für ein grobes aber brauchbares und stabiles Modell wird R^2 mit einem Wert von 0.75 vorgegeben (Eriksson *et al.*, 2008). Ein großer Wert für R^2 impliziert jedoch nicht notwendigerweise ein gutes Regressionsmodell, da durch das Hinzufügen von Modelltermen, egal ob statistisch signifikant oder nicht, ein mehr oder weniger großer Erklärungsanteil hinzugefügt und somit R^2 automatisch erhöht wird. Insbesondere bei einer kleinen Anzahl von Freiheitsgraden aber verschlechtern sich mit der Zahl der Modellterme die Schätzeigenschaften des Modells (Backhaus *et al.*, 2003; Montgomery, 2009).

Diesen Sachverhalt berücksichtigend wird daher in vielen Fällen ein korrigiertes Bestimmtheitsmaß (engl.: adjusted R^2),

$$R_{\text{adj}}^2 = 1 - \left(\frac{DF_{\text{RES}}}{DF_{\text{TC}}} \right) \cdot (1 - R^2) = 1 - \frac{MS_{\text{RES}}}{MS_{\text{TC}}}, \quad (7.38)$$

bevorzugt, das beim Hinzufügen von statistisch nicht signifikanten Modelltermen auch abnehmen kann. Dieses Maß wird insbesondere bei kleinen Stichprobenumfängen empfohlen.

Das erhaltene Regressionsmodell kann für Vorhersagen durch Intrapolation oder Extrapolation eingesetzt werden. Um die Vorhersagegüte Q^2 zu bestimmen, wird zunächst die Quadratsumme der prognostizierten Residuen PRESS (engl.: prediction error sum of squares),

$$\text{PRESS} = \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{1 - h_{ii}} \right)^2, \quad (7.39)$$

aus der Wichtung der normalen Residuen mit den Elementen h_{ii} der Hauptdiagonalen der hat-Matrix aus Gleichung (7.12) berechnet.

Über das Verhältnis von prognostizierter nicht erklärter Streuung aus Gleichung (7.39) zur Gesamtstreuung aus Gleichung (7.24) wird dann die Prognosegüte (engl.: goodness of prediction) Q^2 ,

$$Q^2 = 1 - \frac{\text{PRESS}}{SS_{\text{TC}}}, \quad (7.40)$$

ermittelt, die im Idealfall nah an R^2 liegt, diesen Wert aber niemals überschreiten kann.

Für eine gute Vorhersage sollte die Differenz von R^2 zu Q^2 nicht größer als 0.2 bis 0.3 sein. Sofern $Q^2 > 0.5$ ist wird von einer guten, sofern $Q^2 > 0.9$ ist, wird von einer exzellenten Vorhersagegüte gesprochen (Eriksson *et al.*, 2008).

Der Lack of Fit Test ist die Basis zur Berechnung der Modellvalidität MV,

$$\text{MV} = 1 + 0.57647 \cdot \log_{10}(p_{\text{LOF}}), \quad (7.41)$$

die als generelle Größe eine weitere Aussage über die Adäquatheit des gewählten Modellsatzes macht. Sofern der p-Wert des Lack of Fit Tests p_{LOF} der gewählten Irrtumswahrscheinlichkeit $\alpha = 0.05$ entspricht, nimmt die Modellvalidität den Wert 0.25 an. Dieser Wert kann somit als untere Grenze für ein adäquates Modell mit geringem Modellfehler angesehen werden (Eriksson *et al.*, 2008).

Über die mehrfachen Realisierungen am Zentralpunkt kann die Reproduzierbarkeit der Versuche durch Subtraktion des Verhältnisses der reinen Zufallsstreuung aus Gleichung (7.31) zur Gesamtstreuung aus Gleichung (7.24) vom Maximalwert,

$$RP = 1 - \frac{MS_{PE}}{MS_{TC}}, \quad (7.42)$$

ermittelt werden. Je größer der Wert ist, desto kleiner ist der Wiederholungsfehler im Vergleich zur Gesamtstreuung. Für eine gute Wiederholbarkeit wird $RP > 0.5$ vorgegeben (Eriksson *et al.*, 2008).

Tabelle 7.8 zeigt die zuvor beschriebenen wichtigen Modellparameter vor und nach der Entfernung nicht-signifikanter Modellterme, wobei im ersten Fall weder der Blockeffekt, noch die damit vermengte 3-Faktor-Wechselwirkung berücksichtigt sind.

Tabelle 7.8: Wichtige Modellparameter vor (linke Spalte) und nach (rechte Spalte) der Entfernung nicht-signifikanter Modellterme

Parameter	Modell für <i>E. coli</i> MSD5247				Modell für <i>E. coli</i> MSD5248			
	\hat{y}_1 (löslich)		\hat{y}_2 (unlöslich)		\hat{y}_1 (löslich)		\hat{y}_2 (unlöslich)	
R^2	0.969	0.926	0.993	0.978	0.971	0.947	0.980	0.976
R^2_{adj}	0.922	0.894	0.983	0.968	0.913	0.921	0.939	0.965
Q^2	0.727	0.797	0.854	0.929	-0.044	0.576	0.265	0.853
MV	0.909	0.863	0.691	0.601	0.090	0.278	0.060	0.360
RP	0.892	0.892	0.990	0.990	0.993	0.993	0.996	0.996

Es ist zu erkennen, dass das Bestimmtheitsmaß sowohl für die lösliche als auch für die unlösliche Proteinexpression bei beiden *E. coli* Stämmen stets größer als 0.9 ist, so dass stets mehr als 90 % der Gesamtvariation durch die Modelle erklärt wird. Das Modell für die unlösliche Proteinexpression bei *E. coli* MSD5247 weist mit $Q^2 = 0.929$ eine exzellente Prognosegüte auf, mit den anderen Modellen können gute Vorhersagen mit einer geringen prognostizierten Variation getroffen werden. Die exzellente Reproduzierbarkeit der Experimente weist auf eine robuste und wiederholbare Prozessführung hin.

Die Differenz von R^2 zu Q^2 ist für das Modell der löslichen Proteinexpression bei *E. coli* MSD5248 mit 0.371 größer als 0.3, auch ist die Modellvalidität nur knapp größer als 0.25. Wie bereits bei der Varianzanalyse durch den lack-of-fit Test in Kapitel 7.1.9 bemerkt, zeigt das Modell eine gewisse Tendenz zur Inadäquatheit. Dies kann insbesondere dadurch erklärt werden, dass bedingt durch ein fehlgeschlagenes Experiment (6) wichtige Antwortwerte bei hoher Wachstumsrate und niedriger Temperatur für die Faktoreinstellungen ($\mu_w = 0.22 \text{ h}^{-1}$, $\vartheta_{LW} = 27 \text{ °C}$, $C_{IPTGW} = 1.5 \text{ mmol l}^{-1}$) fehlen.

Bild 7.7 zeigt die grafische Darstellung der beobachteten gegenüber den prognostizierten Werten (vgl. Piñeiro *et al.*, 2008). Auch hieraus wird deutlich, dass die Antwortwerte für die Experimente (2) und (6) einen großen Einfluss auf die multivariaten Modelle haben. Bei MSD5248 wurde über den Grubbs-Test aus Kapitel 7.1.5 Versuch (2) als Ausreißer erkannt, die Nullhypothese jedoch unberechtigt angenommen, da das Ergebnis nicht als Fehlmessung, sondern vielmehr als sachlogisch interessant interpretiert worden ist.

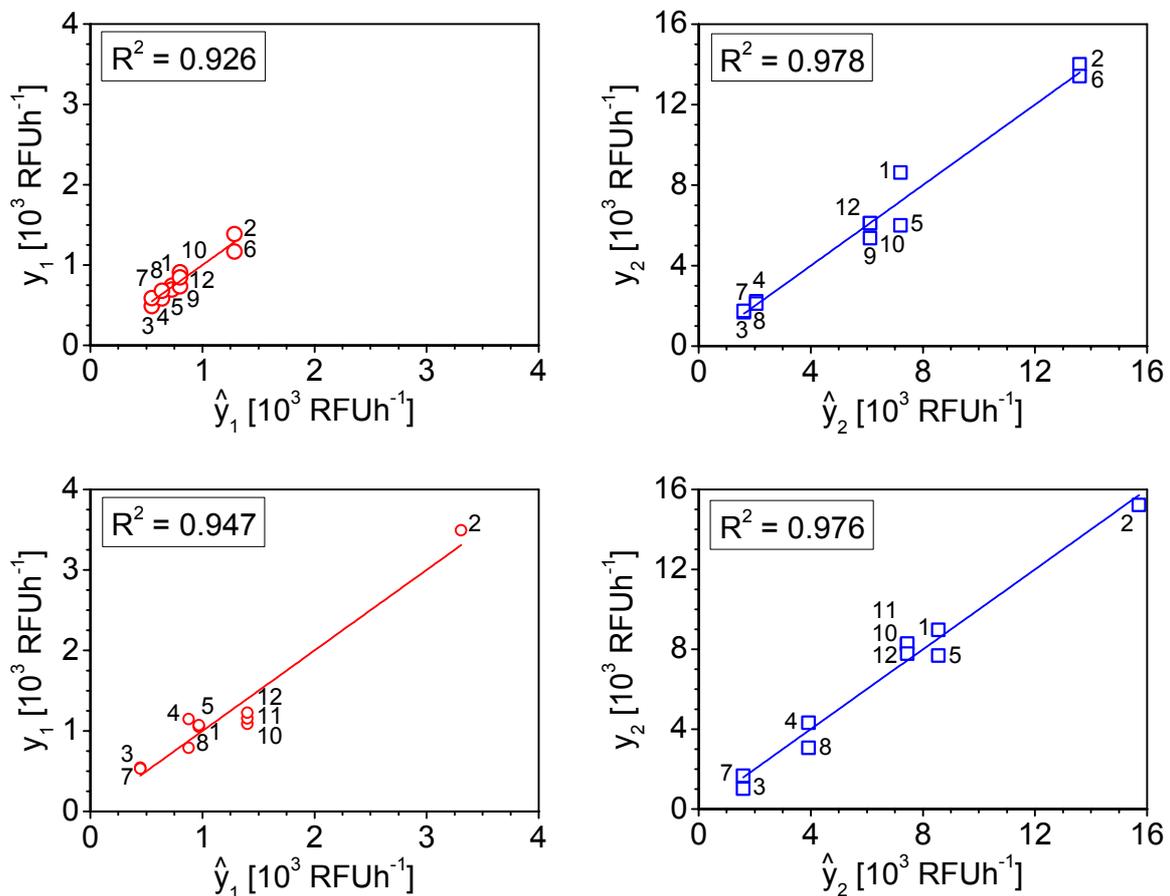


Bild 7.7: Observed versus Predicted Plot für lösliche (\circ , links) und unlösliche (\square , rechts) Proteinexpression für *E. coli* MSD5247 (oben) und MSD5248 (unten)

7.1.11 Auswahl eines *E. coli* Stamms für Optimierungsversuche

Neben der Identifikation von signifikanten Faktoren ist das experimentelle Ziel der Screening-Versuche auch die Bestimmung des interessanteren *E. coli* Stamms für die nachfolgenden Optimierungsexperimente.

Mit den vorhandenen Modellgleichungen (7.20) bis (7.23) ist es möglich, die in Bild 7.8 dargestellten Contourplots zu zeichnen. Für die im Fokus liegende lösliche Raum-Zeit-Ausbeute, aber auch für die RZA des unlöslichen Proteins wird deutlich, dass bei einer hohen Wachstumsrate sowie einer niedrigen Temperatur höhere Werte zu erwarten sind. Subjektiv betrachtet fällt die erste Wahl auf MSD5248, da die lösliche RZA bei hoher Wachstumsrate sowie niedriger Temperatur ca. zweieinhalbmal größer ist als bei MSD5247.

Um diese Entscheidung zu objektivieren und die Fragestellung zu beantworten, ob die beiden *E. coli* Stämme tatsächlich signifikant unterschiedliche Werte für die lösliche RZA liefern, wird ein t-Test benutzt, der die normalverteilten Grundgesamtheiten bezüglich Mittelwert \bar{y} und Varianz MS miteinander vergleicht.

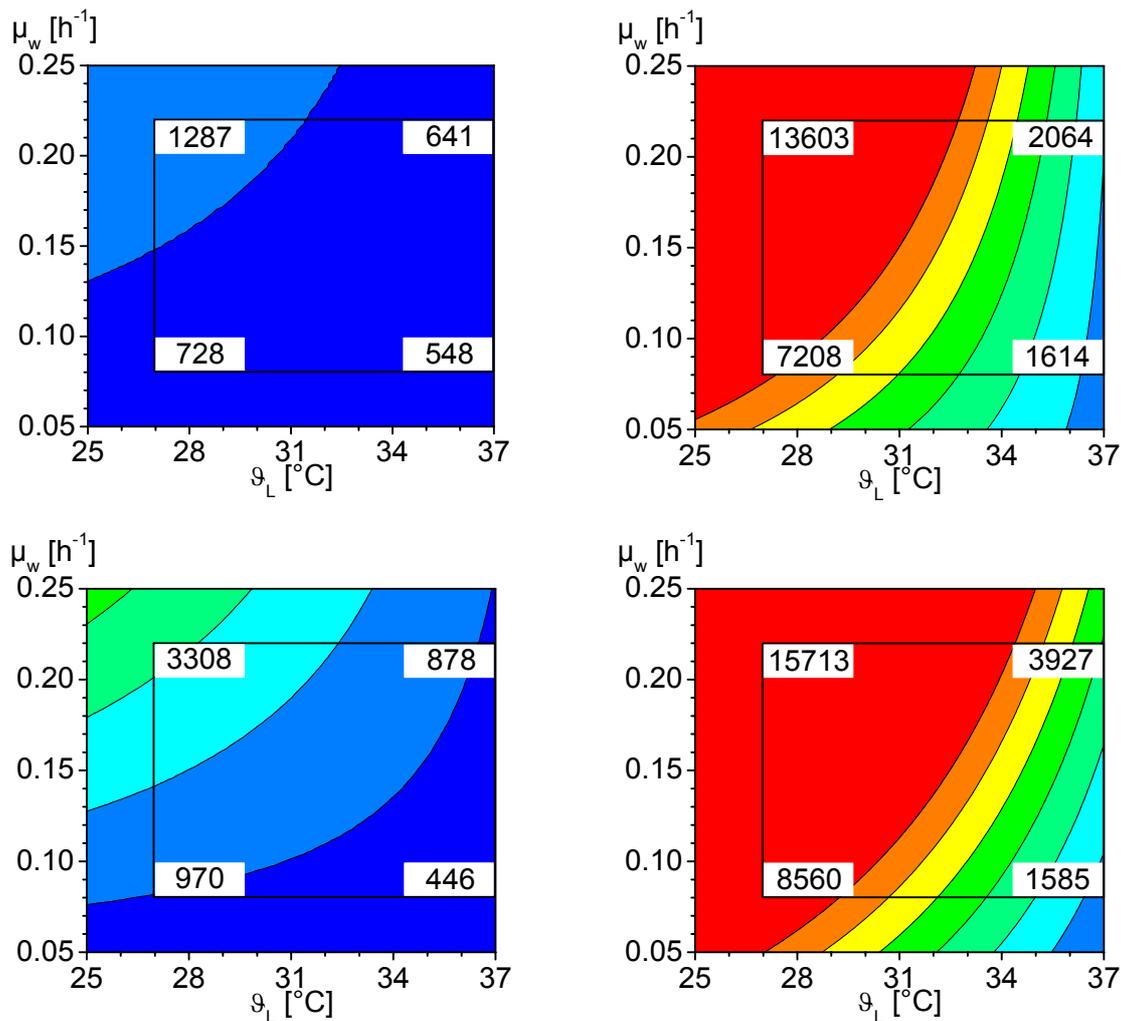


Bild 7.8: Contourplot der Antworten für lösliche (links) und unlösliche (rechts) Proteinexpression für *E. coli* MSD5247 (oben) und MSD5248 (unten)

Vor diesem Mittelwertvergleich ist ein F-Test durchzuführen, der die beiden Gesamtvarianzen MS_{TC} aus Gleichung (7.27) ins Verhältnis setzt.

Im Vergleich zum Lack of Fit Test aus Gleichung (7.35) wird mit dem normalen F-Test die Testgröße F_{TC} ,

$$F_{TC} = \frac{MS_{TC1j}}{MS_{TC1k}} \text{ mit } MS_{TC1j} > MS_{TC1k} \text{ für } j \neq k = 5247 \vee 5248, \quad (7.43)$$

immer mit der größeren Varianz im Zähler berechnet.

Der Vergleich gegen einen kritischen F-Wert $F_{tab}(DF_{TC1j} = 10, DF_{TC1k} = 11, \alpha = 0.05) = 2.85 < F_{TC} = 9.85$ ergibt, dass sich die beiden Varianzen signifikant unterscheiden und deutet somit bereits auf Unterschiede in der biotechnologischen Verfahrensführung beider Screeningversuche hin.

Mit dem t-Test wird nun die einseitige Alternativhypothese H_A geprüft, ob der Mittelwert \bar{y}_{1_5248} signifikant größer als \bar{y}_{1_5247} ist ($H_A: \bar{y}_{1_5248} > \bar{y}_{1_5247}$).

Dafür wird die Testgröße t_{cal} ,

$$t_{cal} = \left| \frac{\bar{y}_{1_5247} - \bar{y}_{1_5248}}{SD_{MD}} \right|, \quad (7.44)$$

als Absolutbetrag aus dem Quotient von Mittelwertdifferenz und Standardabweichung der Mittelwertdifferenz berechnet und gegen einen kritischen t-Wert $t_{tab}(DF_{TC} = 10.7, \alpha)$ verglichen.

Die Berechnung der Freiheitsgrade DF_{TC} sowie der Standardabweichung SD_k hängt dabei von der (Un-) Gleichheit der Gesamtvarianz ($MS_{TC1j} = MS_{TC1j} \vee MS_{TC1j} \neq MS_{TC1j}$) und der Anzahl der betrachteten Versuche ($n_1 = n_2 \vee n_1 \neq n_2$) ab. Auf eine Darstellung der Gleichungen wird an dieser Stelle verzichtet (siehe Scheffler, 1997).

Der t-Test ergibt, dass mit $t_{cal} = 1.473 < t_{tab} = 1.812$ für $DF_{TC} = 10$ und der bevorzugten Irrtumswahrscheinlichkeit $\alpha = 0.05$ die Nullhypothese angenommen wird, bei einer Irrtumswahrscheinlichkeit α von 0.10 mit $t_{cal} = 1.473 > t_{tab} = 1.372$ jedoch die Alternativhypothese.

Da das Risiko für eine irrtümliche Annahme der Alternativhypothese als angemessen betrachtet wird und die Ergebnisse der Regressionsanalyse ebenfalls darauf hindeuten, dass die in die Aminosäuresequenz des Zielproteins eingefügte Punktmutation [L767P] einen positiven Effekt auf die lösliche Proteinexpression hat, wird die subjektive Wahl des Stammes *E. coli* MSD5248 für die Durchführung der Optimierungsexperimente bestätigt.

7.2 Die Durchführung der Optimierungsversuche

7.2.1 Festlegung der Faktorstufen und Auswahl des Optimierungs-Designs

Die Methode des Steilsten Anstiegs (Box and Wilson, 1951) gehört zu den Gradientenverfahren zur Optimumsuche und kann bei der Festlegung von neuen Faktoreinstellungen für die Optimierungsexperimente mit *E. coli* MSD5248 eingesetzt werden. Obwohl ein Verlassen des untersuchten Bereich im Sinne der Regressionsanalyse nicht statthaft ist, erlaubt die Methode des Steilsten Anstiegs senkrecht auf den Isolinien des Contourplots ausgehend vom Zentralpunkt des Screening-Versuchsplans einen Vektor zu errichten, der in die Richtung des vermuteten Maximums zeigt (Scheffler, 1997; Eriksson *et al.*, 2008).

Die Schrittweite für die Temperatur $\Delta\vartheta_{LW}$ wurde mit -2 °C festgelegt, so dass ausgehend vom Zentralpunkt des Screenings bei 32 °C die weiteren Versuchspunkte für die Optimierung mit 30 °C , 28 °C und 26 °C festliegen. Die Schrittweite für die Wachstumsrate $\Delta\mu_w$,

$$\Delta\mu_w = \Delta\vartheta_{LW} \cdot \frac{(\mu_{w\text{high}} - \mu_{w\text{low}}) \cdot \beta_1}{(\vartheta_{L\text{high}} - \vartheta_{L\text{low}}) \cdot \beta_2}, \quad (7.45)$$

kann nun über das Verhältnis der aus dem Screening festgelegten Stufenhöhen $X_{j\text{low}}$ und $X_{j\text{high}}$ für Temperatur ϑ_{LW} und Wachstumsrate μ_w aus Tabelle 7.1 sowie den entsprechenden Regressionskoeffizienten $\beta_1 = 692.7$ und $\beta_2 = -738.7$ für die lösliche Proteinexpression aus Tabelle 7.6 mit $\Delta\mu_w = 0.026\text{ h}^{-1} \approx 0.03\text{ h}^{-1}$ gebildet werden.

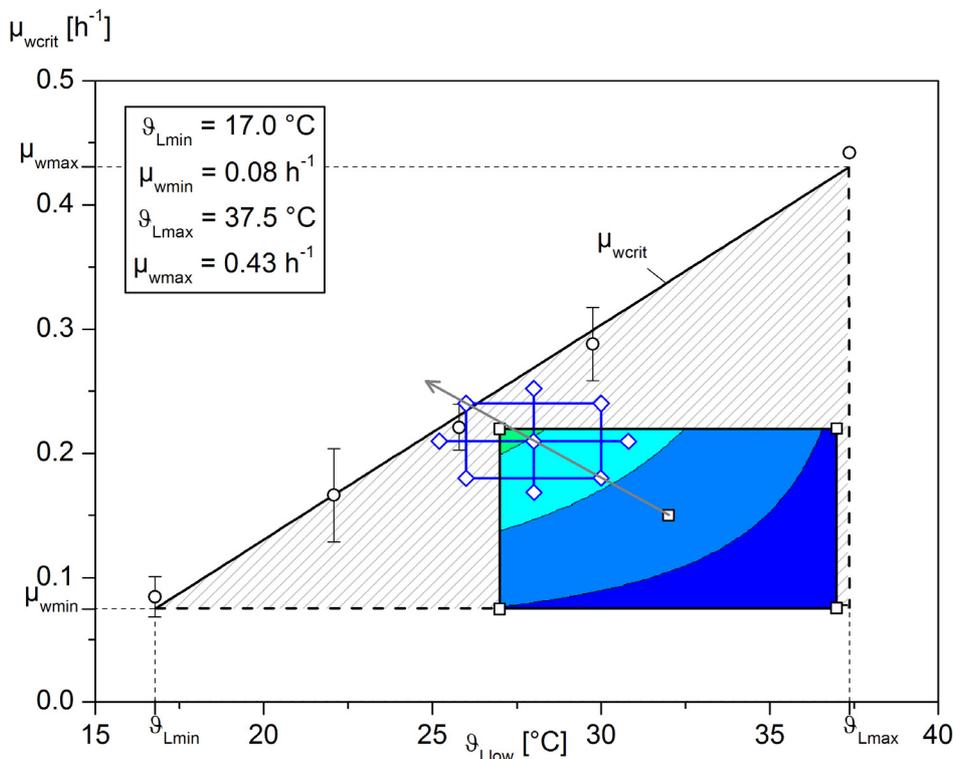


Bild 7.9: Steilster Anstieg mit den neuen Faktoreinstellungen für Temperatur ϑ_{LW} und Wachstumsrate μ_w sowie einem CCC-Optimierungsdesign

Ausgehend vom Zentralpunkt des Screenings bei 0.15 h^{-1} liegen die weiteren Versuchspunkte für die Wachstumsrate mit 0.18 h^{-1} , 0.21 h^{-1} und 0.24 h^{-1} fest. Bei dieser Versuchsanordnung wird bedingt durch die Vorgabe der unteren Kultivierungstemperatur ϑ_{Low} von $26 \text{ }^\circ\text{C}$ der kritische Wert der steuerbaren zellspezifischen Wachstumsrate μ_{crit} von 0.23 h^{-1} nach Gleichung (6.4) überschritten und der erlaubte Versuchsbereich verlassen. Die Überschreitung wird jedoch mit Hinblick auf die zu erwartende Standardabweichung als marginal angesehen und für die Festlegung der neuen Faktoreinstellungen vernachlässigt.

Insgesamt resultieren zwölf Experimente, die in zwei Blöcken durchgeführt werden. Für Screening-Parameter, die die Zielgröße nicht beeinflussen, können in Hinblick auf Optimierungskriterien günstige Werte eingestellt werden. Daher wurde der Sollwert für die molare IPTG-Konzentration für die Durchführung der Optimierungsversuche auf den niedrigsten untersuchten Wert von 0.1 mmol l^{-1} festgesetzt. Mit Hilfe des quadratischen Regressionsmodells aus Gleichung 6.3 können neben den zwei Haupteffekten auch die beiden quadratischen Effekte sowie eine 2-Faktor-Wechselwirkung und der Blockeffekt getrennt erfasst werden.

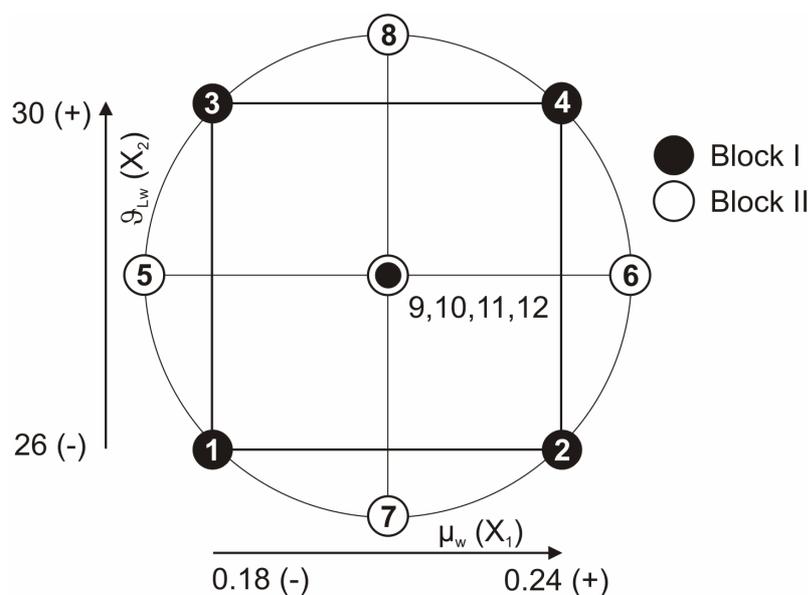


Bild 7.10 zeigt die Geometrie des zentral zusammengesetzten Designs mit den unteren und oberen Faktorstufen $X_{j\text{low}}$ bzw. $X_{j\text{high}}$ sowie vier zusätzlichen Zentralpunkten in der Mitte und vier Sternpunkten, wobei der Sternpunktabstand $\gamma = \sqrt{j} = \sqrt{2} = 1.414$ von der Anzahl j der Faktoren abhängt. Jeder Faktor wird auf fünf Stufen bestimmt.

Bild 7.10: Geometrie des CCC-Optimierungsdesigns

7.2.2 Aufstellung der Planmatrix und Auswertung der Antwortmatrix

Tabelle 7.9 zeigt die randomisierte Planmatrix mit dem Verzeichnis der Einflussgrößen in natürlicher und kodierter Darstellung sowie die Antwortmatrix mit den beobachteten und prognostizierten Zielgrößen für *E. coli* MSD5248. Die untere Stufe der Faktoren ist durch (-), die obere Stufe durch (+), der Zentralpunkt durch (0) und die Sternpunkte durch (-1.414) gekennzeichnet. Die Einzelversuche wurden softwareseitig so eingeteilt, dass im ersten Block die faktorielle Fraktion des Designs und im zweiten Block die Sternpunkte enthalten sind.

Die in Fettdruck gekennzeichneten Werte wurden entweder über den Grubbs-Test (*) mit einer Irrtumswahrscheinlichkeit von 0.10, oder über die Auftragung der studentisierten Residuen (†) im Wahrscheinlichkeitsnetz der Normalverteilung als Ausreißer erkannt und nicht für die weitere Auswertung berücksichtigt. Die Nullhypothese wurde für die RZA des unlöslichen Proteinanteils bei Versuch 8 unberechtigt angenommen, da das Ergebnis nicht als Fehlmessung, sondern vielmehr als sachlogisch interessant interpretiert wurde.

Die Konditionszahl der Designmatrix änderte sich nach Entfernen der Ausreißer von 3.186 auf 3.546. Da die zulässige Änderung der Konditionszahl von ± 0.5 angegeben eingehalten wird, kann weiterhin von einer Orthogonalität des Designs ausgegangen werden.

Tabelle 7.9: Übersicht der randomisierten Planmatrix mit Einflussgrößen in natürlicher und kodierter Darstellung sowie der Antwortmatrix mit den Zielgrößen

Versuch / Abfolge		Faktoren		Antworten MSD5248			
		μ_w [h ⁻¹]	ϑ_{Lw} [°C]	RZA _{sol} [RFUh ⁻¹]		RZA _{IB} [RFUh ⁻¹]	
				observed	predicted	observed	predicted
1	2	0.18 (-)	26 (-)	6279	6515	22433	23714
2	3	0.24(+)	26 (-)	14634	14718	26201	26484
3	6	0.18 (-)	30 (+)	2257	2409	16095	17093
4	4	0.24 (+)	30 (+)	10557	-	34826⁺	-
5	5	0.21 (0)	28 (0)	8247	-	12229⁺⁺	-
6	1	0.21 (0)	28 (0)	7792	6505	29922	29810
7	10	0.167 (-1.414)	28 (0)	4508	4282	27985	26639
8	11	0.252 (+1.414)	28 (0)	13432	13421	50224*	50289
9	9	0.21 (0)	25.2 (-1.414)	10846	10668	14724	13883
10	7	0.21 (0)	30.8 (+1.414)	2461	2402	24694	24254
11	8	0.21 (0)	28 (0)	6595	6505	31702	29810
12	12	0.21 (0)	28 (0)	5128	6505	27806	29810

Die Prüfung auf Normalverteilung der Antwortgrößen nach Shapiro und Wilk wurde ohne Berücksichtigung der vermutlichen Ausreißer durchgeführt. Die Nullhypothese H_0 , dass die Daten aus einer normalverteilten Grundgesamtheit stammen, konnte für alle Antwortgrößen berechtigt angenommen werden. Wie in Tabelle 7.10 gezeigt, ist die berechnete Testgröße W_{cal} stets größer als der kritische Wert W_{tab} .

Tabelle 7.10: Ergebnisse für die Prüfung auf Normalverteilung der Antwortgrößen mit dem W-Test nach Shapiro und Wilk

<i>E. coli</i> Stamm	n	α	$W_{\text{tab}}(n; \alpha)$	$W_{1\text{cal}}$	$W_{2\text{cal}}$
MSD5248	12	0.05	0.859	0.957	0.933
	10		0.842	0.922	0.876

Der Screeningversuch (6) mit *E. coli* MSD5248 erzielte bei den Einstellungen $\mu_w = 0.22 \text{ h}^{-1}$, $\vartheta_{Lw} = 27 \text{ °C}$ sowie $C_{\text{IPTGw}} = 0.1 \text{ mmolL}^{-1}$ Antwortwerte von $RZA_{\text{sol}} = 3492 \text{ RFUh}^{-1}$ bzw. $RZA_{\text{IB}} = 15229 \text{ RFUh}^{-1}$. Im Vergleich zu den Zentralpunktversuchen der Optimierung mit ähnlichen Faktoreinstellungen von $\mu_w = 0.21 \text{ h}^{-1}$, $\vartheta_{Lw} = 28 \text{ °C}$ sowie $C_{\text{IPTGw}} = 0.1 \text{ mmolL}^{-1}$ ist die Raum-Zeit-Ausbeute jedoch ungefähr nur halb so groß.

Die Abweichung kann nicht erklärt werden, da weder eine sorgfältige Analyse der Rohdaten noch die Bewertung mikrobiologischer oder verfahrenstechnischer Kriterien plausible Schlüsse zulassen. Die Daten und die daraus erzielten Ergebnisse werden daher im Folgenden aufgrund der theoretischen Überlegungen als wissenschaftlich korrekt betrachtet.

7.2.3 Ergebnisse der Regressionsanalyse

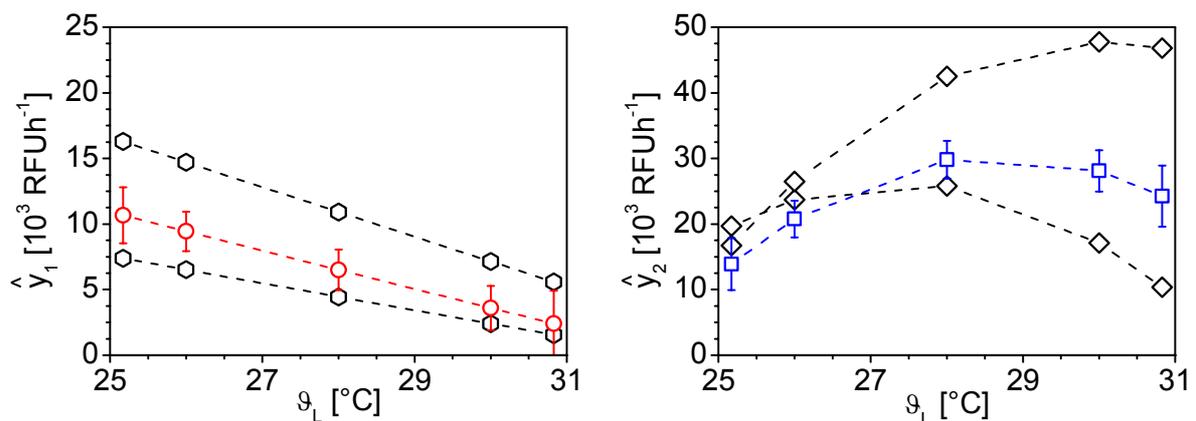
Tabelle 7.11 zeigt die um den Blockeffekt bereinigten Ergebnisse der Regressionsanalyse für die Optimierungsversuche bei einem Konfidenzniveau von 95 %. Der Blockeffekt war mit einem p-Wert von 0.626 bzw. 0.413 für die lösliche bzw. unlösliche Proteinexpression wie bereits beim Screening nicht signifikant und wurde daher aus dem Modell gelöscht. Die Koeffizienten sind skaliert und zentriert dargestellt.

Die Zwei-Faktor-Interaktion (β_{12}) sowie der quadratische Term der Temperatur (β_{22}) sind im Gegensatz zu den Termen der unlöslichen Proteinexpression für die lösliche Proteinexpression mit p-Werten > 0.10 nicht signifikant.

Tabelle 7.11: Ergebnisse der Regressionsanalyse für die Optimierung

Term	lösliche Proteinexpression				unlösliche Proteinexpression			
	Koeff.	Std.Fehl.	p-Wert	Konf.Int.	Koeff.	Std.Fehl.	p-Wert	Konf.Int.
β_0	6505	557.7	<0.001	1549	29810	1042	<0.001	2894
β_1	3232	394.4	0.001	1095	8363	737.2	<0.001	2047
β_2	-2923	394.4	0.002	1095	3667	737.2	0.008	2047
β_{12}	-869.8	623.6	0.236*	1731	4328	782.0	0.005	2171
β_{11}	1174	418.4	0.049	1162	-5373	782.0	0.002	2171
β_{22}	15.13	418.4	0.973*	1162	6978	1166	0.004	3236

Bild 7.11 zeigt die Änderung der prognostizierten Antwortgrößen hervorgerufen durch den Haupteffekt der Temperatur sowie den Einfluss der Zwei-Faktor-Interaktion.

**Bild 7.11:** Grafische Darstellung von Haupt- und Interaktionseffekt für die lösliche (links) und unlösliche (rechts) Proteinexpression für *E. coli* MSD5248

Da für die lösliche Proteinausbeute kein quadratischer Effekt für die Temperatur bzw. kein Wechselwirkungseffekt vorliegt, verlaufen die Kurven ungekrümmt ohne Schnittpunkt und nur durch Zufallseinflüsse gestört nahezu parallel. Es ist klar zu erkennen, dass eine hohe zellspezifische Wachstumsrate über den untersuchten Temperaturbereich einen größeren Einfluss auf die lösliche Proteinausbeute hat als eine niedrige Wachstumsrate. Sofern bei mittlerer Wachstumsrate die Temperatur von 25.2 °C auf 30.8 °C erhöht wird, sinkt die lösliche Proteinausbeute bedingt durch den negativen Haupteffekt um 77.5 %.

Für die unlösliche Proteinexpression ist die Zwei-Faktor-Interaktion mit einem p-Wert ≤ 0.01 sehr signifikant, was auch durch den Schnittpunkt der beiden Kurven im unteren Temperatur-

bereich bestätigt wird. Die Zunahme der unlöslichen Proteinausbeute durch Temperaturerhöhung wird durch eine hohe Wachstumsrate wesentlich größer beeinflusst als durch eine niedrige Wachstumsrate. Der sehr signifikante quadratische Temperatureffekt ist für den gekrümmten Kurvenverlauf von Haupt- und Interaktionseffekt verantwortlich. Dass die unlösliche Proteinausbeute bei höherer Temperatur wieder abnimmt, deutet auf ein Optimum hin.

Tabelle 7.12: Wichtige Parameter für verschiedene Modelle der löslichen Proteinexpression sowie für das Modell der unlöslichen Proteinexpression

Modell	Berücksichtigte Modellterme						R^2	R^2_{adj}	Q^2	MV	RP
\hat{y}_1 (I)	β_0	β_1	β_2	β_{12}	β_{11}	β_{22}	0.978	0.950	0.943	0.988	0.904
\hat{y}_1 (II)	β_0	β_1	β_2	β_{12}	β_{11}		0.978	0.960	0.961	0.997	0.904
\hat{y}_1 (III)	β_0	β_1	β_2		β_{11}		0.966	0.950	0.917	0.964	0.904
\hat{y}_2 (I)	β_0	β_1	β_2	β_{12}	β_{11}	β_{22}	0.985	0.966	0.918	0.865	0.961

Der nicht-signifikante quadratische Term der Temperatur (β_{22}) wurde aus dem Modell für die lösliche Proteinexpression gestrichen. Wie in Tabelle 7.12 gezeigt, verbesserte sich dadurch nicht nur das korrigierte Bestimmtheitsmaß R^2_{adj} um 0.010, sondern auch die Prognosegüte Q^2 um 0.018 und die Modellvalidität um 0.009.

Das Löschen der nicht-signifikanten Zwei-Faktor-Wechselwirkung (β_{12}) aus dem Modell führt einerseits zu schlechteren Modellparametern, andererseits aber auch zu einer besseren Normalverteilung der studentisierten Residuen $e_{1\text{stu}}$, wie in Bild 7.12 dargestellt.

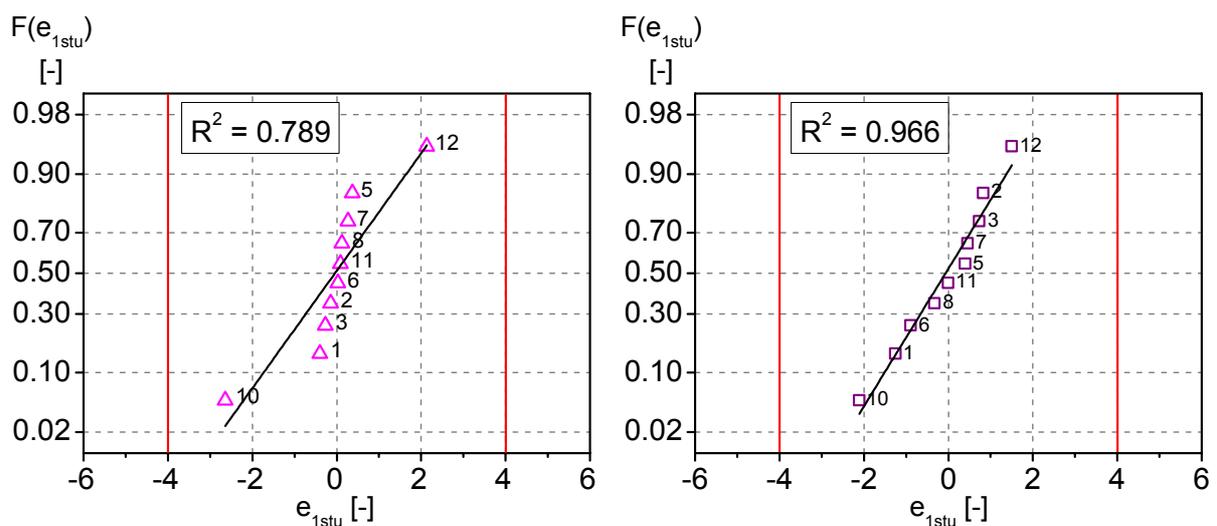


Bild 7.12: Darstellung der studentisierten Residuen für den löslichen Proteinanteil $e_{1\text{stu}}$ bei *E. coli* MSD5248 im Wahrscheinlichkeitsnetz der Normalverteilung

Da die Normalverteilung der Residuen eine Grundvoraussetzung für die statistische Bewertung der erhaltenen Ergebnisse aus der linearen Regressionsanalyse ist (Hartung, 2005; Sachs and Hedderich, 2009), wird die Reduktion der Parameter, insbesondere die von Q^2 , um -0.044, akzeptiert.

Das Modell für die lösliche Raum-Zeit-Ausbeute für *E. coli* MSD5248,

$$\hat{y}_1 = 6598 + 3399 \cdot x_1 - 2756 \cdot x_2 + 1228 \cdot x_1^2, \quad (7.46)$$

reduziert sich damit auf den konstanten Term, die linearen Abhängigkeiten sowie den quadratischen Term der Wachstumsrate.

Das Modell für die unlösliche Raum-Zeit-Ausbeute,

$$\hat{y}_2 = 29810 + 8363 \cdot x_1 + 3667 \cdot x_2 + 4328 \cdot x_1 \cdot x_2 - 5373 \cdot x_1^2 + 6978 \cdot x_2^2, \quad (7.47)$$

enthält dagegen zusätzlich die nicht-lineare Interaktion und den quadratischen Term der Temperatur.

7.2.4 ANOVA für die Optimierungsversuche

In Tabelle 7.13 und Tabelle 7.14 sind die Ergebnisse der Varianzanalyse für die lösliche und unlösliche Proteinexpression aus den Optimierungsversuchen zusammengefasst.

Beide Regressionsmodelle zeigen p-Werte, die mit < 0.001 unter der Irrtumswahrscheinlichkeit α von 0.05 liegen, so dass die empirischen F-Werte von 57.59 für die lösliche Proteinexpression und 52.43 für die unlösliche Proteinexpression mit höchster Signifikanz die kritischen F-Werte $F_{\text{tab}}(DF_{\text{REG}}, DF_{\text{RES}}, \alpha)$ von 4.76 bzw. 6.26 überschreiten. Da die erklärte Abweichung der Regressionswerte vom Mittelwert signifikant höher ist als die nicht erklärte Abweichung der Antwortwerte von den Regressionswerten, ist die Erklärungsleistung der Modelle somit kein Zufallsergebnis.

Auch wenn sich der p-Wert nach Löschen des nicht-signifikanten quadratischen Terms der Temperatur und der nicht-signifikanten Zwei-Faktor-Wechselwirkung aus dem Modell für die lösliche Proteinexpression von 0.954 (Modell I) über 0.990 (Modell II) auf 0.867 (Modell III) verschlechtert, weisen die Ergebnisse aus dem Lack-of-Fit Test auch für das Modell der unlöslichen Proteinexpression mit einem p-Wert von 0.583 eindeutig auf einen äußerst geringen Anpassungsdefekt hin. Die ermittelten F-Werte liegen mit 0.287 für die lösliche und 0.714 für die unlösliche Proteinexpression unter den kritischen tabellierten Werten von 19.25 bzw. 19.00, so dass ein signifikanter Teil der Variation der Daten über adäquate Regressionsmodelle erklärt wird.

Tabelle 7.13: ANOVA für die lösliche Proteinexpression bei *E. coli* MSD5248

	DF [-]	SS·10 ⁶ [RFUh ⁻¹]	MS·10 ⁶ [RFUh ⁻¹]	SD·10 ³ [RFUh ⁻¹]	F-Wert [-]	p-Wert [-]
Regression	3	161.4	53.80	7.335	57.59 (4.76)	< 0.001
Residual	6	5.605	0.934	0.967		
Lack of Fit	4	2.045	0.511	0.715	0.287 (19.25)	0.867
Pure Error	2	3.561	1.780	1.334		
Total Corr.	9	167.0	18.56	4.308		

Tabelle 7.14: ANOVA für die unlösliche Proteinexpression bei *E. coli* MSD5248

	DF [-]	SS·10 ⁶ [RFUh ⁻¹]	MS·10 ⁶ [RFUh ⁻¹]	SD·10 ³ [RFUh ⁻¹]	F-Wert [-]	p-Wert [-]
Regression	5	854.7	170.9	13.07	52.43 (6.26)	< 0.001
Residual	4	13.04	3.260	1.806		
Lack of Fit	2	5.432	2.716	1.648	0.714 (19.00)	0.583
Pure Error	2	7.608	3.804	1.950		
Total Corr.	9	867.7	96.42	9.819		

Die guten Ergebnisse der ANOVA spiegeln sich in der grafischen Darstellung der beobachteten gegenüber den prognostizierten Werten wider. Bild 7.13 zeigt die Korrelation der beiden Werte sowohl für die lösliche Proteinexpression mit einer Residuen-Standardabweichung von $0.966 \cdot 10^3$ RFUh⁻¹ als auch für die unlösliche Proteinexpression mit einer Residuen-Standardabweichung von $1.806 \cdot 10^3$ RFUh⁻¹. Experiment (6) wurde für die unlösliche Proteinexpression über den Grubbs-Test als Ausreißer erkannt, jedoch weiter für die Auswertung berücksichtigt.

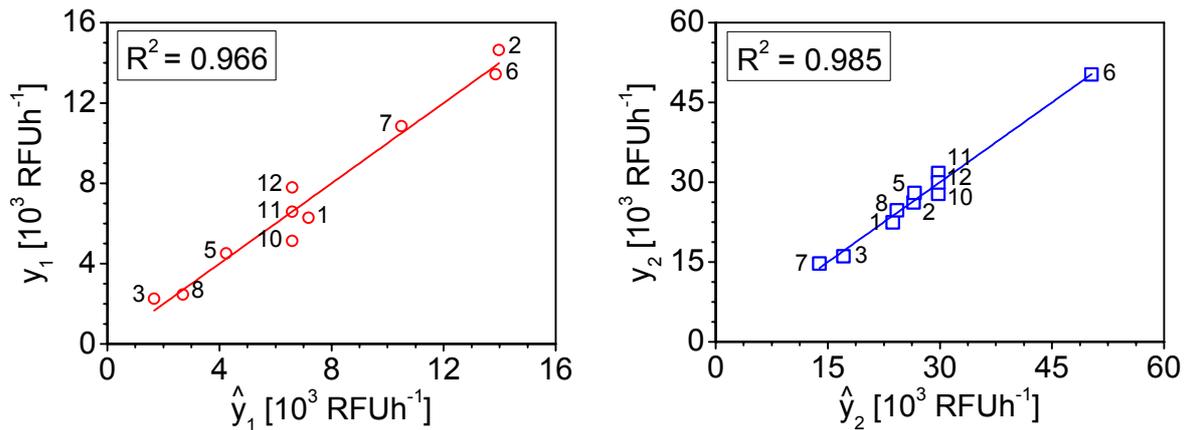


Bild 7.13: Observed versus Predicted Plot für lösliche (○, links) und unlösliche (□, rechts) Proteinexpression für *E. coli* MSD5248

7.2.5 Nichtlineare Optimierung zur Identifikation von optimalen Betriebsbedingungen

Mit den in Gleichung (7.46) und (7.47) dargestellten Regressionsmodellen ist es möglich, einen Punkt, der in optimalen Betriebsbedingungen resultiert, den sog. Set-Point, zu identifizieren (Montgomery, 2009).

Dafür wird zunächst das vollständige quadratische Polynom aus Gleichung (6.3) in Matrixschreibweise überführt,

$$\hat{y} = \hat{\beta}_0 + \underline{x}^T \cdot \underline{b} + \underline{x}^T \cdot \underline{B} \cdot \underline{x}, \quad (7.48)$$

wobei die Spaltenvektoren \underline{x} und \underline{b} bzw. die Matrix \underline{B} ,

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \underline{b} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \quad \text{und} \quad \underline{B} = \begin{bmatrix} \hat{\beta}_{11} & \frac{\hat{\beta}_{12}}{2} \\ \frac{\hat{\beta}_{12}}{2} & \hat{\beta}_{22} \end{bmatrix}, \quad (7.49)$$

unter Berücksichtigung der finalen Zielfunktion (7.46) und (7.47) mit zwei Faktoren notiert werden. Der Spaltenvektor \underline{x} enthält in jeder Zeile die kodierten Stufen der unabhängigen Variablen, der Spaltenvektor \underline{b} die unskalierten Regressionskoeffizienten erster Ordnung und die symmetrische Matrix \underline{B} die quadratischen unskalierten Regressionskoeffizienten in der Hauptdiagonale sowie die halbierten nicht-linearen unskalierten Interaktionsterme.

Die Ableitung der beobachteten Antwort nach \underline{x} ,

$$\frac{\partial \hat{y}}{\partial \underline{x}} = \underline{b} + 2 \cdot \underline{B} \cdot \underline{x} = 0, \quad (7.50)$$

und Null setzen resultiert nach Umstellung der impliziten Form in eine Gleichung,

$$\underline{x}_{SP} = -\frac{1}{2} \cdot \underline{B}^{-1} \cdot \underline{b}, \quad (7.51)$$

mit der die stationären Punkte x_{1SP} und x_{2SP} identifiziert werden können.

Durch Einsetzen von Gleichung (7.51) in Gleichung (7.48) kann die prognostizierte Antwort am stationären Punkt \hat{y}_{SP} ,

$$\hat{y}_{SP} = \hat{\beta}_0 + \frac{1}{2} \cdot \underline{x}_{SP}^T \cdot \underline{b} \quad (7.52)$$

gefunden werden.

Die Lösung der Gleichungen (7.51) und (7.52) resultiert ohne die Definition von Nebenbedingungen in ein undefiniertes Wertepaar $x_{1SPsol} = x_{2SPsol} = \text{undef.}$ für die lösliche Proteinexpression, bzw. mit $x_{1SPIB} = 0.186 \text{ h}^{-1}$ und $x_{2SPIB} = 27.6 \text{ }^\circ\text{C}$ in der Identifikation eines Sattelpunkts für die unlösliche Proteinexpression mit einer RZA_{IB} von 26056 RFUh^{-1} . Dass es sich bei dem Wertepaar tatsächlich um einen Sattelpunkt handelt, kann mit Hilfe einer kanonischen Analyse bewiesen werden (Montgomery, 2009).

Dafür wird zunächst die kanonische Form des Regressionsmodells aus Gleichung (6.3) mit zwei unabhängigen Variablen aufgestellt,

$$\hat{y}_{can} = \hat{y}_{SP} + \lambda_1 \cdot w_1^2 + \lambda_2 \cdot w_2^2, \quad (7.53)$$

wobei w_i die transformierten unabhängigen Variablen und λ_i die Eigenwerte der Matrix \underline{B} darstellen, die über Lösung der charakteristischen Gleichung,

$$\lambda^2 + (\hat{\beta}_{11} + \hat{\beta}_{22}) \cdot \lambda + \left(\hat{\beta}_{11} \cdot \hat{\beta}_{22} + \frac{\hat{\beta}_{12} \cdot \hat{\beta}_{12}}{2} \right) = 0, \quad (7.54)$$

mit $\lambda_{1IB} = 4.812 \cdot 10^6$ und $\lambda_{2IB} = -2.044 \cdot 10^3$ erhalten werden können, womit sich die kanonische Form des Modells aus Gleichung (7.47) für die unlösliche Proteinausbeute,

$$\hat{y}_{2can} = 2.606 \cdot 10^5 + 4.812 \cdot 10^6 \cdot w_1^2 - 2.044 \cdot 10^3 \cdot w_2^2, \quad (7.55)$$

notieren lässt.

Da die Eigenwerte λ_i unterschiedliche Vorzeichen besitzen, handelt es sich bei dem identifizierten Wertepaar $x_{1SPIB} = 0.186 \text{ h}^{-1}$ und $x_{2SPIB} = 27.6 \text{ }^\circ\text{C}$ um einen Sattelpunkt. Sofern die Eigenwerte positiv sind, liegt ein Minimum vor, sind beide negativ, ein Maximum.

Eine direkte Berechnung des Optimums unter Berücksichtigung der aus Gleichung (6.4) resultierenden allgemeinen Nebenbedingung,

$$x_1 - 0.017 \cdot x_2 \geq - 0.210 \quad (7.56)$$

sowie der speziellen Nebenbedingungen für die lösliche Proteinexpression,

$$0.167 \leq x_{1\text{SPsol}} \leq 0.252 \quad (7.57)$$

$$25.2 \leq x_{2\text{SPsol}} \leq 30.8 \quad (7.58)$$

und der speziellen Nebenbedingungen für die unlösliche Proteinexpression,

$$0.186 < x_{1\text{SPIB}} \leq 0.252 \text{ und} \quad (7.59)$$

$$27.6 < x_{2\text{SPIB}} \leq 30.8, \quad (7.60)$$

gestaltet sich aufgrund der Nichtlinearität der endgültigen Modelle sowie dem vorliegenden Ungleichungssystem der Nebenbedingungen als aufwendig.

Zur Lösung nichtlinearer Modelle existiert derzeit kein generelles Verfahren, wenn auch für bestimmte Probleme Algorithmen vorhanden sind. Mögliche Lösungen basieren z. B. auf dem Lagrangeschen Multiplikatorverfahren, dem Kuhn-Tucker-Verfahren und dem Gradienten- oder Hilfsfunktionsverfahren (Ellinger *et al.*, 2003).

Für die Lösung des vorliegenden quadratischen Optimierungsproblems kann die MATLAB[®]-Funktion *quadprog*, aber auch die Funktion *fmincon* verwendet werden. Beide Funktionen erlauben eine Maximierung oder ggfs. auch Minimierung der jeweiligen Zielfunktion mit Nebenbedingungen. Die Umsetzung des Optimierungsproblems erfolgte mit einem MATLAB[®]-Skript.

Beide Funktionen konvergieren sowohl für die lösliche Proteinexpression mit $x_{1\text{SPsol}} = 0.25 \text{ h}^{-1}$ und $x_{2\text{SPsol}} = 27.2 \text{ °C}$ bei $\text{RZA}_{\text{sol}} = 14866 \text{ RFUh}^{-1}$ als auch für die unlösliche Proteinexpression mit $x_{1\text{SPIB}} = 0.25 \text{ h}^{-1}$ und $x_{2\text{SPIB}} = 30.5 \text{ °C}$ bei $\text{RZA}_{\text{IB}} = 58402 \text{ RFUh}^{-1}$ zu einer Lösung.

Bild 7.14 zeigt den Contour Plot und den Response Surface Plot für beide Raum-Zeit-Ausbeuten sowie die Lage der identifizierten optimalen Faktorkombinationen. Der schraffierte Bereich im Contour Plot zeigt die irreguläre experimentelle Region durch den funktionalen Zusammenhang zwischen Kultivierungstemperatur und Wachstumsrate.

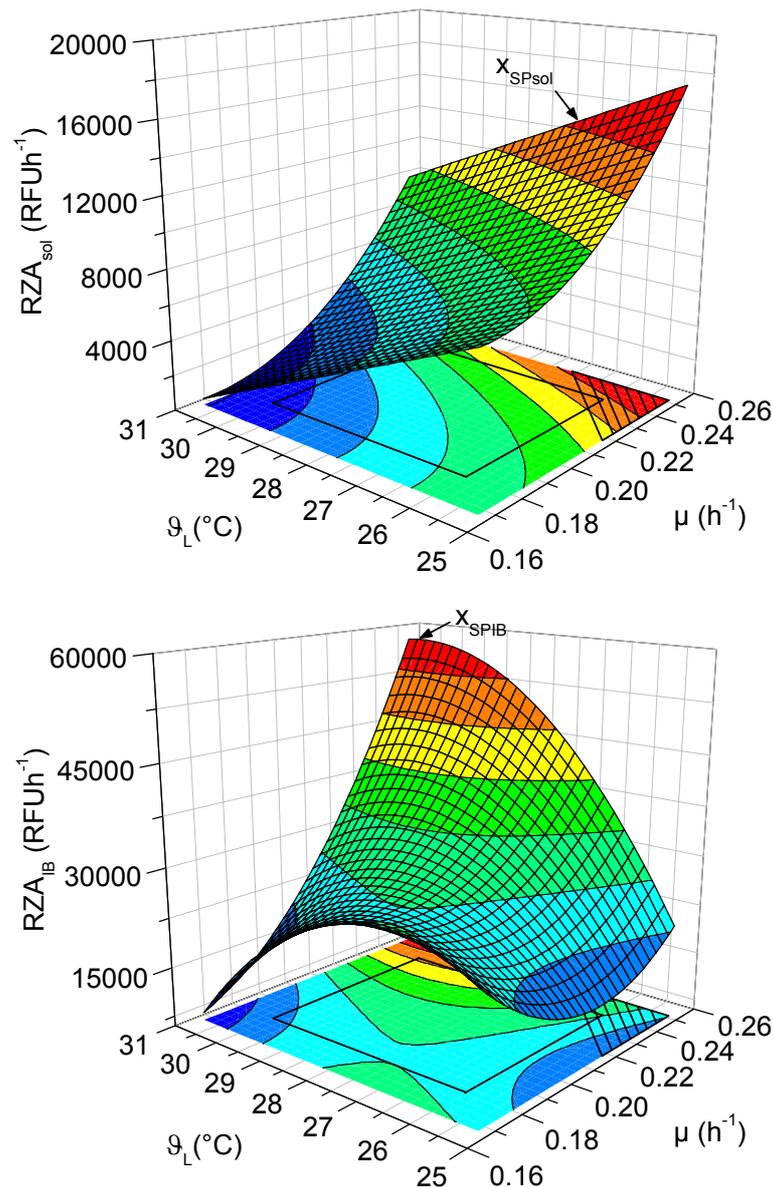


Bild 7.14: Contour Plot und Response Surface Plot für RZA_{sol} (oben) und RZA_{IB} (unten)

7.2.6 Wunschfunktionen und das Simplexverfahren zur Optimumsuche

Nicht immer ist es gewünscht, eine spezifische Zielgröße zu maximieren oder zu minimieren, sondern eine oder mehrere Zielgrößen gleichzeitig hinsichtlich eines spezifizierten Bereichs, in dem unterschiedliche Faktorstufenkombinationen möglich sind, zu optimieren. Daher hat sich im Rahmen der Optimierung technischer Prozesse mithilfe der statistischen Versuchsplanung der Einsatz von Wunschfunktionen⁵ durchgesetzt (Harrington, 1965; Derringer and Suich, 1980, Del Castillo *et al.*, 1996).

⁵ engl.: desirability functions

Der Grundgedanke in der Anwendung von individuellen Wunschfunktionen liegt darin begründet, ein multivariates Optimierungsproblem durch mathematische Transformation auf die univariate Maximierung oder Minimierung einer gemeinsamen Wunschfunktion zu überführen.

In der DoE-Software MODDE wird die gemeinsame Wunschfunktion mit einem Nelder-Mead-Simplex-Algorithmus optimiert, der mit bis zu acht Simplexen simultan startet, um nicht in einem lokalen Extremum gefangen zu sein (Nelder and Mead, 1965). Im Folgenden wird der Optimierungsansatz beschrieben, wie er in der DoE-Software MODDE hinterlegt ist.

Da die Zielgrößen im Allgemeinen hinsichtlich ihrer physikalischen Werte nicht miteinander vergleichbar sind, wird zunächst mit einer individuellen Normierungsfunktion g_k ,

$$g_k = -\ln\left(\frac{100}{100 - v_{\text{kwtg}}}\right) \cdot \frac{\hat{y}_k - y_{\text{kspX}}}{y_{\text{kS}} - y_{\text{kspX}}}, \quad S = \text{low, high}, \quad (7.61)$$

für jede Antwort eine gleiche Größenordnung geschaffen.

Um der Anforderung nach unterschiedlicher Wichtung der einzelnen Zielgrößen gerecht zu werden, wird der Wichtungparameter v_{kwtg} ,

$$v_{\text{kwtg}} = 90 + 80 \cdot \log_{10}(w_k), \quad 0.1 \leq w_k \leq 1, \quad (7.62)$$

eingeführt.

Der individuelle Wichtungsfaktor w_k jeder Zielgröße wird wie die individuellen Spezifikationsgrenzen y_{kS} durch den Benutzer definiert. Sofern für alle Zielgrößen dieselbe Wichtung vorgegeben wird, sucht der Algorithmus nach einer Lösung, die alle Zielgrößen in den Spezifikationsgrenzen liegen lässt.

Die Spezifikationsgrenzen y_{kS} ändern sich mit dem verfolgten Anliegen. Bei der Maximierung einer individuellen Zielgröße wird $y_{\text{kS}} = y_{\text{klow}}$ als unterer noch zu akzeptierender Wert vorgegeben ($y_{\text{klow}} < y_{\text{ktar}}$), bei der Minimierung hingegen $y_{\text{kS}} = y_{\text{khigh}}$ als oberer noch zu akzeptierender Wert ($y_{\text{khigh}} > y_{\text{ktar}}$). Soll die Zielgröße einen bestimmten Wert annehmen, sind beide Spezifikationsgrenzen anzugeben. Der Zielwert y_{ktar} ist dann üblicherweise das Nennmaß des Spezifikationsintervalls.

Die Größe y_{kspX} repräsentiert den niedrigsten von den Startsimplexen berechneten Antwortwert und wird durch MODDE vorgegeben. Die Werte der Zielgröße werden mit Hilfe der unskalierten Regressionsfunktion \hat{y}_k analog Gleichung (7.46) und (7.47) geschätzt.

Soll die Antwort maximiert werden, erhält die Wunschfunktion d_k ,

$$d_k = \begin{cases} 0 & \hat{y}_k \leq y_{k\text{spx}} \\ 100 \cdot (e^{\hat{y}_k} - 1) & y_{k\text{spx}} < \hat{y}_k < y_{k\text{tar}} \\ -100 & \hat{y}_k \geq y_{k\text{tar}} \end{cases} \quad (7.63)$$

den Wert $d_{k\text{min}} = 0$, falls der Schätzwert der Zielgröße kleiner als der niedrigste von den Startsimplex berechnete Antwortwert $y_{k\text{spx}}$ ist.

Sofern der Schätzwert den Zielwert $y_{k\text{tar}}$ der Zielgröße überschreitet, liegt $d_{k\text{max}}$ bei -100. Innerhalb der Spezifikationsgrenzen sinkt der individuelle Wunschwert d_k asymptotisch mit abnehmender Differenz von Schätzwert und Zielwert.

Soll die Antwort minimiert werden, gilt für die Wunschfunktion d_k ,

$$d_k = \begin{cases} -100 & \hat{y}_k \leq y_{k\text{tar}} \\ 100 \cdot (e^{\hat{y}_k} - 1) & y_{k\text{tar}} < \hat{y}_k < y_{k\text{spx}} \\ 0 & \hat{y}_k \geq y_{k\text{spx}} \end{cases} \quad (7.64)$$

Obiges entsprechend umgekehrt.

Sofern der Schätzwert der Zielgröße \hat{y}_k der unteren bzw. oberen Spezifikationsgrenze $y_{k\text{low}}$ bzw. $y_{k\text{high}}$ entspricht, resultiert aus der Definition der Normierungsfunktion (7.61) für die Wunschfunktion d_k der Wert $-v_{k\text{wgt}}$.

Sind zweiseitige Spezifikationsgrenzen vorgegeben, nimmt die Wunschfunktion d_k ,

$$d_k = \begin{cases} 0 & \hat{y}_k \geq y_{k\text{high}} \\ 100 \cdot (e^{\hat{y}_k} - 1) & y_{k\text{low}} < \hat{y}_k < y_{k\text{high}} \\ 0 & \hat{y}_k \leq y_{k\text{low}} \end{cases} \quad (7.65)$$

den Wert $d_{k\text{min}} = 0$ an, sofern der Schätzwert der Zielgröße außerhalb der Grenzen liegt. Im Intervall $y_{k\text{tar}} \leq \hat{y}_k < y_{k\text{high}}$ ist $y_{k\text{S}} = y_{k\text{high}}$, im Intervall $y_{k\text{low}} < \hat{y}_k \leq y_{k\text{tar}}$ ist $y_{k\text{S}} = y_{k\text{low}}$.

Da sich jeder Schätzwert aus der entsprechenden Regressionsfunktion bei vorgegebenen Einflussgrößen berechnet, gelten die gleichen Einstellungen der Einflussgrößen auch für die individuellen Wunschwerte. Damit erlauben die individuellen Wunschwerte eine Zusammenfassung der Zielgrößen zu einem gemeinsamen Wunschwert bzw. einer Gesamterwünschtheit D ,

$$D = \frac{\sum_{j=1}^k v_{j\text{wgt}} \cdot d_j}{\sum_{j=1}^k v_{j\text{wgt}}}, \quad (7.66)$$

die sich als gewichtetes Mittel der individuellen Wunschfunktionen ergibt. Der Nelder-Mead-Simplex-Algorithmus benutzt diese Funktion, um eine Faktorkombination innerhalb der Antwortspezifikationen möglichst nah an den Zielwerten y_{ktar} vorausszusagen.

Die Entfernung zum Gesamtziel (engl.: overall distance to target) wird über den Indikator $\text{Log}(D)$,

$$\text{Log}(D) = \log_{10} \left[\frac{\sum_{j=1}^k v_{j\text{wgt}} \cdot \left(\frac{\hat{y}_j - y_{jtar}}{y_{jtar} - y_{js}} \right)^2}{k} \right] \quad (7.67)$$

berechnet, der zwar nicht als Gütewert in der Optimierung verwendet, aber dennoch dem Benutzer als finales und einziges Bewertungskriterium in der Software zur Verfügung gestellt wird.

Positive $\text{Log}(D)$ -Werte deuten darauf hin, dass das Gesamtziel nicht erreicht und somit keine der vorgegeben Antwortspezifikationen erfüllt wurde. Wird nach einer Lösung gesucht, die viele Kriterien erfüllen soll, wird das Ergebnis ein Kompromiss sein, und je besser ein Kompromiss ist, desto kleiner der $\text{Log}(D)$ -Wert. Bei $\text{Log}(D) = -10$ werden alle Antwortspezifikationen erfüllt.

Mit dem Ziel die Gesamtausbeute aktiver Proteine in der löslichen Fraktion zu erhöhen, wurde für das vorliegende Optimierungsproblem eine Maximierung der löslichen Raum-Zeit-Ausbeute durchgeführt. Die untere Spezifikationsgrenze y_{low} wurde mit 10500 RFU h^{-1} festgelegt, wobei als Orientierung das 75 %-Quantil der beobachteten Antwortgröße mit 10846 RFU h^{-1} galt. Der Zielwert y_{tar} wurde mit 14500 RFU h^{-1} definiert. Die Intervallgrenzen für die Suche des Maximums wurden für Wachstumsrate und Temperatur mit $0.17 \text{ h}^{-1} \leq x_1 \leq 0.25 \text{ h}^{-1}$ bzw. $27.0 \text{ }^\circ\text{C} \leq x_2 \leq 30.8 \text{ }^\circ\text{C}$ festgelegt.

Bild 7.15 zeigt den Verlauf des $\text{Log}(D)$ -Wertes sowie den Verlauf der prognostizierten Antwort mit zunehmender Anzahl der Iterationen als Ergebnis der Optimierung auf Basis der gemeinsamen Wunschfunktion und des Nelder-Mead-Simplex-Algorithmus. Nach maximal 118 Iterationen war die Optimierung beendet und resultierte für die lösliche Proteinexpression mit $x_{1SPsol} = 0.25 \text{ h}^{-1}$ und $x_{2SPsol} = 27.1 \text{ }^\circ\text{C}$ in einer optimalen Faktorkombination bei einer maximalen Raum-Zeit-Ausbeute von $RZA_{sol} = 14611 \text{ RFU}h^{-1}$.

Sowohl die Lösung des Optimierungsproblems mit den MATLAB[®]-Funktionen *quadprog* bzw. *fmincon* als auch die Lösung über die Wunschfunktion und das Simplexverfahren resultieren in einem identischen Set-Point für die Wachstumsrate von $x_{1SPsol} = 0.25 \text{ h}^{-1}$. Der Set-Point für die Temperatur weicht nur minimal um $0.1 \text{ }^\circ\text{C}$ ab.

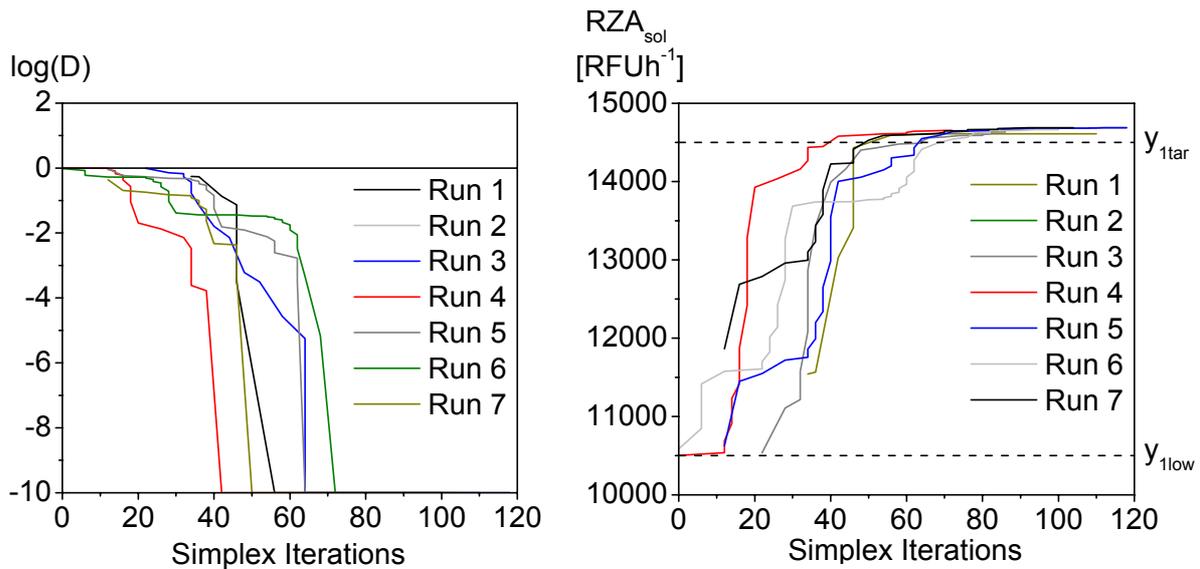


Bild 7.15: Verlauf des Log(D)-Wertes (links) und der prognostizierten Antwort (rechts) mit zunehmender Anzahl der Iterationen

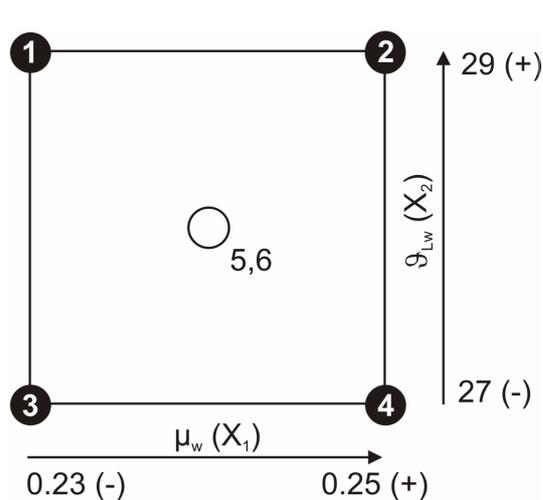
7.3 Verifikation der optimalen Betriebsparameter durch den Robustheitstest

7.3.1 Festlegung der Faktorabweichungen und Erstellung des Designs

Idealerweise sollte ein optimierter Prozess resistent gegenüber äußeren Einflüssen sein und leichte Schwankungen in den Prozessbedingungen tolerieren. Das primäre Ziel für einen Robustheitstest ist daher die gezielte Überprüfung der definierten Antworten auf Robustheit oder (bevorzugt fehlende) Sensitivität auf kleine Faktoränderungen.

Das ideale Ergebnis einer Robustheitsstudie sind nahezu identische Antwortwerte für jeden Versuch (Eriksson *et al.*, 2008). Dies erfordert jedoch eine exakte Einstellung der Faktorstufen, so dass der Antwortwert nur infolge der Versuchsstreuung schwankt. Die zielgerichtete Beeinflussung der Variabilität ist nur für quantitative Faktoren möglich, für deren Einfluss auf die Antwortgröße ein entsprechendes Regressionspolynom vorliegt. Der Wechsel der Faktoreinstellungen soll bei einem Robustheitstest jedoch idealerweise nicht per Regressionsanalyse erfassbar sein, sondern wird vielmehr von der Annahme ausgegangen, dass die Faktoren Zufallsgrößen sind und die Variabilität nur durch Einstellschwankungen verursacht wird.

Wie in Bild 7.16 dargestellt, wurde ein einfaches vollfaktorielles Design mit vier Versuchspunkten sowie zwei Zentralpunkten für den Robustheitstest verwendet, so dass mit einem Kultivierungslauf im BIOSTAT[®] Qplus die Versuche für den Robustheitstest abgeschlossen sind. Das Design unterstützt ein Interaktionsmodell 1. Ordnung nach Gleichung (6.2), so dass die im Hauptaugenmerk liegenden linearen Effekte und deren Auswirkung auf die Zielgrößen evaluiert werden können.



Das Design wurde in der Art um den Set-Point ($x_{1SPsol} = 0.25 \text{ h}^{-1}$, $x_{2SPsol} = 27.1 \text{ °C}$) aus der Optimierung zentriert, dass die vorhandene Grenze durch den Zusammenhang von Wachstumsrate und Temperatur sowie die des experimentellen Bereiches nicht verletzt wurde. Die Schrittweite für die Festlegung der Faktorabweichungen wurde mit 5 % vom Set-Point angenommen, so dass mit $\Delta\vartheta_{Lw} = 1.36 \text{ °C} \approx 1 \text{ °C}$ und $\Delta\mu_w = 0.0125 \text{ h}^{-1} \approx 0.01 \text{ h}^{-1}$ die weiteren Versuchspunkte festliegen.

Bild 7.16: Geometrie des Designs für den Robustheitstest

Tabelle 7.15 zeigt die Planmatrix mit Einflussgrößen in natürlicher und kodierter Darstellung sowie der Antwortmatrix mit den beobachteten und erwarteten Zielgrößen sowie die prozentuale Abweichung der Experimental- von den Modellwerten. Experiment 5 wurde über die Auftragung der studentisierten Residuen ($^+$) im Wahrscheinlichkeitsnetz der Normalverteilung als Ausreißer erkannt und nicht für die weitere Auswertung berücksichtigt. Die Konditionszahl änderte sich durch die Entfernung des fehlgeschlagenen Experiments von 1.225 auf 1.118 innerhalb der zulässigen Grenze von ± 0.5 .

Tabelle 7.15: Übersicht der randomisierten Planmatrix sowie der Antwortmatrix mit den beobachteten und prognostizierten Zielgrößen für den Robustheitstest

Versuch / Abfolge		Faktoren		Antworten MSD5248					
		μ_w [h ⁻¹]	ϑ_{Lw} [°C]	RZA _{sol} [RFUh ⁻¹]			RZA _{IB} [RFUh ⁻¹]		
				observed	predicted	differ.	observed	predicted	differ.
1	2	0.23 (-)	27 (-)	12407	10787	+15.0 %	38598	31808	+21.3 %
2	5	0.25(+)	27 (-)	10984	14691	-25.2 %	32684	40828	-19.9 %
3	4	0.23 (-)	29 (+)	6715	8032	-16.4 %	39703	40126	-1.1 %
4	6	0.25 (+)	29 (+)	9208	11935	-22.8 %	38644	53798	-28.2 %
5	1	0.24(0)	28 (0)	3502⁺	11225	-68.8 %	29101	42502	-31.5 %
6	3	0.24(0)	28 (0)	11398	11225	+1.5 %	38563	42502	-9.3 %

7.3.2 Regressionsanalyse und ANOVA

Im Gegensatz zur ausführlichen Regressionsanalyse für Screening- oder Optimierungsversuche liegt beim Robustheitstest der Fokus auf der Betrachtung der Gültigkeit bzw. Signifikanz der Regressionsmodelle über die Güteparameter R^2 und Q^2 . Generell gilt eine niedrige Prognosegüte Q^2 als ein guter Indikator für Robustheit, da eine niedrige Gesamtstreuung kleine Antwortvariation bedingt und so in einer Unempfindlichkeit der Antworten auf kleine Faktoränderungen resultiert (Mandenius and Brundin, 2008; Eriksson *et al.*, 2008).

Bild 7.17 zeigt einen Vergleich der Regressionskoeffizienten für das Modell der löslichen und unlöslichen Proteinexpression bei *E. coli* MSD5248. Die Koeffizienten sind skaliert und zentriert mit einem Konfidenzniveau von 95 % aufgetragen. Da die Konfidenzintervalle den Wert Null einschließen ist der Einfluss der Regressionskoeffizienten auf die Zielgrößen nur zufällig und damit vernachlässigbar.

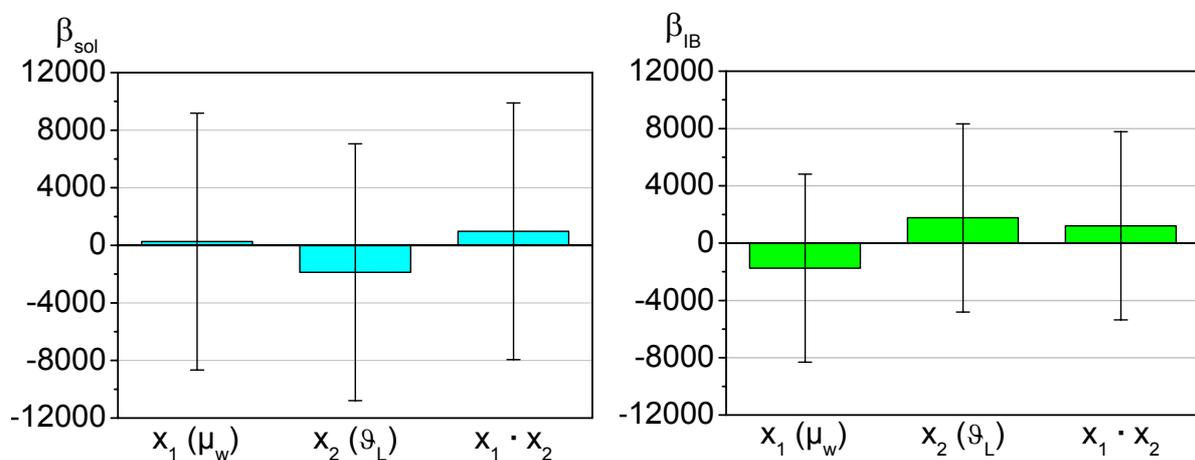


Bild 7.17: Vergleich der Regressionskoeffizienten für die lösliche (links) und unlösliche (rechts) Proteinexpression

Tabelle 7.16 zeigt die Ergebnisse der Regressionsanalyse für die durchgeführten Versuche.

Tabelle 7.16: Ergebnisse der Regressionsanalyse für den Robustheitstest

Term	lösliche Proteinexpression				unlösliche Proteinexpression			
	Koeff.	Std.Fehl.	p-Wert	Konf.Int.	Koeff.	Std.Fehl.	p-Wert	Konf.Int.
β_0	10142	627.8	0.039	7977	37638.4	462.3	0.008	5874
β_1	267.5	701.9	0.768*	8919	-1743.25	516.867	0.183*	6567
β_2	-1867	701.9	0.229*	8919	1766.25	516.867	0.181*	6567
β_{12}	979	701.9	0.396*	8919	1213.75	516.867	0.256*	6567

Sowohl die linearen Terme für die zellspezifische Wachstumsrate μ_w (β_1) und die Flüssigphasentemperatur (β_2) als auch ihre Zwei-Faktor-Interaktion (β_{12}) sind für die lösliche bzw. unlösliche Proteinexpression mit p-Werten > 0.10 nicht signifikant. Beide Antworten sind somit unempfindlich gegen Änderungen der untersuchten Faktoren und bestätigen Robustheit.

Trotz des hohen korrigierten Bestimmtheitsmaßes für die Modelle der löslichen und unlöslichen Proteinexpression ist in beiden Fällen die Prognosegüte Q^2 negativ, das wie bereits beschrieben auf die Robustheit der Antworten auf kleine Faktoränderungen hindeutet. Die Gütemaße für die nicht im Fokus liegende Modellvalidität und Reproduzierbarkeit können nicht in Tabelle 7.17 dargestellt werden, da die Ergebnisse für einen Zentralpunktversuch als Ausreißer gewertet wurden.

Tabelle 7.17: Wichtige Modellparameter als Gütemaße der Regression

Modell	R^2	R_{adj}^2	Q^2
y_1 (löslich)	0.902	0.607	-6.992
y_2 (unlöslich)	0.966	0.865	-1.748

Die Resultate der Varianzanalyse sind in Tabelle 7.18 und Tabelle 7.19 wiedergegeben. Die beiden F-Werte deuten darauf hin, dass sich Regressionsvarianz und Restvarianz nicht signifikant unterscheiden und somit die Erklärungsleistung der Modelle als Zufallsergebnis zu akzeptieren ist. Beide Regressionsmodelle zeigen p-Werte die mit > 0.10 über der Irrtumswahrscheinlichkeit α von 0.05 liegen. Daher kann angenommen werden, dass beide Varianzen aus der gleichen Verteilung stammen und somit die gezielt herbeigeführte Veränderung der Antwortgrößen innerhalb der Versuchsstreuung liegt.

Tabelle 7.18: ANOVA für die lösliche Proteinexpression bei *E. coli* MSD5248

	DF	SS·10 ⁶	MS·10 ⁶	SD·10 ³	F-Wert	p-Wert
	[-]	[RFU·h ⁻¹]	[RFU·h ⁻¹]	[RFU·h ⁻¹]	[-]	[-]
Regression	3	18.06	6.021	2.454	3.055 (215.7)	0.393*
Residual	1	1.971	1.971	1.404		
Total Corr.	4	20.03	5.008	2.238		

Tabelle 7.19: ANOVA für die unlösliche Proteinexpression bei *E. coli* MSD5248

	DF [-]	SS·10 ⁶ [RFUh ⁻¹]	MS·10 ⁶ [RFUh ⁻¹]	SD·10 ³ [RFUh ⁻¹]	F-Wert [-]	p-Wert [-]
Regression	3	30.53	10.18	3.190	9.523 (215.7)	0.233*
Residual	1	1.069	1.069	1.034		
Total Corr.	4	31.60	7.899	2.811		

7.3.3 Beurteilung der Variabilität der Antworten

Die Ergebnisse der Regressionsanalyse und ANOVA haben gezeigt, dass sowohl für die lösliche als auch für die unlösliche Raum-Zeit-Ausbeute nicht-signifikante Regressionsmodelle vorliegen. Daher kann angenommen werden, dass die Variabilität in den Experimenten als Streuung um den Mittelwert der jeweiligen RZA vorliegt. Auf dieser Annahme basiert eine weitere Definition für Robustheit, die dann gewährleistet ist, wenn einerseits der Mittelwert der beobachteten Antwortgrößen einen gewünschten Zielwert erreicht und andererseits die Variabilität um den Mittelwert so gering wie möglich ist (Montgomery, 2009).

Mit Hilfe der Regressionsmodelle für die Raum-Zeit-Ausbeuten aus den Gleichungen (7.46) und (7.47) konnten die in Tabelle 7.15 prognostizierten Antwortwerte für den Robustheitstest berechnet werden. Als zu erreichender Zielwert wird der Mittelwert der erwarteten Antwortwerte für die lösliche und die unlösliche RZA definiert. Wie in Tabelle 7.20 zusammengefasst, weicht der beobachtete Mittelwert für die lösliche RZA um -10.5 % und für die unlösliche RZA um -10.0 % vom prognostizierten und gewünschten Zielwert ab.

Tabelle 7.20: Vergleich der beobachteten und prognostizierten Antwortmittelwerte

Antwort	predicted	observed	Abweichung
\bar{y}_1 [RFUh ⁻¹]	11334	10142	-10.5 %
\bar{y}_2 [RFUh ⁻¹]	41812	37638	-10.0 %

Die Variabilität um den Mittelwert kann durch ein zweiseitiges Konfidenzintervall ausgedrückt werden (Eriksson *et al.*, 2008). Ähnlich wie bei der Berechnung des zweiseitigen Konfidenzintervalls der Regressionskoeffizienten nach Gleichung (7.17) werden die Konfidenzgrenzen für den wahren Mittelwert μ ,

$$\bar{y}_k - t_{\alpha/2, n-1} \cdot SE_{MEA} \leq \mu_k \leq \bar{y}_k + t_{\alpha/2, n-1} \cdot SE_{MEA} \quad (7.68)$$

ebenfalls mit Hilfe der t-Verteilung ($\alpha = 0.05$) festgelegt.

Der Standardfehler des Mittelwerts (engl.: standard error of the mean) SE_{MEA} ,

$$SE_{MEA} = \frac{SD_{TC}}{\sqrt{n}}, \quad (7.69)$$

wird aus dem Verhältnis von Standardabweichung SD_{TC} und Wurzel aus $n = 5$ durchgeführten Versuchen berechnet.

Robustheit kann demnach angenommen werden, wenn sich die Streuung der beobachteten Antwortgrößen im 95 % Konfidenzintervall $10142 \pm 2778 \text{ RFUh}^{-1}$ für die lösliche RZA bzw. $47638 \pm 3490 \text{ RFUh}^{-1}$ für die unlösliche RZA bewegt. Bild 7.18 zeigt den Replicate Plot für eine weitere Beurteilung der Robustheit.

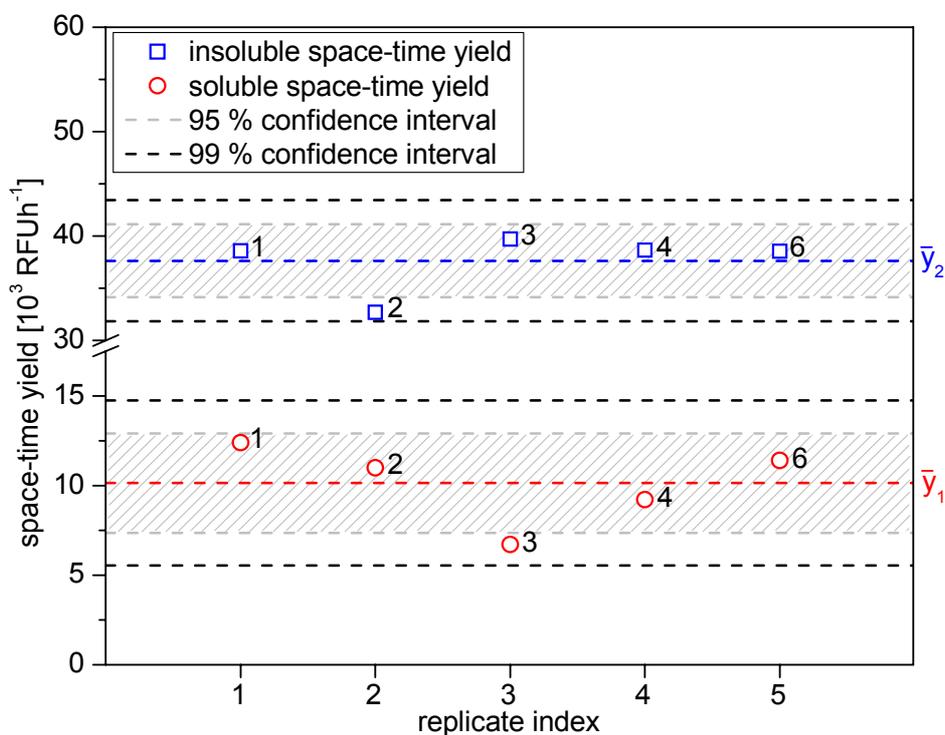


Bild 7.18: Replicate Plot für die Beurteilung der Robustheit mit Lage der Antworten im vorgegebenen Toleranzlimit des Konfidenzintervalls

Bereits anhand der Rohwerte aus Tabelle 7.15 ist zu erkennen, dass die beobachtete RZA_{sol} für Experiment 3 mit 6715 RFUh^{-1} und die beobachtete RZA_{IB} für Experiment 2 mit 32684 RFUh^{-1} nicht die nach Gleichung (7.68) berechneten unteren 95 % Konfidenzgrenzen überschreitet. Die restlichen zur Auswertung herangezogenen Antwortwerte erfüllen dagegen die Vorgabe. Bei einer kleiner gewählten Irrtumswahrscheinlichkeit $\alpha = 0.01$ werden die Konfidenzintervalle größer und die Robustheitsannahme bestätigt, damit aber auch die Angaben der Mittelwerte ungenauer.

8 Zusammenfassung

Die vorliegende Arbeit dokumentiert einerseits die Umsetzung einer Strategie zur vollautomatischen Herstellung rekombinanter Proteine in Integrierter Prozessführung mit der methylo-trophen Hefe *Pichia pastoris*, andererseits die Optimierung eines Herstellungsprozesses über Statistische Versuchsplanung mit zwei Mutanten des Bakteriums *Escherichia coli*.

Die Kultivierungen des sekretorischen Expressionssystems *P. pastoris* wurden mit dem Bioreaktor BIOSTAT[®] C-Xflow durchgeführt. Neben der handelsüblichen MSR-Technik wurden zwei unterschiedliche Systeme zur Bestimmung der Methanolkonzentration sowie eine Trübungssonde zur Bestimmung der Zelldichte eingesetzt. Die Anbindung einer HPLC ermöglichte eine Atline-Beobachtung der Expression des Zielproteins 1-3del I-TAC.

Die Prozessführungsstrategie sah zunächst die Zellanzucht in einem Glycerol-Batch unter nicht limitierenden Bedingungen und in einem Glycerol-Fed Batch mit reduzierter spezifischer Wachstumsrate unter Substratlimitierung vor. Zur Induktion der Proteinexpression folgte ein Wechsel des zugefütterten Substrats von Glycerol zu Methanol. Die Regelung der Methanolkonzentration in der Medienphase verlief mit beiden Methanol-Messsystemen sehr gut, das HPLC-System lieferte sowohl in der Produktionsphase als auch in den nachfolgenden Sequenzen über lange Zeit ein stabiles Messsignal.

Das Konzept der Sequentiell Integrierten Prozessführung beinhaltet einen zyklischen Wechsel von Produktions-, Filtrations-, Ernte- und Refreshphasen. Für die Steuerung des Ablaufs der einzelnen Sequenzen wurden die aus dem Trübungssignal S_{turb} berechnete Zelldichte c_{XLturb} , die Masse des Reaktionsvolumens m_L und die Masse des produktbehafteten Mediums im Permeattank m_P verwendet. Die Waagensignale m_L und m_P bildeten ein sicheres Umschaltkriterium für die Steuerung des vollautomatischen Ablaufs. Die Zelldichte c_{XLturb} hingegen war als Indikator nur bedingt einsetzbar. Durch mathematische Anpassung an die experimentellen Daten konnte kein geeigneter Parametersatz für die automatisierte Prozessführung ermittelt werden.

Die integrierte Bioprozessführung wurde erfolgreich mit dem in der Prozessleitsoftware MFCS/win implementierten S88-Rezepturmodul umgesetzt. Der vollautomatische Prozess ermöglichte dabei die Abtrennung des Proteins noch während der laufenden Kultivierung. Darüber hinaus wurde ein modulares Visual Basic Tool zur Regelung des Gelöstsauerstoffpartialdrucks pO_2 , zur automatischen Batchende-Erkennung und zur μ -Steuerung in der Fed Batch-Phase verwendet. Die Module wurden über zyklische Variablen aus dem S88-Rezept heraus an- und abgeschaltet.

Für die Optimierung der Herstellung von zwei sich durch eine Punktmutation unterscheidenden zytoplasmatischen Fragmenten des humanen EphB2-Rezeptors mit *E. coli* wurde zunächst ein Sechsfach-Parallelbioreaktorsystem aufgebaut und in Betrieb genommen. Das System wurde um verschiedene Analysen- und Automatisierungskomponenten erweitert und er-

folgreich an die SCADA-Software MFCS/win angebunden. Mit dem S88-Rezepturmodul und durch die Anbindung des im Rahmen dieser Arbeit entwickelten MATLAB[®]-Programms MAC4MFCS wurde einerseits der vollautomatische Parallelbetrieb der einzelnen Kessel gewährleistet sowie andererseits ein erweitertes Prozessmonitoring etabliert.

Im Fokus der optimierten Proteinproduktion wurden die gängigsten verfahrenstechnischen Strategien gewählt, bei denen ein Einfluss auf die Steigerung der Syntheserate des Zielproteins vermutet wurde. Mit Hilfe der Statistischen Versuchsplanung wurde der Einfluss der zellspezifischen Wachstumsrate μ_w im Fed Batch, der Flüssigphasentemperatur ϑ_{Lw} ab Induktion, der molaren IPTG-Konzentration C_{IPTGw} und der eingesetzten *E. coli* Mutante MSD5247 bzw. MSD5248 auf die Raum-Zeit-Ausbeute an zellinternen löslichen Protein untersucht.

Aus Vorversuchen, die den Zusammenhang von Kultivierungstemperatur und Wachstumsrate sowie Acetatbildung und Wachstumsrate verdeutlichen, konnte der reguläre Versuchsbereich durch minimale und maximale Faktorstufen der quantitativen Einflussgrößen festgelegt werden. Durch die Verwendung des Reef Coral Fluorescent Protein ZsGreen als Expressionsreporter wurde eine Zielproteinanalytik über den Einsatz eines Fluoreszenzreaders erfolgreich etabliert.

Die Durchführung der Screeningversuche ergab, dass die molare IPTG-Konzentration im untersuchten Bereich von 0.1-1.5 mmol⁻¹ keinen statistisch signifikanten Einfluss auf die lösliche Proteinexpression hat. Des Weiteren konnte für den Stamm MSD5248 bestätigt werden, dass die in die Aminosäuresequenz des Zielproteins eingefügte Punktmutation [L767P] einen positiven Effekt auf die lösliche Proteinexpression hat. Die lösliche Raum-Zeit-Ausbeute war bei hoher Wachstumsrate sowie niedriger Temperatur circa zweieinhalbmal größer als bei MSD5247.

Die Optimierungsexperimente wurden daher mit dem Stamm MSD5248 durchgeführt, wobei die Festlegung der Faktorstufen für die signifikanten Faktoren zellspezifische Wachstumsrate und Flüssigphasentemperatur nach der Methode des Steilsten Anstiegs erfolgte. IPTG wurde mit der niedrigsten untersuchten Konzentration von 0.1 mmol⁻¹ zugegeben. Die in maximaler Raum-Zeit-Ausbeute an zellinternen löslichen Protein resultierende Faktorkombination wurde mit $\mu_w = 0.25 \text{ h}^{-1}$ und $\vartheta_{Lw} = 27.1 \text{ °C}$ ermittelt.

Im abschließenden Robustheitstest konnte Robustheit der Antworten auf kleine Faktoränderungen bestätigt werden. Für sämtliche Schritte der Statistischen Versuchsplanung wurden die Antwortgrößen mit dem W-Test auf Normalverteilung und mit dem Grubbs-Test auf Ausreißer geprüft.

9 Anhang

9.1 Eingesetzte Kultivierungsmedien

9.1.1 Verwendete Medien für die Kultivierung von *P. pastoris*

Es werden insgesamt fünf unterschiedliche Medien für die Kultivierung benötigt. Das VK II Medium ist für die Herstellung des Inokulums aus einer Kryokultur vorgesehen. Für die weiteren Prozessphasen werden ein Batchmedium, zwei Fed Batch-Medien und ein Refresh-Medium benötigt. In Tabelle 9.1 ist die Zusammensetzung der eingesetzten Medien aufgeführt. Zum Herstellen der verschiedenen Medien wurden zunächst Stammlösungen angesetzt. Der pH-Wert des VK II Mediums, des Batch- und des Fed Batch-Mediums wird mit H_3PO_4 (85 %) und NH_3 (25 %) auf pH = 5 eingestellt.

Tabelle 9.1: Zusammensetzung der eingesetzten Kultivierungsmedien für *P. pastoris*

Medienkomponente		Konzentration				
		VK II	Batch	Fed Batch I	Fed Batch II	Refresh
$\text{C}_3\text{H}_8\text{O}_3$	[gl^{-1}]	25.4	30.0	350.0	-	-
CH_3OH	[gl^{-1}]	-	-	-	790.0	2.0
H_3PO_4	[ml^{-1}]	3.5	26.0	-	-	26.0
K_2SO_4	[gl^{-1}]	2.4	18.0	-	-	18.0
KOH	[gl^{-1}]	0.65	4.0	-	-	4.0
$\text{CaSO}_4 \cdot 2 \text{H}_2\text{O}$	[gl^{-1}]	0.15	0.9	-	-	0.9
$\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$	[gl^{-1}]	1.95	14.0	-	-	14.0
Biotin	[ml^{-1}]	0.4	8.0	10.0	10.0	10.0
PTM1	[ml^{-1}]	1.1	8.0	10.0	10.0	10.0

Das Glycerol und die Grundalze des VK II Mediums werden bei 121 °C für 20 min autoklaviert und nach Abkühlen mit Biotin und PTM1 versetzt. Die verdünnte Zellsuspension aus der Glycerolkonserve wird auf 4x600 ml des VK II Mediums aufgeteilt und in 2000 ml Erlenmeyerkolben mit Schikanen und Kapsenbergkappe im Inkubationsschüttler (CERTOMAT® BS-1, Sartorius) bei 200 min^{-1} und 30 °C für 24 h kultiviert. Nach einer mikroskopischen Betrachtung der jeweiligen Vorkulturen wird die optische Dichte ermittelt. Liegen die Messwerte zwischen 1 und 3, werden die zwei Vorkulturen mit der höchsten OD vereinigt. Nun stehen 1200 ml Inokulum mit einer Biotrockenmasse von 1.5 bis 3.0 gl^{-1} zur Verfügung.

Für das VK II Medium wird eine geringere Glycerolkonzentration als für das Batch-Medium gewählt, um die Gefahr eines Metabolischen Overflows im Schüttelkolben zu vermeiden.

Ansonsten bietet es sich an, die weitere Zusammensetzung genau so wie für das Batch-Medium zu wählen, um eine schnelle Adaption zu gewährleisten und damit die lag-Phase so kurz wie möglich zu halten. Das Glycerol und die Grundsalze des Batch-Mediums werden im Fermenter vorgelegt und *in-situ* bei 121 °C für 30 min sterilisiert. Nach Abkühlung wird das fehlende Biotin und PTM1 hinzu gegeben.

Die beiden Fed Batch-Medien sowie das Refresh-Medium werden bei 121 °C für 20 min im Autoklaven sterilisiert und nach Abkühlen mit Biotin und PTM1 versetzt.

Die PTM1-Stammlösung und die 0.2 g^l⁻¹ Biotin-Stammlösung (Vitamin) werden unter Verwendung von Cellulose-Nitrat Membranfiltern mit einer Porengröße von 0.45 µm sterilfiltriert und zur weiteren Verwendung bei 8 °C im Kühlschrank gelagert.

Tabelle 9.2: Biotin-/ PTM1-Stammlösung (trace salts – Spurenelemente)

Komponente	Konzentration	Komponente	Konzentration
CoCl ₂ · 6 H ₂ O	0.5 g ^l ⁻¹	KI	0.1 g ^l ⁻¹
ZnSO ₄ · 7 H ₂ O	20.0 g ^l ⁻¹	H ₂ SO ₄	5.0 g ^l ⁻¹
FeSO ₄ · 7 H ₂ O	65.0 g ^l ⁻¹	H ₃ BO ₃	0.02 g ^l ⁻¹
CuSO ₄ · 5 H ₂ O	6.0 g ^l ⁻¹	Na ₂ MoO ₄ · 2 H ₂ O	0.2 g ^l ⁻¹
MnSO ₄ · H ₂ O	3.0 g ^l ⁻¹	C ₁₀ H ₁₆ N ₂ O ₃ S	0.2 g ^l ⁻¹

9.1.2 Verwendete Medien für die Kultivierung von *E. coli*

Die Zusammensetzung des synthetischen Mediums basiert im Wesentlichen auf einem Medium nach Riesenberg *et al.* (1991). Es wurden lediglich Änderungen an der Glukosekonzentration, um die Acetatproduktion im Batch zu begrenzen, und der Ammoniumkonzentration, um eine N-Limitierung zu vermeiden, vorgenommen.

Zum Herstellen der verschiedenen Medien wurden zunächst Stammlösungen angesetzt. Die Stammlösungen werden, mit Ausnahme der Glukose- und Magnesiumsulfatlösung, zusammengegeben und für 15 min bei 121 °C im Autoklaven sterilisiert. Um eine starke Karamellisierung der Glukose, sowie eine Maillard-Reaktion und Ausfällungen zu verhindern, werden Glukose und Magnesiumsulfat getrennt sterilisiert und unter der Sterilwerkbank mit dem restlichen Medium zusammengeführt.

Tetracyclin wird, aufgrund seiner Temperaturinstabilität, per Sterilfiltration hinzugefügt. Anschließend wird der pH-Wert mit 5 M NaOH auf 6.8 eingestellt. Tabelle 9.3 zeigt die Zusammensetzung der verschiedenen Kultivierungsmedien.

Tabelle 9.3: Zusammensetzung der eingesetzten Kultivierungsmedien für *E. coli*

Medienkomponente		Konzentration			
		Stammlsg.	VKII	Batch	Fed Batch
$C_6H_{12}O_6$	[g l^{-1}]	200 ¹ /400 ²	15.0	20.0	300.0
$MgSO_4 \cdot 7 H_2O$	[g l^{-1}]	400	1.2	1.2	20.0
KH_2PO_4	[g l^{-1}]	266	13.3	13.3	-
$(NH_4)_2HPO_4$	[g l^{-1}]	80	8.0	7.0	-
Citronensäure	[g l^{-1}]	170	1.7	1.7	-
Fe(III)Citrat	[mg l^{-1}]	10000	100.0	100.0	40.0
$CoCl_2 \cdot 6 H_2O$	[mg l^{-1}]	250	2.5	2.5	4.0
$MnCl_2 \cdot 4 H_2O$	[mg l^{-1}]	1500	15.0	15.0	23.5
$CuCl_2 \cdot 2 H_2O$	[mg l^{-1}]	150	1.5	1.5	2.5
H_3BO_3	[mg l^{-1}]	300	3.0	3.0	5.0
$Na_2MoO_4 \cdot 2 H_2O$	[mg l^{-1}]	250	2.5	2.5	4.0
$Zn(CH_3COO)_2 \cdot 2 H_2O$	[mg l^{-1}]	1300	13.0	13.0	16.0
EDTA	[mg l^{-1}]	840	8.4	8.4	13.0
Tetracyclin	[mg l^{-1}]	-	50.0	50.0	50.0

¹ Konzentration der Stammlösung für VKII und Batch

² Konzentration der Stammlösung für Fed Batch

9.2 Bestimmung der optischen Dichte, der Biotrockenmasse und der Biofeuchtmasse

Die prozessbegleitende Messung der optischen Dichte (OD) erfolgte photometrisch bei einer Wellenlänge von 600 nm mit einem UV/vis-Photometer (Ultrospec 3000, Fa. Amersham Biosciences) in Polystyrol-Einmalküvetten mit einer Schichtdicke von 1 cm. Die Proben wurden gegebenenfalls mit VE-Wasser verdünnt, so dass die Extinktion zwischen 0.1 und 0.8 AU (*P. pastoris*) beziehungsweise 0.1 und 0.6 AU (*E. coli*) lag. Es erfolgte eine Doppelbestimmung.

Zur Bestimmung der Biotrockenmasse (BTM) wurden 1.5 ml Mikroreaktionsgefäße für 24 h bei 80 °C getrocknet und anschließend ausgewogen. 1 ml der zu bestimmenden Probe wurde in die Mikroreaktionsgefäße pipettiert und für 5 min bei 14000 min⁻¹ zentrifugiert (Centrifuge 5417C, Fa. Eppendorf). Die Probenahme erfolgte als Doppelbestimmung. Der Überstand wurde für die Offline-Substratanalytik in separate Mikroreaktionsgefäße dekantiert und das Zellpellet für 24 h bei 80 °C bis zur Gewichtskonstanz getrocknet.

Zur Bestimmung der Biofeuchtmasse (BFM) wurde 1 ml Probe in ein vorher ausgewogenes Mikroreaktionsgefäß pipettiert und für 5 min bei 14000 min^{-1} zentrifugiert. Der Überstand wurde verworfen, das Zellpellet einmal mit 1 ml VE-Wasser resuspendiert und erneut für 5 min bei 14000 min^{-1} zentrifugiert. Anschließend wurde das Mikroreaktionsgefäß mit den feuchten Zellen ausgewogen. Es erfolgte eine Doppelbestimmung.

Basierend auf der BTM- und BFM-Bestimmung wurde das Biofeucht- zu Biotrockenmasseverhältnis $\alpha_{Z/X}$ über lineare Regression ermittelt und für beide *E. coli* Stämme mit $\alpha_{Z/X} = 5.65$ zeitinvariant angenommen (Tatge, 2008).

9.3 Acetatmessung über enzymatisches Testkit

Um den in Kapitel 6.3.2 beschriebenen Zusammenhang zwischen Acetatbildung und Wachstumsrate zu untersuchen, wurde die Acetatkonzentration mit einem enzymatischen UV-Test (Nr. 0148 261) der Fa. Roche Diagnostics bestimmt. Dabei wurde die enzymatische Umsetzung über die NADH-Extinktion mit einem UV/vis-Photometer (Ultrospec 3000, Fa. Amersham Biosciences) bei einer Wellenlänge von 340 nm detektiert.

9.4 Vorgehensweise für die Zielproteinanalytik bei *E. coli* Kultivierungen

Während der Fed Batch- und Produktionsphase einer Kultivierung wurden bei jeder Probenahme drei Proben (a, b, c) für die Zielproteinanalytik bei einer Temperatur von -20 °C zurückgestellt. Dafür wurde je 1 ml der Kulturbrühe in Mikroreaktionsgefäße pipettiert und für 5 min bei 14000 min^{-1} zentrifugiert. Der Überstand verblieb auf dem abzentrifugierten Zellpellet.

Bild 9.1 zeigt die schematische Vorgehensweise zur Probenaufbereitung für die Analytik auf lösliches zellinternes (Überstände Ü1 bis Ü3) und unlösliches zellinternes/zellexternes (Überstand Ü4) Zielprotein. Wie bereits in Kapitel 6.4.2 beschreiben, wird die unlösliche zellexterne Proteinfraction durch die Zentrifugation wieder in die feuchte Zellmasse zurückgeführt, zusammen mit der zellinternen Inclusion Body Fraktion analysiert und daher als Gesamtfraction in der Zellphase Z notiert.

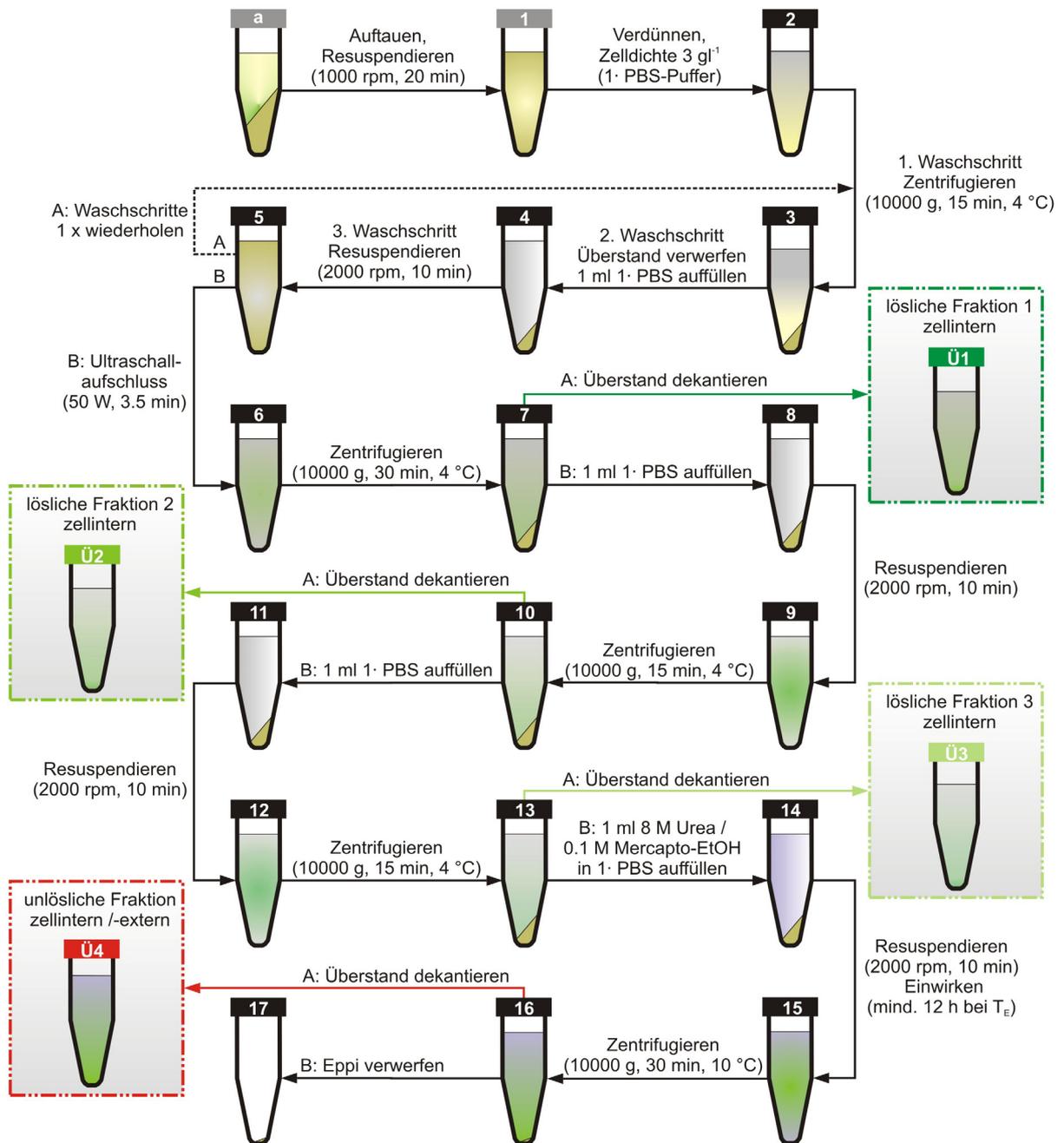


Bild 9.1: Schematische Vorgehensweise zur Probenaufbereitung für die Analytik auf zellinternes lösliches und zellinternes /-externes unlösliches Zielprotein

Im den folgenden Schritten wird die Vorgehensweise zur Probenaufbereitung für die Zielproteinanalytik detailliert beschrieben.

1. Die Proben werden zunächst im gefrorenen Zustand mit einem Schüttler (Vibrax[®] VXR basic, Fa. IKA) bei 1000 min^{-1} für ca. 20 min resuspendiert. Die Proben sollten dabei möglichst weit außen im Schüttelaufsatz platziert werden.

Das Flüssigvolumen der aufgetauten und resuspendierten Probe im Mikroreaktionsgefäß E1,

$$V_{E1}(t_{Sj}) = V_{M1}(t_{Sj}) + V_{Z1}(t_{Sj}) \stackrel{!}{=} V_E = 1000 \mu\text{l}, \quad (9.1)$$

wird als Summe aus dem Medienvolumen V_{M1} und dem Biofeuchtvolumen V_{Z1} zum Zeitpunkt $t = t_{Sj}$ definiert.

V_{E1} entspricht immer 1000 μl für jede genommene Probe und wird daher mit V_E als konstant definiert.

2. Im nächsten Schritt werden alle Proben auf eine einheitliche Zelldichte von 3 gl^{-1} verdünnt. Als Lösungsmittel wird bei Raumtemperatur aufbewahrter 1·PBS-Puffer mit einem pH-Wert von 7.4 verwendet. Alle Verdünnungen werden für die weiteren Schritte auf Eis zwischengelagert.

Die im Reaktionsgefäß E1 vorhandene Zelldichte c_{XE1} ,

$$c_{XE1}(t_{Sj}) = c_{XL}(t_{Sj}), \quad (9.2)$$

entspricht der für diese Probe offline bestimmten Zelldichte c_{XL} .

Das für den ersten Verdünnungsschritt erforderliche Teilvolumen der zellhaltigen Suspension aus Reaktionsgefäß E1,

$$V'_{E1}(t_{Sj}) = \frac{c_{XE2}}{c_{XL}(t_{Sj})} \cdot V_E, \quad (9.3)$$

errechnet sich aus dem Verhältnis der gewählten Zelldichte c_{XE2} von 3 gl^{-1} und der offline bestimmten Zelldichte c_{XL} multipliziert mit dem konstanten Lösungsvolumen $V_{E2} = V_E$.

Für Gleichung (9.3) gilt die Bedingung $c_{XE2} < c_{XL}$. Diese Bedingung ist für die durchgeführten Versuche immer erfüllt, da erst ab Beginn der Fed Batch Phase Proben genommen wurden und die offline bestimmte Zelldichte c_{XL} ab diesem Zeitpunkt den Wert von 3 gl^{-1} stets übersteigt.

Das für den zweiten Verdünnungsschritt erforderliche Volumen an 1·PBS-Puffer in Gefäß E2,

$$V_{B2}(t_{Sj}) = V_E - V'_{E1}(t_{Sj}) = V_E - \frac{c_{XE2}}{c_{XL}(t_{Sj})} \cdot V_E = \frac{c_{XL}(t_{Sj}) - c_{XE2}}{c_{XL}(t_{Sj})} \cdot V_E, \quad (9.4)$$

kann aus der Differenz von Lösungsvolumen V_E und dem in Gleichung (9.3) berechneten Teilvolumen V'_{E1} bestimmt werden.

Für jede Probe ergibt sich so ein individueller Verdünnungsfaktor $K_{E1/2}$,

$$K_{E1/2}(t_{Sj}) = \frac{c_{XL}(t_{Sj})}{c_{XE2}} > 1, \quad (9.5)$$

der aus dem reziproken Zelldichteverhältnis in Gleichung (9.3) ermittelt werden kann.

3. In den unter Punkt 3 bis Punkt 5 aufgeführten Waschschritten der verdünnten Proben sollen Verunreinigungen aus der zellhaltigen Suspension entfernt werden. Im ersten Waschschriff werden die Proben 15 min bei einer relativen Zentrifugalbeschleunigung von 10000 g gekühlt zentrifugiert. Die Mikro-Kühlzentrifuge 5417R (Fa. Eppendorf) wird dabei mit einer Temperiereinstellung von 4 °C betrieben.
4. Im zweiten Waschschriff wird der PBS- sowie medienhaltige Überstand aus dem Reaktionsgefäß 3 verworfen und das Gefäß auf Zellstoff ausgeklopft. Anschließend wird 1000 µl frischer 1·PBS-Puffer aufgefüllt.
5. Im dritten Waschschriff werden die Proben erneut mit dem Schüttler bei 2000 min⁻¹ für 10 min resuspendiert. Die Waschprozedur wird einmal wiederholt.
6. Die zweimal gewaschenen Proben stehen nun für den Aufschluss im Ultraschallhomogenisator Labsonic U (Fa. Sartorius BBI Systems) zur Verfügung. Zur besseren Wärmeabfuhr werden die Reaktionsgefäße auf Eis gekühlt. Alle zu analysierenden Proben werden dann für 3.5 min bei 50 W Leistungseintrag aufgeschlossen.

Bedingt durch die in Punkt 4 gewählte Zugabe von 1000 µl 1·PBS-Puffer, wächst das Flüssigvolumen der aufzuschließenden Probe im Reaktionsgefäß E5,

$$V_{E5} = V_{B4} + V_{Z5}, \quad (9.6)$$

um das Biofeuchtvolumen V_{Z5} ,

$$V_{Z5} = V_{Z2} = \frac{\alpha_{Z/X} \cdot m_{Z2}}{\rho_Z} = \frac{\alpha_{Z/X} \cdot V_E \cdot c_{XE2}}{\rho_Z}, \quad (9.7)$$

das dem nach der Verdünnung in Punkt 2 erhaltenen Biofeuchtvolumen V_{Z2} entspricht.

Unter Kenntnis des konstanten Lösungsvolumens V_E und der fest vorgegebenen Zelldichte c_{XE2} sowie dem experimentell bestimmten Biofeucht- zu Biotrockenmasseverhältnis $\alpha_{Z/X}$ von 5.65 und einer definierten Dichte der Biofeuchtmasse ρ_Z von 1000 gl⁻¹ kann das Biofeuchtvolumen V_{Z5} ,

$$V_{Z5} = \frac{5.65 \cdot 1 \cdot 10^{-3} \cdot 1.3 \text{ gl}^{-1}}{1000 \text{ gl}^{-1}} = 16.95 \text{ µl}, \quad (9.8)$$

als zeitinvarianter Parameter ermittelt werden.

Das durch den Zellaufschluss freigesetzte Zielprotein befindet sich dann in einem Gesamtvolumen V_{E5} ,

$$V_{E5} = V_E + V_{Z5} = 1016.95 \mu\text{l}, \quad (9.9)$$

und ist nunmehr in einem Volumen über 1 ml gelöst.

Mit der in Reaktionsgefäß E5 vorhandenen Zelldichte c_{XE5} ,

$$c_{XE5} = \frac{m_{XE5}}{V_{E5}} = \frac{m_{XE2}}{V_{E5}} = \frac{V_E \cdot c_{XE2}}{V_{E5}} = 2.95 \text{ gl}^{-1}, \quad (9.10)$$

wird daher der in Gleichung (9.5) berechnete Faktor $K_{E1/2}$ durch einen neuen Verdünnungsfaktor $K_{E1/5}$,

$$K_{E1/5}(t_{Sj}) = \frac{c_{XL}(t_{Sj})}{c_{XE5}}, \quad (9.11)$$

ersetzt. Die zum Ende der Analytik ermittelten Messergebnisse müssen mit diesem Verdünnungsfaktor korrigiert werden.

- Nach dem Ultraschallaufschluss schließen sich erneute Waschschriffe an. Im ersten Waschschriff werden die auf Eis gelagerten Proben für 30 min bei 10000 g gekühlt zentrifugiert. Die Mikro-Kühlzentrifuge 5417R wird dabei mit 4 °C betrieben.

Über die Zentrifugation wird die lösliche und unlösliche Zielproteinfraktion voneinander getrennt. Im Überstand befindet sich durch den Dichteunterschied von Lysat und Zelldebris bzw. Inclusion Bodies nur zellinternes lösliches Protein.

Im Pellet hingegen befindet sich unlösliches Protein, das einerseits erst durch den Ultraschallaufschluss, andererseits aber auch bereits während der Kultivierung durch abgestorbene und lysierte Zellen freigesetzt wurde.

- Im zweiten Waschschriff wird die im Überstand Ü1 befindliche lösliche Fraktion in ein neues entsprechend gekennzeichnetes Reaktionsgefäß überführt und bis zur Messung bei -20 °C gelagert. Das Gefäß wird auf Zellstoff ausgeklopft. Anschließend wird 1000 µl frischer PBS-Puffer aufgefüllt.
- Im dritten Waschschriff werden die Proben erneut mit dem Schüttler bei 2000 min⁻¹ für 10 min resuspendiert.
- Die in den Punkten 7 bis 9 erläuterte Waschprozedur wird einmal wiederholt, so dass sich im Überstand Ü2 eventuelle Rückstände der löslichen Fraktion befinden. Die Zentrifugationsdauer der auf Eis gelagerten Proben wird hier jedoch auf 15 min reduziert.

- siehe 8.

12. siehe 9.
13. Nach dem Resuspendieren werden die Proben wie in Punkt 10 erneut für 15 min bei einer relativen Zentrifugalbeschleunigung von 10000 g gekühlt bei 4 °C zentrifugiert.
14. Im letzten Schritt zur Trennung der löslichen von der unlöslichen Fraktion wird der Überstand Ü3 in ein neues entsprechend gekennzeichnetes Reaktionsgefäß dekantiert und bis zur Messung bei -20 °C gelagert. Das Gefäß wird auf Zellstoff ausgeklopft.

Die Reaktionsgefäße werden nun ungekühlt bei Raumtemperatur mit 1000 µl Inclusion Body Aufschlusspuffer versetzt, so dass die darin befindliche unlösliche Proteinfraction für die Messung zugänglich wird. Der IB-Aufschlusspuffer besteht aus in 1·PBS gelöstem 8 M Harnstoff sowie 0.1 M β-Mercaptoethanol mit einem pH-Wert von 9.5.
15. Die mit dem Schüttler bei 2000 min⁻¹ für 10 min resuspendierten Proben werden mindestens 12 h bei Raumtemperatur aufbewahrt.
16. Nach dem mehrstündigen Einwirken werden die Proben 30 min mit 10000 g gekühlt zentrifugiert. Die Mikro-Kühlzentrifuge 5417R wird dabei mit 10 °C betrieben, da es bei einer zu niedrigen Temperatur zu Harnstoffausfällungen kommen kann.
17. Der erhaltene Überstand Ü4 wird in ein neues entsprechend gekennzeichnetes Reaktionsgefäß überführt und direkt mit dem Tecan GENiosTM Reader gemessen. Das eventuell vorhandene Pellet wird verworfen.

Die Überstände Ü1 bis Ü4 werden mit der Wellenlängenkombination $\lambda_{\text{ex}} = 480 \pm 20$ nm und $\lambda_{\text{em}} = 535 \pm 20$ nm zunächst unverdünnt gemessen. Falls der Messbereich des Readers (max. 60000 RFU) überschritten wird, werden die Überstände Ü1 bis Ü3 mit 1·PBS-Puffer verdünnt, der Überstand Ü4 dagegen mit IB-Aufschlusspuffer.

Da die Messung als Doppelbestimmung erfolgt, werden je 200 µl Probe in zwei Wells der verwendeten 96-Well-Platten (MicroWellTM MaxiSorpTM Platten, Fa. Nunc) pipettiert. Die ersten beiden Wells sind für die Leerwertmessung reserviert.

Zur Messung wird die Methode "ZsGreen_STD.mth" mit den in Tabelle 9.4 aufgeführten Einstellungen verwendet.

Tabelle 9.4: Parameter für die Messung mit der Methode "ZsGreen_STD.mth"

Parameter	Adjustment	Parameter	Adjustment
measurement mode	fluorescence top	number of kinetic cycles	3
excitation wavelength	480 nm	kinetic interval	minimal
emission wavelength	535 nm	valid temperature range	36 - 38 °C
gain	manual, 65	incubation time	300 s
number of flashes	100	shake duration	orbital, normal, 60 s
lag time	0 µs	shake settle time	120 s
integration time	75 µs	unit	RFU
plate definition file	GRE96ft		

Das optische Signal der Fluoreszenzmessung für die lösliche zellinterne Fraktion,

$$S_{48/53Zsol}(t_{Sj}) = K_{E1/5}(t_{Sj}) \cdot [S_{48/53Zsol_raw}(t_{Sj}) - S_{48/53Zsol_blk}], \quad (9.12)$$

resultiert nach Abzug des Leerwertes $S_{48/53Zsol_blk}$ vom Messwert $S_{48/53Zsol_raw}$ und anschließender Multiplikation mit dem für jede Probe individuellen Verdünnungsfaktor $K_{E1/5}$. Der Leerwert $S_{48/53Zsol_blk}$ stammt aus der Messung des 1·PBS-Puffers.

Analog dazu resultiert das optische Signal der Fluoreszenzmessung für die unlösliche auf die Zellphase Z bezogene Gesamtfraction $S_{48/53Z_IB}$,

$$S_{48/53Z_IB}(t_{Sj}) = K_{E1/5}(t_{Sj}) \cdot [S_{48/53Z_IB_raw}(t_{Sj}) - S_{48/53Z_IB_blk}], \quad (9.13)$$

nach Abzug des Leerwertes $S_{48/53Z_IB_blk}$ vom Messwert $S_{48/53Z_IB_raw}$ und Multiplikation mit dem Verdünnungsfaktor $K_{E1/5}$. Der Leerwert $S_{48/53Z_IB_blk}$ resultiert aus der Messung des aus Harnstoff und Mercapto-EtOH bestehenden IB-Aufschlusspuffers.

9.5 Automatisierungsaufgaben

In Tabelle 9.5 sind die Automatisierungsaufgaben während des dreiphasigen Produktionsablaufs tabellarisch zusammengefasst. Die Reglermodi (OFF, MAN, AUTO, CAS), Sollwerte und Stellgrößen sind für jeden DCU-Regler und virtuellen MFCS-Regler aufgeführt.

Tabelle 9.5: Automatisierungsaufgaben während des dreiphasigen Produktionsablaufs

Controller		Process Phase				
Name	Property	Stand-By	Inoculation	Batch	Fed Batch	Production
liquid temperature (PI)	mode	AUTO	AUTO	AUTO	AUTO	AUTO
	setpoint	$\vartheta_{LW} = 30 \text{ }^\circ\text{C}$				
	output	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}
jacket temperature (PI)	mode	CASCADE	CASCADE	CASCADE	CASCADE	CASCADE
	setpoint	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}
	output	HEAT / COOL				
pH (PI)	mode	MAN	MAN	AUTO	AUTO	AUTO
	setpoint	$pH_w = 5$				
	output	$y_{pH} = 0 \%$	$y_{pH} = 0 \%$	ACID / BASE	ACID / BASE	ACID / BASE
pressure (PI)	mode	AUTO	MAN	AUTO	AUTO	AUTO
	setpoint	$p_{GW} = 500 \text{ mbar}$	-	$p_{GW} = 500 \text{ mbar}$	$p_{GW} = 500 \text{ mbar}$	$p_{GW} = 500 \text{ mbar}$
	output	y_{pG}	$y_{pG} = 100 \%$	y_{pG}	y_{pG}	y_{pG}
pO2 (PI)	mode	OFF	OFF	AUTO	AUTO	AUTO
	setpoint	-	-	$pO_{2w} = 10 \%$	$pO_{2w} = 10 \%$	$pO_{2w} = 10 \%$
	output	-	-	$F_{nAIR_SP}/N_{St_SP}/X_{OGin_SP}$	$F_{nAIR_SP}/N_{St_SP}/X_{OGin_SP}$	$F_{nAIR_SP}/N_{St_SP}/X_{OGin_SP}$
AIR aeration (SPC)	mode	AUTO	AUTO	CASCADE	CASCADE	CASCADE
	setpoint	$F_{nAIR_SET} = 12 \text{ lmin}^{-1}$	$F_{nAIR_SET} = 0 \text{ lmin}^{-1}$	F_{nAIR_SP}	F_{nAIR_SP}	F_{nAIR_SP}
	output	$F_{nAIRw} = F_{nAIR_SET}$	$F_{nAIRw} = F_{nAIR_SET}$	$F_{nAIRw} = F_{nAIR_SP}$	$F_{nAIRw} = F_{nAIR_SP}$	$F_{nAIRw} = F_{nAIR_SP}$
stirrer speed (SPC)	mode	AUTO	AUTO	CASCADE	CASCADE	CASCADE
	setpoint	$N_{St_SET} = 200 \text{ rpm}$	$N_{St_SET} = 200 \text{ rpm}$	N_{St_SP}	N_{St_SP}	N_{St_SP}
	output	$N_{Stw} = N_{St_SET}$	$N_{Stw} = N_{St_SET}$	$N_{Stw} = N_{St_SP}$	$N_{Stw} = N_{St_SP}$	$N_{Stw} = N_{St_SP}$
xOGin_SP (SPC)	mode	OFF	OFF	CASCADE	CASCADE	CASCADE
	setpoint	-	-	X_{OGin_SP}	X_{OGin_SP}	X_{OGin_SP}
	output	-	-	$X_{OGinw} = X_{OGin_SP}$	$X_{OGinw} = X_{OGin_SP}$	$X_{OGinw} = X_{OGin_SP}$
oxygen aeration (SPC)	mode	OFF	OFF	AUTO	AUTO	AUTO
	setpoint	-	-	F_{nO2_SET}	F_{nO2_SET}	F_{nO2_SET}
	output	-	-	$F_{nO2_w} = F_{nO2_SET}$	$F_{nO2_w} = F_{nO2_SET}$	$F_{nO2_w} = F_{nO2_SET}$
feed rate 1 (SPC)	mode	OFF	OFF	OFF	CASCADE	OFF
	setpoint	-	-	-	F_{R2rel_SP}	-
	output	-	-	-	F_{R2relw}	-
Methanol-FMC (PI)	mode	OFF	OFF	OFF	OFF	AUTO
	setpoint	-	-	-	-	$C_{S2Mw} = 2 \text{ g}^{-1}$
	output	-	-	-	-	F_{R2rel}
feed rate 2 (SPC)	mode	OFF	OFF	OFF	OFF	CASCADE
	setpoint	-	-	-	-	F_{R2rel}
	output	-	-	-	-	F_{R2relw}

Tabelle 9.6 gibt eine detaillierte Übersicht der Reglereinstellungen während der unterschiedlichen Phasen der Sequentiell Integrierten Bioprozessführung.

Tabelle 9.6: Reglereinstellungen während der Integrierten Prozessführung

Controller		Process Phase				
Name	Property	Cake Formation	Microfiltration	Cell Harvest	Media Refresh	Production
liquid weight (PI)	mode	OFF	AUTO	MAN	MAN	MAN
	setpoint	-	$m_{Lw} = 12 \text{ kg}$	-	-	-
	output	-	F_{R3rel}	$F_{R3rel} = 0 \%$	$F_{R3rel} = 50 \%$	$F_{R3rel} = 0 \%$
feed rate 3 (SPC)	mode	OFF	CASCADE	CASCADE	CASCADE	CASCADE
	setpoint	-	F_{R3rel}	F_{R3rel}	F_{R3rel}	F_{R3rel}
	output	-	F_{R3relw}	F_{R3relw}	F_{R3relw}	F_{R3relw}
harvest rate (SPC)	mode	OFF	OFF	AUTO	OFF	OFF
	setpoint	-	-	$F_{Hrel_SET} = 50 \%$	-	-
	output	-	-	$F_{Hrelw} = F_{Hrel_SET}$	-	-
liquid temperature (PI)	mode	AUTO	AUTO	AUTO	AUTO	AUTO
	setpoint	$\vartheta_{Lw} = 30 \text{ }^\circ\text{C}$				
	output	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}
jacket temperature (PI)	mode	CASCADE	CASCADE	CASCADE	CASCADE	CASCADE
	setpoint	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}	ϑ_{DJ_SP}
	output	HEAT / COOL				
pH (PI)	mode	AUTO	AUTO	AUTO	AUTO	AUTO
	setpoint	$pH_w = 5$				
	output	ACID / BASE				
pressure (PI)	mode	AUTO	AUTO	AUTO	AUTO	AUTO
	setpoint	$p_{Gw} = 500 \text{ mbar}$				
	output	y_{pG}	y_{pG}	y_{pG}	y_{pG}	y_{pG}
cross flow rate (SPC)	mode	CASCADE	CASCADE	CASCADE	CASCADE	CASCADE
	setpoint	y_{feed}	y_{feed}	y_{feed}	y_{feed}	y_{feed}
	output	$F_{feedrelw}$	$F_{feedrelw}$	$F_{feedrelw}$	$F_{feedrelw}$	$F_{feedrelw}$
feed pressure (PI)	mode	AUTO	AUTO	AUTO	AUTO	AUTO
	setpoint	$p_{feedw} = 1.2 \text{ bar}$	$p_{feedw} = 2.0 \text{ bar}$	$p_{feedw} = 1.2 \text{ bar}$	$p_{feedw} = 1.2 \text{ bar}$	$p_{feedw} = 1.2 \text{ bar}$
	output	y_{feed}	y_{feed}	y_{feed}	y_{feed}	y_{feed}
retentate pressure (PI)	mode	AUTO	AUTO	AUTO	AUTO	AUTO
	setpoint	$p_{retw} = 0.8 \text{ bar}$	$p_{retw} = 1.0 \text{ bar}$	$p_{retw} = 0.8 \text{ bar}$	$p_{retw} = 0.8 \text{ bar}$	$p_{retw} = 0.8 \text{ bar}$
	output	y_{ret}	y_{ret}	y_{ret}	y_{ret}	y_{ret}
permeate pressure (PI)	mode	MAN	AUTO	MAN	MAN	MAN
	setpoint	-	$p_{perw} = 0.5 \text{ bar}$	-	-	-
	output	$y_{per} = 0 \%$	y_{per}	$y_{per} = 0 \%$	$y_{per} = 0 \%$	$y_{per} = 0 \%$
pO2 (PI)	mode	AUTO	AUTO	AUTO	AUTO	AUTO
	setpoint	$pO_{2w} = 10 \%$				
	output	$F_{nAIR_SP}/N_{St_SP}/X_{OGin_SP}$	$F_{nAIR_SP}/N_{St_SP}/X_{OGin_SP}$	$F_{nAIR_SP}/N_{St_SP}/X_{OGin_SP}$	$F_{nAIR_SP}/N_{St_SP}/X_{OGin_SP}$	$F_{nAIR_SP}/N_{St_SP}/X_{OGin_SP}$
AIR aeration (SPC)	mode	CASCADE	CASCADE	CASCADE	CASCADE	CASCADE
	setpoint	F_{nAIR_SP}	F_{nAIR_SP}	F_{nAIR_SP}	F_{nAIR_SP}	F_{nAIR_SP}
	output	$F_{nAIRw} = F_{nAIR_SP}$				
stirrer speed (SPC)	mode	CASCADE	CASCADE	CASCADE	CASCADE	CASCADE
	setpoint	N_{St_SP}	N_{St_SP}	N_{St_SP}	N_{St_SP}	N_{St_SP}
	output	$N_{Stw} = N_{St_SP}$				
xOGin_SP (SPC)	mode	CASCADE	CASCADE	CASCADE	CASCADE	CASCADE
	setpoint	X_{OGin_SP}	X_{OGin_SP}	X_{OGin_SP}	X_{OGin_SP}	X_{OGin_SP}
	output	$X_{OGinw} = X_{OGin_SP}$				
oxygen aeration (SPC)	mode	AUTO	AUTO	AUTO	AUTO	AUTO
	setpoint	F_{nO2_SET}	F_{nO2_SET}	F_{nO2_SET}	F_{nO2_SET}	F_{nO2_SET}
	output	$F_{nO2_w} = F_{nO2_SET}$				
methanol conc. (PI)	mode	AUTO	AUTO	AUTO	AUTO	AUTO
	setpoint	$C_{S2Mw} = 2 \text{ gl}^{-1}$				
	output	F_{R2rel}	F_{R2rel}	F_{R2rel}	F_{R2rel}	F_{R2rel}
feed rate 2 (SPC)	mode	CASCADE	CASCADE	CASCADE	CASCADE	CASCADE
	setpoint	F_{R2rel}	F_{R2rel}	F_{R2rel}	F_{R2rel}	F_{R2rel}
	output	F_{R2relw}	F_{R2relw}	F_{R2relw}	F_{R2relw}	F_{R2relw}

10 Literaturverzeichnis

- Abdelrahim, M.; Konduri, S.; Basha, R.; Philip, P.A.; Baker, C.H. (2010): Angiogenesis: an update and potential drug approaches, *Int J Oncol* 36(1): 5–18.
- Adams, C.P. and Brantner, V.V. (2006): Estimating the cost of new drug development: is it really 802 million dollars? *Health Aff (Millwood)* 25(2): 420–428.
- Adams, C.P. and Brantner, V.V. (2010): Spending on new drug development, *Health Econ* 19(2): 130–141.
- Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R. (2003): *Multivariate Analysemethoden: Eine Anwendungsorientierte Einführung*, 10. Auflage, Springer Verlag Berlin Heidelberg New York, ISBN 3540004912.
- Baeriswyl, V. and Christofori, G. (2009): The angiogenic switch in carcinogenesis, *Semin Cancer Biol* 19(5): 329–337.
- Baggiolini, M. (2001): Chemokines in pathology and medicine, *J Intern Med* 250(2): 91–104.
- Baggiolini, M.; Dewald, B.; Moser, B. (1997): Human chemokines: an update. *Annu Rev Immunol* 15: 675–705.
- Baneyx, F. (1999): Recombinant protein expression in *Escherichia coli* *Curr Opin Biotechnol* 10(5): 411–421.
- Baneyx, F. and Georgiou, G. (1990): In vivo degradation of secreted fusion proteins by the *Escherichia coli* outer membrane protease OmpT, *J Bacteriol* 172(1): 491–494.
- Baneyx, F. and Mujacic, M. (2004): Recombinant protein folding and misfolding in *Escherichia coli*, *Nat Biotechnol* 22(11): 1399–1408.
- Barondeau D.P.; Putnam, C.D.; Kassmann, C.J.; Tainer, J.A.; Getzoff, E.D. (2003): Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures, *Proc Natl Acad Sci* 100(21): 12111–12116.
- Becker, T.; Hitzmann, B.; Muffler, K.; Pörtner, R.; Reardon, K.F.; Stahl, F.; Ulber, R. (2007): Future aspects of bioprocess monitoring, *Adv Biochem Eng Biotechnol* 105: 249–293.
- Becker, T. and Krause, D. (2010): Softsensorysysteme – Mathematik als Bindeglied zum Prozessgeschehen, *Chemie Ingenieur Technik*, 82: 429–440.
- Bezerra, M.A.; Santelli, R.E.; Oliveira, E.P.; Villar, L.S.; Escalera, L.A. (2008): Response surface methodology (RSM) as a tool for optimization in analytical chemistry, *Talanta* 76(5): 965–977.
- Binns, K.L.; Taylor, P.P.; Sicheri, F.; Pawson, T.; Holland, S.J. (2000): Phosphorylation of tyrosine residues in the kinase domain and juxtamembrane region regulates the biological and catalytic activities of Eph receptors, *Mol Cell Biol* 20(13): 4791–4805.

- Boerlage, S. F. E.; Kennedy, M. D.; Bonne, P. A. C.; Galjaard, G.; Schippers, J. C. (1997): Prediction of flux decline in membrane systems due to particulate fouling, *Desalination* 113: 231–233.
- Booth, V.; Clark-Lewis, I.; Sykes, B.D. (2004): NMR structure of CXCR3 binding chemokine CXCL11 (ITAC), *Protein Sci* 13(8): 2022–2028.
- Bowden, G.A.; Paredes, A.M.; Georgiou, G. (1991): Structure and morphology of protein inclusion bodies in *Escherichia coli*, *Biotechnology* 9(8): 725–730.
- Box, G.E.P. and Wilson, K.B. (1951): On the experimental attainment of optimum conditions, *J Roy Statist Soc Ser B Metho* 13: 1–45.
- Buckholz, R.G. and Gleeson, M.A. (1991): Yeast systems for the commercial production of heterologous proteins, *Nat Biotechnol* 9(11): 1067–1072.
- Carrió, M.M.; Corchero, J.L.; Villaverde, A. (1998): Dynamics of in vivo protein aggregation: building inclusion bodies in recombinant bacteria, *FEMS Microbiol Lett* 169(1): 9–15.
- Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W.W.; Prasher, D.C. (1994): Green fluorescent protein as a marker for gene expression, *Science* 263(5148): 802–805.
- Chattaway, T.; Montague, G.A.; Morris, A.J. (2008): Fermentation Monitoring and Control, in *Biotechnology: Bioprocessing, Volume 3, Second Edition* (eds Rehm, H.-J. and Reed, G.), Wiley-VCH Verlag GmbH, Weinheim, Germany.
- Chéry, A. (1997): Software sensors in bioprocess engineering, *J Biotechnol* 52 (3): 193–199.
- Chopra, I. and Roberts, M. (2001): Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance, *Microbiol Mol Biol Rev* 65: 232–260.
- Clark-Lewis, I.; Mattioli, I.; Gong, J.H.; Loetscher, P. (2003): Structure-function relationship between the human chemokine receptor CXCR3 and its ligands, *J Biol Chem* 278(1): 289–295.
- Cleff, T. (2008): *Deskriptive Statistik und moderne Datenanalyse: Eine computergestützte Einführung mit Excel, PASW (SPSS) und STATA*, 1. Auflage, Gabler Verlag Wiesbaden, ISBN 3834902020.
- Cody, C.W.; Prasher, D.C.; Westler, W.M.; Prendergast, F.G.; Ward, W.W. (1993): Chemical structure of the hexapeptide chromophore of the *Aequorea* green fluorescent protein, *Biochemistry* 32: 1212–1218.

- Cole, K.E.; Strick, C.A.; Paradis, T.J.; Ogborne, K.T.; Loetscher, M.; Gladue, R.P.; Lin, W.; Boyd, J.G.; Moser, B.; Wood, D.E.; Sahagan, B.G.; Neote, K. (1998): Interferon-inducible T Cell Alpha Chemoattractant (I-TAC): A Novel Non-ELR CXC Chemokine with Potent Activity on Activated T Cells through Selective High Affinity Binding to CXCR3, *J Exp Med* 187(12): 2009–2021.
- Contiero, J.; Beatty, C.; Kumari, S.; DeSanti, C.L.; Strohl, W.R.; Wolfe, A. (2000): Effects of mutations in acetate metabolism on high-cell-density growth of *Escherichia coli*, *J Ind Microbiol Biotechnol* 24(6): 421–430.
- Cornelissen, G.; Bertelsen, H.-P.; Hahn, B.; Schultz, M.; Scheffler, U.; Werner, E.; Leptien, H.; Krüß, S.; Jansen, A.-K.; Gliem, T.; Hielscher, M.; Wilhelm, B.-U.; Sowa, E.; Radeke, H.-H.; Luttmann, R. (2003): Herstellung rekombinanter Proteine mit *Pichia pastoris* in integrierter Prozessführung, *Chemie Ingenieur Technik* 75: 281–290.
- Cornelissen, G. (2004): Integrierte Bioprozessentwicklung zur Herstellung pharmakologisch wirksamer Proteine mit *Pichia pastoris*, VDI-Verlag, Reihe 17, Nr. 249.
- Couderc, R. and Baratti, J. (1980): Oxidation of methanol by the yeast *Pichia pastoris*. Purification and properties of alcohol oxidase, *Agric Biol Chem* 44: 2279–2289.
- Crabtree, H.G. (1929): Observations on the carbohydrate metabolism of tumours, *Biochem J* 23(3): 536–545.
- Cregg, J.M.; Barringer, K.J.; Hessler, A.Y.; Madden, K.R. (1985): *Pichia pastoris* as a host system for transformations, *Mol Cell Biol* 5(12): 3376–3385.
- Cregg, J.M.; Cereghino, J.L.; Shi, J.; Higgins, D.R. (2000): Recombinant protein expression in *Pichia pastoris*, *Mol Biotechnol* 16(1): 23–52.
- Cregg, J.M.; Vedvick, T.S.; Raschke, W.C. (1993): Recent Advances in the Expression of Foreign Genes in *Pichia pastoris*, *Biotechnol* 11(8): 905–910.
- Daly, R. and Hearn, M.T. (2005): Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production, *J Mol Recognit* 18(2):119–138.
- Danzer, K.; Hobert, H.; Fischbacher, C.; Jagemann, K.U. (2001): *Chemometrik: Grundlagen und Anwendungen*, Springer Verlag Berlin Heidelberg New York, ISBN 3-540-41291-3.
- Davanloo, P.; Rosenberg, A.H.; Dunn, J.J.; Studier, F.W. (1984): Cloning and expression of the gene for bacteriophage T7 RNA polymerase, *Proc Natl Acad Sci* 81(7): 2035–2039.
- De Groot, N.S. and Ventura, S. (2006): Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates, *J Biotechnol* 125(1): 110–113.
- Del Castillo, E.; Montgomery, D.L.; McCorville, D.R. (1996): Modified Desirability Functions for Multiple Response Optimization, *Journal of Quality Technology* 28(3): 337–345.

- De Mey, M.; De Maeseneire, S.; Soetaert, W.; Vandamme, E. (2007): Minimizing acetate formation in *E. coli* fermentations, *J Ind Microbiol Biotechnol* 34(11): 689–700.
- Derringer, G. and Suich, R. (1980): Simultaneous Optimization of Several Response Variables, *Journal of Quality Technology* 12(4): 214–219.
- Dickson, M. and Gagnon, J.P. (2004): Key factors in the rising cost of new drug discovery and development, *Nat Rev Drug Discov* 3(5): 417–429.
- DiMasi, J.A.; Hansen, R.W.; Grabowski, H.G. (2003): The price of innovation: new estimates of drug development costs, *J Health Econ* 22(2): 151–185.
- Doelle, H.W.; Ewings, K.N.; Hollywood, N.W. (1982): Regulation of glucose metabolism in bacterial systems, *Adv Biochem Eng* 23: 1–35.
- Donovan, R.S.; Robinson, C.W.; Glick, B.R. (1996): Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter, *J Ind Microbiol* 16(3): 145–154.
- Dubendorff, J.W. and Studier, F.W. (1991): Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor, *J Mol Biol* 219: 45–59.
- Dünnebier, G. and Tups, H. (2007): FDA PAT Initiative – Eine Anwendersicht zu technischen Möglichkeiten und aktueller industrieller Umsetzung, *CIT* 79(12): 2019–2028.
- Duller, C. (2008): Einführung in die nichtparametrische Statistik mit SAS und R: Ein anwendungsorientiertes Lehr- und Arbeitsbuch, Physica-Verlag Heidelberg, ISBN 978-3-7908-2059-1.
- Eicke, M. (2007): Automation and Optimization of Integrated Bioprocesses for the Production of Active Pharmaceutical Ingredient with *Pichia pastoris*, Masterarbeit an der HAW Hamburg, unveröffentlicht.
- Ellert, A.; Thiesing, D.; Tatge, F.; Borchert, S.; Ganzlin, M.; Luttmann, R. (2010): Optimisation of Fluorescent Fusion Protein Expression in *Escherichia coli* via Design of Experiments, Vortrag und Proceedings der CAB 2010 – 11th International Symposium on Computer Applications in Biotechnology, Leuven, Belgien; Banga, J.R.; Bogaerts, P.; van Impe, J.; Dochain, D.; Smets, I. (Eds.), doi: 10.3182/20100707-3-BE-2012.0068.
- Ellinger, T.; Beuermann, G.; Leisten, R.: Operations Research: Eine Einführung, Springer Verlag Berlin Heidelberg New York, ISBN 3-540-00477-7.
- El-Mansi, E.M. and Holms, W.H. (1989): Control of carbon flux to acetate excretion during growth of *Escherichia coli* in batch and continuous cultures. *J Gen Microbiol* 135(11): 2875–2883.
- Elsner, J.; Petering, H.; Höchstetter, R.; Kimmig, D.; Wells, T.N.; Kapp, A.; Proudfoot, A.E. (1997): The CC chemokine antagonist Met-RANTES inhibits eosinophil effector functions through the chemokine receptors CCR1 and CCR3, *Eur J Immunol* 27(11): 2892–2898.

- Endres, C.; Haake, C.; Landgrebe, D.; Beutel, S.; Stahl, F.; Hitzmann, B.; Scheper, T.; Friehs, K. (2009): Moderne Bioprozessanalytik – eine kurze Übersicht, *BIOspektrum* 06.09 15. Jahrg.: 662–664.
- Eph Nomenclature Committee (1997): Unified nomenclature for Eph family receptors and their ligands the ephrins, *Cell* 90: 403–404.
- Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wikström, C.; Wold, S. (2008): *Design of Experiments: Principles and Applications*, 3rd Edition, MKS Umetrics AB, ISBN 91-973730-4-4.
- Farza, M. and Chéry, A. (1994): BIOESTIM: software for automatic design of estimators in bioprocess engineering, *Comput Appl Biosci* 10(5): 477–488.
- Ferrer, M.; Chernikova, T.N.; Yakimov, M.M.; Golyshin, P.N.; Timmis, K.N. (2003): Chaperonins govern growth of *Escherichia coli* at low temperatures, *Nat Biotechnol* 21(11): 1266–1267.
- FDA - Food and Drug Administration (2004): *Guidance for Industry. PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*.
- Fiechter, A.; Meiners, M.; Moser, A. (1994): *Biologische Regulation und Prozessführung*, in *Handbuch der Biotechnologie*, 4. Auflage, (eds Präve, P.; Faust, U.; Sittig, W.; Sukatsch, D.A.), R. Oldenbourg Verlag GmbH, München, Germany.
- Fiedler, E. (1993): *Chromatographische Methoden zur Protein-Reinigung Teil 2: Ionenaustausch-Chromatographie*, *BIO forum* 16: 218–222.
- Fisher, R.A. (1935): *The Design of Experiments*, ISBN 0-02-844690-9.
- Folkman, J. (2006): Angiogenesis, *Annu Rev Med* 57: 1–18.
- Folkman, J. (2007): Angiogenesis: an organizing principle for drug discovery?, *Nat Rev Drug Discov* 6(4): 273–286.
- Friehs, K.; Hitzmann, B.; Scheper, T. (2006): *Prozessanalytik*, in *Angewandte Mikrobiologie* (ed Antranikian, G.), Springer-Verlag Berlin Heidelberg.
- García-Fruitós, E.; Martínez-Alonso, M.; González-Montalbán, N.; Valli, M.; Mattanovich, D.; Villaverde, A. (2007): Divergent genetic control of protein solubility and conformational quality in *Escherichia coli*, *J Mol Biol* 374(1): 195–205.
- Gaspar, H. (2000): *Handbuch der industriellen Fest/Flüssig-Filtration*, WILEY-VCH Verlag, ISBN 3-527-29796-0.
- Gellissen, G. (2000): Heterologous protein production in methylotrophic yeasts, *Appl Microbiol Biotechnol* 54(6): 741–750.
- Gleeson, M.A.; White, C.E.; Meininger, D.P.; Komives, E.A. (1998): Generation of protease-deficient strains and their use in heterologous protein expression, *Methods Mol Biol* 103: 81–94.
- Glick, B.R. (1995): Metabolic load and heterologous gene expression, *Biotechnol Adv* 13(2): 247–261.

- Gnoth, S.; Jenzsch, M.; Simutis, R.; Lübbert, A. (2008): Control of cultivation processes for recombinant protein production: a review, *Bioprocess Biosyst Eng* 31:21–39.
- Goldberg, A. L.; Moerschell, R. P.; Chung, C. H.; Maurizi, M. R. (1994): ATP-dependent protease La (lon) from *Escherichia coli*, *Methods Enzymol* 244, 350–375.
- Gong, J.H. ; Ratkay, L.G.; Waterfield, J.D.; Clark-Lewis, I. (1997): An antagonist of monocyte chemoattractant protein 1 (MCP-1) inhibits arthritis in the MRL-lpr mouse model, *J Exp Med* 186(1): 131–137.
- González-Montalbán, N.; Natalello, A.; García-Fruitós, E.; Villaverde, A.; Doglia, S.M. (2008): *In situ* protein folding and activation in bacterial inclusion bodies, *Biotechnol Bioeng* 100(4): 797–802.
- Grabowski, H. (2008): Follow-on biologics: data exclusivity and the balance between innovation and competition, *Nat Rev Drug Discov* 7(6): 479–488.
- Gregory, M.E. and Turner, C. (1993): Open-loop control of specific growth rate in fed-batch cultures of recombinant *E. coli*, *Biotechnol Tech* 7(12): 889–894.
- Grodberg, J. and Dunn, J.J. (1988): ompT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification, *J Bacteriol* 170(3): 1245–1253.
- Grubbs, F.E. (1969): Procedures for detecting outlying observations in samples, *Technometrics* 11: 1–21.
- Hahn, B. (2001): Optimierung einer FPLC-Anlage zur Aufreinigung rekombinanter Chemokine aus Integrierter Prozessführung mit *Pichia pastoris*, Diplomarbeit an der HAW Hamburg, unveröffentlicht.
- Han, K.; Lim, H.C.; Hong, J. (1992): Acetic acid formation in *Escherichia coli* fermentation, *Biotechnol Bioeng* 39: 663–671.
- Hanahan, D. and Folkman, J. (1996): Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis, *Cell* 86(3): 353–364.
- Hantelmann, K.; Kollecker, M.; Hüll, D.; Hitzmann, B. Scheper, T. (2006): Two-dimensional fluorescence spectroscopy: a novel approach for controlling fed-batch cultivations, *J Biotechnol* 121(3): 410–417.
- Harrington, E.C. (1965): The Desirability Function. *Industrial Quality Control* 21(10): 494–498.
- Hartung, J. (2005): Statistik: Lehr- und Handbuch der angewandten Statistik, 14. Auflage, Oldenbourg Wissenschaftsverlag München, ISBN 3-486-57890-1.
- Heddle, C. and Mazaleyrat, S.L. (2007): Development of a screening platform for directed evolution using the reef coral fluorescent protein ZsGreen as a solubility reporter, *Protein Eng Des Sel* 20(7): 327–337.
- Herwig, C. (2010): Prozess Analytische Technologie in der Biotechnologie, *Chemie Ingenieur Technik* 82(4): 405–414.
- Herzenberg, L.A. (1959): Studies on the induction of beta-galactosidase in a cryptic strain of *Escherichia coli*, *Biochim Biophys Acta* 31(2): 525–538.

- Hillen, W. and Berens, C. (1994): Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu Rev Microbiol* 48: 345–369.
- Himanen, J.P.; Henkemeyer, M.; Nikolov, D.B. (1998): Crystal structure of the ligand-binding domain of the receptor tyrosine kinase EphB2. *Nature* 396(6710): 486–491.
- Himanen, J.P.; Rajashankar, K.R.; Lackmann, M.; Cowan, C.A.; Henkemeyer, M.; Nikolov, D.B. (2001): Crystal structure of an Eph receptor-ephrin complex, *Nature* 414(6866): 933–938.
- Himanen, J.P.; Saha, N.; Nikolov, D.B. (2007): Cell-cell signaling via Eph receptors and ephrins, *Curr Opin Cell Biol* 19(5): 534–542.
- Hirai, H.; Maru, Y.; Hagiwara, K.; Nishida, J.; Takaku, F. (1987): A novel putative tyrosine kinase receptor encoded by the eph gene, *Science* 238(4834): 1717–1720.
- Hoaglin, D.C. and Welsch, R.E. (1978): The Hat Matrix in Regression and ANOVA, *Am Stat* 32(1): 17–22.
- Höpfner, T.; Bluma, A.; Rudolph, G.; Lindner, P.; Scheper, T. (2010): A review of non-invasive optical-based image analysis systems for continuous bioprocess monitoring, *Bioprocess Biosyst Eng* 33(2): 247–256.
- Hoffmann, F. and Rinas, U. (2004): Stress induced by recombinant protein production in *Escherichia coli*, *Adv Biochem Eng Biotechnol* 89: 73–92.
- Hoffmann, F.; van den Heuvel, J.; Zidek, N.; Rinas, U. (2004): Minimizing inclusion body formation during recombinant protein production in *Escherichia coli* at bench and pilot plant scale, *Enzyme Microb Technol* 34 (3-4): 235–241.
- Holms, H. (1996): Flux analysis and control of the central metabolic pathways in *Escherichia coli*, *FEMS Microbiol Rev* 19(2): 85–116.
- Hubbard, S.R. and Till, J.H. (2000): Protein tyrosine kinase structure and function, *Annu Rev Biochem* 69: 373–398.
- Hunke, S. and Betton, J.M. (2003): Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli*, *Mol Microbiol* 50(5): 1579–1589.
- ICH - International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (2009): *Pharmaceutical Development Q8(R2)*.
- Ingraham, J. (1987): Effect of Temperature, pH, Water Activity, and Pressure on Growth, in Neidhardt, F.C. (Ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, American Society for Microbiology, Washington, D.C. (2): 1543–1554.
- Jacob, F.; Monod, J. (1961): Genetic regulatory mechanisms in the synthesis of proteins, *J Mol Biol* 3: 318–356.
- Jahic, M.; Veide, A.; Charoenrat, T.; Teeri, T.; Enfors, S.O. (2006): Process Technology for Production and Recovery of Heterologous Proteins with *Pichia pastoris*, *Biotechnol Prog* 22(6): 1465–1473.

- Jahn, D. and Jahn, M. (2006): Rekombinante Proteinproduktion in Bakterien und Pilzen, in: G. Antranikian (Hrsg.): Angewandte Mikrobiologie, Springer-Verlag Berlin Heidelberg New York, ISBN 3-540-24083-7.
- James, S.; Legge, R.; Budman, H. (2002): Comparative study of black-box and hybrid estimation methods in fed-batch fermentation, *J Process Control* 12: 113–121.
- Jensen, P.R.; Westerhoff, H.V.; Michelsen, O. (1993): The use of lac-type promoters in control analysis, *Eur J Biochem* 211(1-2): 181–191.
- Jones, P.G.; VanBogelen, R.A.; Neidhardt, F.C. (1987): Induction of proteins in response to low temperature in *Escherichia coli*, *J Bacteriol* 169(5): 2092–2095.
- Kaiser, C. (2009): Methoden zur Prozessüberwachung und Modellierung von Herstellungsprozessen rekombinanter Proteine mit *Escherichia coli*, Hannover, Universität Hannover, Dissertation.
- Kaiser, C.; Pototzki, T.; Ellert, A.; Luttmann, R. (2008): Applications of PAT - Process Analytical Technology in Recombinant Protein Processes with *Escherichia coli*, *Eng Life Sci* 8(2): 132–138.
- Kessler, W. (2007): Multivariate Datenanalyse für die Pharma-, Bio- und Prozessanalytik, 1. Auflage, Wiley-VCH Verlag, ISBN 978-3-527-31262-7.
- Khlebnikov, A. and Keasling, J.D. (2002): Effect of lacY expression on homogeneity of induction from the P(tac) and P(trc) promoters by natural and synthetic inducers, *Biotechnol Prog* 18(3): 672–674.
- Kiefer, J. and Wolfowitz, J. (1959): Optimum Designs in Regression Problems, *Ann Math Statist* 30(2): 271–294.
- Kiefhaber, T.; Rudolph, R.; Kohler, H.H.; Buchner, J. (1991): Protein aggregation *in vitro* and *in vivo*: a quantitative model of the kinetic competition between folding and aggregation, *Biotechnology* 9(9): 825–829.
- Kiviharju, K.; Salonen, K.; Moilanen, U.; Eerikäinen, T. (2008): Biomass measurement online: the performance of in situ measurements and software sensors, *J Ind Microbiol Biotechnol* 35(7): 657–665.
- Kleppmann, W. (2009): Taschenbuch Versuchsplanung. Produkte und Prozesse optimieren. 5. überarbeitete Auflage, Carl Hanser Verlag München Wien, ISBN 978-3-446-41595-9.
- Klockowa, C.; Hülla, D.; Hitzmann, B. (2008): Model based substrate set point control of yeast cultivation processes based on FIA measurements, *Analytica Chimica Acta* 623(1): 30–37.
- Knippers, R. (2006): Molekulare Genetik, 9. Auflage, Georg Thieme Verlag Stuttgart New York, ISBN 3-13-477009-1.
- Konstantinov, K.; Kishimoto, M.; Seki, T.; Yoshida, T. (1990): A balanced DO-stat and its application to the control of acetic acid excretion by recombinant *Escherichia coli*, *Biotechnol Bioeng* 36(7): 750–758.

- Kosinski, M.J.; Rinas, U.; Bailey, J.E. (1992): Isopropyl- β -d-thiogalactopyranoside influences the metabolism of *Escherichia coli*, *Appl Microbiol Biotechnol* 36(6): 782–784.
- Kraume, M. (2003): Transportvorgänge in der Verfahrenstechnik: Grundlagen und apparative Umsetzungen, Springer Verlag Berlin Heidelberg New York, ISBN 3-540-40105-9.
- Kullander, K. and Klein, R. (2002): Mechanisms and functions of Eph and ephrin signalling, *Nat Rev Mol Cell Biol* 3(7): 475–486.
- Lange, J. (2005): Zehn Jahre OPC: Von Data Access zu Unified Architecture, *SPS Magazin* 8: 64–67.
- Lee, S.Y. (1996): High cell-density culture of *Escherichia coli*, *Trends Biotechnol* 14(3): 98–105.
- Lemaux, P.G.; Herendeen, S.L.; Bloch, P.L.; Neidhardt, F.C. (1978): Transient rates of synthesis of individual polypeptides in *E. coli* following temperature shifts, *Cell* 13(3): 427–434.
- Levisauskas, D.; Simutis, R.; Borvitz, D.; Lübbert, A. (1996): Automatic control of the specific growth rate in fed-batch cultivation processes based on an exhaust gas analysis. *Bioprocess Biosyst Eng* 15: 145–150.
- Lewis, M. (2005): The lac repressor, *C R Biol* 328(6): 521–548.
- Li, P.; Anumanthan, A.; Gao, X.G.; Ilangovan, K.; Suzara, V.V.; Düzgüneş, N.; Renugopalakrishnan, V. (2007): Expression of recombinant proteins in *Pichia pastoris*, *Appl Biochem Biotechnol* 142(2): 105–124.
- Li, F.; Hashimura, Y.; Pendleton, R.; Harms, J.; Collins, E.; Lee, B. (2006): A systematic approach for scale-down model development and characterization of commercial cell culture processes, *Biotechnol Prog* 22(3): 696–703.
- Lindemann, C. (1998): Anwendungen der 2D-Fluoreszenzspektroskopie zur Bioprozessbeobachtung, VDI-Verlag, Reihe 17, Nr. 171.
- Luli, G.W. and Strohl, W.R. (1990): Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations, *Appl Environ Microbiol* 56(4): 1004–1011.
- Luttmann, R.; Bitzer, G.; Hartkopf, J. (1994): Development of control strategies for high cell density cultivations, *Math Comp Simul* 37: 153–164.
- Luttmann, R.; Stilkenbäumer, A.; Roß, A. (1990): Regelung intrazellulärer Aktivitätsgrößen in Bioreaktoren - Theorie, Konzept und technische Realisierung, *BioEngineering* 6: 12–22.
- Maiwald, M. (2010): Prozessanalytik als Instrument des Informationsmanagements in der Chemischen und Pharmazeutischen Industrie, *CIT* 82(4): 383–390.
- Martínez-Alonso, M.; González-Montalbán, N.; García-Fruitós, E.; Villaverde, A. (2009): Learning about protein solubility from bacterial inclusion bodies, *Microb Cell Fact* 8:4.
- Mandenius, C.F. and Brundin, A. (2008): Bioprocess Optimization Using Design-of-Experiments Methodology, *Biotechnol Prog* 24: 1191–1203.

- Matsui, T.; Akahoshi, T.; Namai, R.; Hashimoto, A.; Kurihara, Y.; Rana, M.; Nishimura, A.; Endo, H.; Kitasato, H.; Kawai, S.; Takagishi, K.; Kondo, H. (2001): Selective recruitment of CCR6-expressing cells by increased production of MIP-3 alpha in rheumatoid arthritis, *Clin Exp Immunol* 125(1): 155–161.
- Matz, M.V.; Fradkov, A.F.; Labas, Y.A.; Savitsky, A.P.; Zaraisky, A.G.; Markelov, M.L.; Lukyanov, S.A. (1999): Fluorescent proteins from nonbioluminescent Anthozoa species, *Nature Biotech* 17: 969–973.
- Mendenhall, W.; Beaver, R.J.; Beaver, B.M. (2009): Introduction to Probability and Statistics, 13th Edition, Brooks/Cole USA, ISBN 0495389536.
- Meyer, F.; Gehmlich, I.; Guthke, R.; Górak, A.; Knorre, W. A. (1998): Analysis and Simulation of Complex Interactions During Dynamic Microfiltration of *Escherichia coli* Suspensions, *Biotechnol Bioeng* 59: 189–202.
- Meyer, M.; Hensbergen, P.J.; van der Raaij-Helmer, E.M.; Brandacher, G.; Margreiter, R.; Heufler, C.; Koch, F.; Narumi, S.; Werner, E.R.; Colvin, R.; Luster, A.D.; Tensen, C.P.; Werner-Felmayer, G. (2001): Cross reactivity of three T cell attracting murine chemokines stimulating the CXC chemokine receptor CXCR3 and their induction in cultured cells and during allograft rejection, *Eur J Immunol* 31(8): 2521-2527.
- Michl, D.; Heinemann, A. (2010): Medizinische Biotechnologie in Deutschland 2010, Studie der Boston Consulting Group (BCG) für die Interessengruppe Biotechnologie im Verband Forschender Arzneimittelhersteller e.V. (vfa bio).
- Mogk, A.; Mayer, M.P.; Deuerling, E. (2002): Mechanisms of protein folding: molecular chaperones and their application in biotechnology, *Chembiochem* 3(9): 807–814.
- Montgomery, D.C. (2009): Design and Analysis of Experiments, 7th Edition, John Wiley & Sons Canada, ISBN 0470409193.
- Moser, A. (1994): Operations-Strategien für Bioprozesse, *Handbuch der Biotechnologie, Faust/Praeve/Sittig/Sukatsch (Hrsg.), 4. Auflage*, 405-441.
- Moserle, L.; Amadori, A.; Indraccolo, S. (2009): The angiogenic switch: implications in the regulation of tumor dormancy, *Curr Mol Med* 9(8): 935–941.
- Murphy, P.M.; Baggiolini, M.; Charo, I.F.; Hébert, C.A.; Horuk, R.; Matsushima, K.; Miller, L.H.; Oppenheim, J.J.; Power, C.A. (2000): International union of pharmacology. XXII. Nomenclature for chemokine receptors, *Pharmacol Rev* 52(1): 145–176.
- Nelder, J.A. and Mead, R. (1965): A Simplex Method for Function Minimization, *Computer Journal* 7: 308–313.
- Noel, J.T.; Pilyugin, S.S.; Narang, A. (2009): The diffusive influx and carrier efflux have a strong effect on the bistability of the lac operon in *Escherichia coli*, *J Theor Biol* 256(1): 14–28.
- Oesterle, G. (1994): Messung von Konzentration bzw. Zusammensetzung, Kapitel 10.2.4 „Flammenionisationsmessung“, *Handbuch der industriellen Meßtechnik, Profos/Pfeifer (Hrsg.), 6. Auflage*, 946–951.

- Offermanns, S. and Rosenthal, W. (2004): Encyclopedic reference of molecular pharmacology, ISBN 3-540-42843-7, 1026–1027.
- Olsson, L. and Nielsen, J. (1997): On-line and *in situ* monitoring of biomass in submerged cultivations, Trends Biotechnol 15: 517–522.
- Ormö, M.; Cubitt, A.B.; Kallio, K.; Gross, L.A.; Tsien, R.Y.; Remington, S.J. (1996): Crystal structure of the *Aequorea victoria* green fluorescent protein, Science 273(5280): 1392–1395.
- Panzer, U.; Schneider, A.; Wilken, J.; Thompson, D.A.; Kent, S.B.; Stahl, R.A. (1999): The chemokine receptor antagonist AOP-RANTES reduces monocyte infiltration in experimental glomerulonephritis, Kidney Int 56(6): 2107–2115.
- Pasquale, E.B. (2008): Eph-ephrin bidirectional signaling in physiology and disease, Cell 133(1): 38–52.
- Pearson, K. (1909): Determination of the coefficient of correlation, Science 30(757): 23–25.
- Peuker, T. (2006): Modellgestützte Verfahren zur Beobachtung und Steuerung bioverfahrenstechnischer Prozesse, VDI-Verlag, Reihe 17, Nr. 261.
- Peuker, T.; Riedel, M.; Kaiser, C.; Ellert, A.; Lenz, K.; Elsholz, O.; Luttmann, R. (2004): At-line Determination of Glucose, Ammonia, and Acetate in High Cell Density Cultivations of *Escherichia coli*, Engineering in Life Sciences 4(2): 138–143.
- Piñeiro, G.; Perelman, S.; Guerschman, J.P.; Paruelo, J.M. (2008): How to evaluate models: Observed vs. predicted or predicted vs. observed?, Ecol Model 216(3-4): 316–322.
- Pioli, D.; Hockney, R.C.; Kara, B.V.; Bundell, K.R. (1998): T7 promoter-based expression system, United States Patent Application Publication US6537779 B1.
- Pletneva, N.; Pletnev, V.; Tikhonova, T.; Pakhomov, A.A.; Popov, V.; Martynov, V.I.; Wlodawer, A.; Dauter, Z.; Pletnev, S. (2007): Refined crystal structures of red and green fluorescent proteins from the button polyp *Zoanthus*, Acta Crystallogr D Biol Crystallogr 63(10): 1082–1093.
- Polier, S. and Bracher, A. (2009): Proteinfaltung in der Zelle – eine komplexe Angelegenheit, BIOSpektrum 3: 255–257.
- Rautenbach, R. (1997): Membranverfahren, Grundlagen der Modul- und Anlagenauslegung, Springer Verlag Berlin Heidelberg New York, ISBN 978-3-540-61573-6.
- Rehbock, C.; Beutel, S.; Brückerhoff, T.; Hitzmann, B.; Riechers, D.; Rudolph, G.; Stahl, F.; Scheper, T.; Friehs, K. (2008): Bioprozessanalytik, Chemie Ingenieur Technik 80(3): 267–286.
- Reiling, H.E.; Laurila, H.; Fiechter, A. (1985): Mass culture of *Escherichia coli*: Medium development for low and high density cultivation of *Escherichia coli* B/r in minimal and complex media, J Biotechnol 2(3-4): 191–206.
- Riesenberg, D. and Guthke, R. (1999): High-cell-density cultivation of microorganisms, Appl Microbiol Biotechnol 51(4): 422–430.

- Riesenberg, D.; Menzel, K.; Schulz, V.; Schumann, K.; Veith, G.; Zuber, G.; Knorre, W.A. (1990): High cell density fermentation of recombinant *Escherichia coli* expressing human interferon alpha 1, *Appl Microbiol Biotechnol* 34(1): 77–82.
- Riesenberg, D.; Schulz, V.; Knorre, W.A.; Pohl, H.D.; Korz, D.; Sanders, E.A.; Ross, A.; Deckwer, W.D. (1991): High cell density cultivation of *Escherichia coli* at controlled specific growth rate, *J Biotechnol* 20(1): 17–27.
- Rinas, U. (2006): Mikrobielle Herstellung von Pharmaproteinen, Kapitel 6 in *Angewandte Mikrobiologie*, Antranikian, G. (Hrsg.), DOI: 10.1007/3-540-29456-2_6.
- Rinas, U.; Kracke-Helm, H.A.; Schügerl, K. (1989): Glucose as a substrate in recombinant strain fermentation technology, *Appl Microbiol Biotechnol* 31: 163–167.
- Rinne, H. (2008): Taschenbuch der Statistik, 4. Auflage, Verlag Harry Deutsch, ISBN 978-3-8171-1827-4.
- Ripperger, S. (1993): Berechnungsansätze zur Crossflow-Filtration, *Chemie Ingenieur Technik* 65: 533–540.
- Roderick, S.L. (2005): The lac operon galactoside acetyltransferase, *Biol* 328(6): 568–575.
- Roggenkamp, R.; Janowicz, Z.; Stanikowski, B.; Hollenberg, C.P. (1984): Biosynthesis and regulation of the peroxisomal methanol oxidase from the methylotrophic yeast *Hansenula polymorpha*, *Mol Gen Genet* 194(3): 489–493.
- Rosenberg, A.H.; Lade, B.N.; Chui, D.S.; Lin, S.W.; Dunn, J.J.; Studier, F.W. (1987): Vectors for selective expression of cloned DNAs by T7 RNA polymerase, *Gene* 56(1): 125–135.
- Rothen, S.A.; Sauer, M.; Sonnleitner, B.; Witholt, B. (1998): Growth characteristics of *Escherichia coli* HB101[pGEc47] on defined medium, *Biotechnol Bioeng* 58(1): 92–100.
- Rubant, S. (2005): Molekulare Wirkmechanismen rekombinant hergestellter Chemokinrezeptor-Antagonisten auf entzündungsrelevante Immunzellen, Dissertation Technische Universität Darmstadt.
- Sachs, L. and Hedderich, J. (2009): *Angewandte Statistik. Methodensammlung mit R*, 13. aktualisierte und erweiterte Auflage, Springer Verlag Berlin Heidelberg New York, ISBN 978-3-5403-2160-6.
- Sahdev, S.; Khattar, S.K.; Saini, K.S. (2008): Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol Cell Biochem* 307(1-2): 249–264.
- Scheffler, E. (1997): *Statistische Versuchsplanung und -auswertung. Eine Einführung für Praktiker*, 3. Auflage, Deutscher Verlag für Grundstoffindustrie, ISBN 3-342-00366-9.
- Schein, C.H. (1989): Production of Soluble Recombinant Proteins in Bacteria, *Nat Biotechnol* 7: 1141–1149.
- Schmidt, F.R. (2004): Recombinant expression systems in the pharmaceutical industry, *Appl Microbiol Biotechnol* 65(4): 363–372.
- Schmidt, F.R. (2005): Optimization and scale up of industrial fermentation processes, *Appl Microbiol Biotechnol* 68(4): 425–435.

- Schmidt-Bader, T. (2010): PAT und QbD im regulatorischen Umfeld der Pharmazeutischen Industrie, CIT Special Issue: Prozessanalytik 82(4): 415–428.
- Schügerl, K. (2000): Integrated processing of biotechnology products, Biotechnol Adv 18(7): 581–599.
- Schügerl K. (2001): Progress in monitoring, modeling and control of bioprocesses during the last 20 years, J Biotechnol 85(2): 149–173.
- Schultz, M. (2001): Automatisierung eines neuartigen Bioreaktorsystems zur Integrierten Prozessführung, Diplomarbeit an der HAW Hamburg, unveröffentlicht.
- Scorer, C.A.; Clare, J.J.; McCombie, W.R.; Romanos, M.A.; Sreekrishna, K. (1994): Rapid selection using G418 of high copy number transformants of *Pichia pastoris* for high-level foreign gene expression, Biotechnology 12(2): 181–184.
- Scott, C. and McLeod, L.D. (2010): The Time Has Come for Automation in Bioprocessing, BioProcess International 8(2): 16–25.
- Seo, J.H. and Bailey, J.E. (1986): Continuous cultivation of recombinant *Escherichia coli*: Existence of an optimum dilution rate for maximum plasmid and gene product concentration, Biotechnol Bioeng 28(10): 1590–1594.
- Shapiro, S.S. and Wilk, M.B. (1965): An analysis of variance test for normality (complete samples), Biometrika 52(3 and 4): 591–611.
- Shiloach, J. and Fass, R. (2005): Growing *E.coli* to high cell density - A historical perspective on method development, Biotechnol Adv 23(5): 345–357.
- Shimomura, O. (1979): Structure of the chromophore of *Aequorea* green fluorescent protein, FEBS Lett 104: 220–222.
- Shimomura, O.; Johnson, F.H.; Saiga, Y. (1962): Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan *Aequorea*, J Cell Comp Physiol 59: 223–239.
- Sigler, A.; Schubert, P.; Hillen, W.; Niederweis, M. (2000): Permeation of tetracyclines through membranes of liposomes and *Escherichia coli*, Eur J Biochem 267 (2): 527–534.
- Simons, A; Tils, D.; von Wilcken-Bergmann, B.; Müller-Hill, B. (1984): Possible ideal lac operator: *Escherichia coli* lac operator-like sequences from eukaryotic genomes lack the central G X C pair, Proc Natl Acad Sci 81(6): 1624–1628.
- Simutis,R.; Oliveira, R.; Manikowski, M.; Feyer de Azevedo, S.; Lübbert, A. (1997): How to increase the performance of models for process optimization and control, J Biotechnol 59: 73–89.
- Sørensen, H.P. and Mortensen, K.K. (2005): Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*, Microb Cell Fact 4(1):1.
- Sonnleitner, B. (2000): Instrumentation of biotechnological processes. Adv Biochem Eng Biotechnol 66: 1–64.
- Spieß, C.; Beil, A.; Ehrmann, M.A. (1999): Temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein, Cell 97(3): 339–347.

- Stärk, E. (2002): Bioprozessanalytik durch Online-Vorhersage von Bioprozessgrößen mittels 2D-Fluoreszenzspektroskopie und multivariater Auswerteverfahren, Hannover, Universität Hannover, Dissertation.
- Stark, D. and von Stockar, U. (2003): In Situ Product Removal (ISPR) in Whole Cell Biotechnology During the Last Twenty Years, *Adv Biochem Eng Biotechnol* 80: 149–175.
- Steinke, L.; Wessely, B.; Ripperger, S. (2009): Optische Extinktionsmessverfahren zur Inline-Kontrolle disperser Stoffsysteme, *Chemie Ingenieur Technik* 81(6): 735–747.
- Stellato, C.; Brummet, M.E.; Plitt, J.R.; Shahabuddin, S.; Barody, F.M.; Liu, M.C.; Ponath, P.D.; Beck, L.A. (2001): Expression of the C-C chemokine receptor CCR3 in human airway epithelial cells, *J Immunol* 166(3): 1457–1461.
- Straková, M.; Matisová, E.; Šimon, P.; Annus, J.; Mikuláš Lisý, J. (1998): Silicone membrane measuring system with SnO₂ gas sensor for on-line monitoring of volatile organic compounds in water, *Sensors and Actuators B: Chemical* 52(3): 274–282.
- Striedner, G.; Cserjan-Puschmann, M.; Pötschacher, F.; Bayer, K. (2003): Tuning the Transcription Rate of Recombinant Protein in Strong *Escherichia coli* Expression Systems through Repressor Titration, *Biotechnol Prog* 19(5): 1427–1432.
- Studier, F.W. and Moffatt, B.A. (1986): Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *J Mol Biol* 189(1): 113–130.
- Studier, F.W.; Rosenberg, A.H.; Dunn, J.J.; Dubendorff, J.W. (1990): Use of T7 RNA polymerase to direct expression of cloned genes, *Methods Enzymol* 185: 60–89.
- Summers, D.K. and Sherratt, D.J. (1984): Multimerisation of high copy number plasmids causes instability: ColE1 encodes a determinant for plasmid monomerisation and stability, *Cell* 36: 1097–1103.
- Summers, D.; Yaish, S.; Archer, J.; Sherrat, D. (1985): Multimer resolution system of ColE1 and ColK: localization of crossover site, *Mol Gen Genet* 201: 334–338.
- Surawska, H.; Ma, P.C.; Salgia, R. (2004): The role of ephrins and Eph receptors in cancer, *Cytokine Growth Factor Rev* 15(6): 419–433.
- Tatge, F. (2008): Screening von Expressionsparametern für die optimale Herstellung rekombinanter Proteine auf Basis der statistischen Versuchsplanung (SVP), Diplomarbeit an der HAW Hamburg, unveröffentlicht.
- Taylor, G.; Hoare, M.; Gray, D.R.; Marston, F.A.O. (1986): Size and Density of Protein Inclusion Bodies, *Nat Biotechnol* 4: 553–557.
- Tropp, B.E. (2008): *Molecular Biology: genes to proteins*, 3rd Edition, Jones and Bartlett Publishers, ISBN 0-7637-0916-6.
- Tschopp, J.F.; Brust, P.F.; Cregg, J.M.; Stillman, C.A.; Gingeras, T.R. (1987): Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*, *Nucleic Acids Res* 15(9): 3859–3876.
- Tsien, R.Y. (1998): The green fluorescent protein, *Annu Rev Biochem* 67: 509–544.

- Twigg, A.J. and Sherratt, D. (1980): Trans-complementable copy-number mutants of plasmid ColE1, *Nature* 283: 216–218.
- van der Geer, P.; Hunter, T.; Lindberg, R.A. (1994): Receptor protein-tyrosine kinases and their signal transduction pathways, *Annu Rev Cell Biol* 10: 251–337.
- van der Voorn, P.; Tekstra, J.; Beelen, R.H.; Tensen, C.P.; Van Der Valk, P.; De Groot, C.J. (1999): Expression of MCP-1 by reactive astrocytes in demyelinating multiple sclerosis lesions, *Am J Pathol* 154(1): 45–51.
- Veith, D. and Veith, M. (2005): Biologie fluoreszierender Proteine: Ein Regenbogen aus dem Ozean, *Biologie in unserer Zeit* 35(6): 394–404.
- Vernon, J.A.; Golec, J.H.; DiMasi J.A. (2010): Drug development costs when financial risk is measured using the Fama-French three-factor model, *Health Econ* 19(8): 1002–1005.
- Voedisch, B.; Menzel, C.; Jordan, E.; El-Ghezal, A.; Schirrmann, T.; Hust, M.; Jostock, T. (2005): Heterologe Expression von rekombinanten Proteinpharmazeutika, *Laborwelt* 6(3): 31–35.
- Wagner, L.W.; Matheson, N.H.; Heisey, R.F.; Schneider, K. (1997): Use of a silicone tubing sensor to control methanol concentration during fed batch fermentation of *Pichia pastoris*, *Biotechnology Techniques* 11: 791–795.
- Wang, L. and Song, L. (1999): Flux decline in crossflow microfiltration and ultrafiltration: experimental verification of fouling dynamics, *J Membrane Sci* 160: 41–50.
- Weiss, M. and Reiländer, H. (1999): Das *Pichia-pastoris*-Expressionssystem, in: Gassen, H. and G. Schrimpf (Hrsg.): *Gentechnische Methoden*, Spektrum Akademischer Verlag Heidelberg Berlin, ISBN 3-8274-0798-2.
- Weiss, H.M.; Haase, W.; Michel, H.; Reiländer, H. (1995): Expression of functional mouse 5-HT_{5A} serotonin receptor in the methylotrophic yeast *Pichia pastoris*: pharmacological characterization and localization, *FEBS Lett* 377(3): 451–456.
- Werner, R.G. (2004): Economic aspects of commercial manufacture of biopharmaceuticals, *J Biotechnol* 113(1-3): 171–182.
- Wilrich, P.T. and Henning, H.J. (1998): *Formeln und Tabellen der angewandten mathematischen Statistik*, 3. völlig neu bearbeitete Auflage, 2. korrigierter Nachdruck, Springer Verlag Berlin Heidelberg New York, ISBN 3-540-16901-6.
- Wybenga-Groot, L.E.; Baskin, B.; Ong, S.H.; Tong, J.; Pawson, T.; Sicheri, F. (2001): Structural basis for autoinhibition of the Ephb2 receptor tyrosine kinase by the unphosphorylated juxtamembrane region, *Cell* 106(6): 745–757.
- Xu, B.; Jahic, M.; Blomsten, G.; Enfors, S.O. (1999a): Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*, *Appl Microbiol Biotechnol* 51(5): 564–571.
- Xu, B.; Jahic, M.; Enfors, S.O. (1999b): Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*, *Biotechnol Prog* 15(1): 81–90.

Yang, F.; Moss, L.G.; Phillips, G.N., Jr. (1996): The molecular structure of green fluorescent protein, *Nat Biotechnol* 14: 1246–1251.

Yanushevich, Y.G.; Staroverov, D.B.; Savitsky, A.P.; Fradkov, A.F.; Gurskaya, N.G.; Bulina, M.E.; Lukyanov, K.A.; Lukyanov, S.A. (2002): A strategy for the generation of non-aggregating mutants of *Anthozoa* fluorescent proteins, *FEBS Lett* 511(1-3):11–14.

Yu, L.X. (2008): Pharmaceutical quality by design: product and process development, understanding, and control, *Pharm Res* 25(4): 781–791.

Zabin, I.; Kepes, A.; Monod, J. (1962): Thiogalactoside transacetylase, *J Biol Chem* 237: 253–257.

Zhang, W.; Smith, L.A.; Plantz, B.A.; Schlegel, V.L.; Meagher, M.M. (2002): Design of methanol Feed control in *Pichia pastoris* fermentations based upon a growth model, *Biotechnol Prog* 18(6): 1392–1399.

Zlotnik, A. and Yoshie, O. (2000): Chemokines: a new classification system and their role in immunity, *Immunity* 12(2): 121–127.