

Abschlussbericht zum Max-Buchner Antrag

FZK 2703: Anwendung und Auswertung von DNA-Mikroarrays zur Analyse prozesstechnisch relevanter Einflussfaktoren bei der *E.coli*-Kultivierung

Abstract:

Im Rahmen des Projekts wurden mit speziell konstruierten *E. coli* Stämmen und selbstentwickelten, sekretionsspezifischen Mikroarrays Genexpressionsanalysen durchgeführt. Die Ergebnisse dieser Analysen können zur Optimierung der sekretierten Proteinmenge genutzt werden. Zur leichteren Interpretation der Daten wurde eine spezielle Auswertesoftware mit direkter Kegg-Einbindung entwickelt und eine Datenbank aufgebaut.

Bericht:

Das Ziel dieser Forschungsarbeiten war es mittels grundlegender Genexpressionsanalysen ein besseres Bild der Zusammenhänge im Sekretionsgeschehen zu erlangen.

Für ein tiefergehendes mechanistisches Verständnis der Hauptsekretionspathways Sec und Tat ist die Kenntnis des Transkriptionstatus der Pathwaygene zu jedem Zeitpunkt der Sekretion entscheidend. Zusätzlich waren die Auswirkungen der BRP-Expression auf die Transkription der Zelle Ziel der Analysen. Aufgrund der Komplexität der involvierten Vorgänge ist eine globale Betrachtungsweise auf Transkriptomebene indiziert. Für die angestrebten Evaluierungen sollten daher DNA-Microarrayanalysen mittels *whole-genome* Arrays fungieren. Aufgrund des hohen Parallelisierungsgrades und der Flexibilität des Analysendesigns stellt die DNA-Microarray-Technologie eine ideale Methodik für die gewünschte Applikation dar. Aus den Daten der *whole-genome* Experimente sollte ein sekretionsspezifischer *low-density* Microarray entwickelt und zur routinemäßigen Applikation gebracht werden. Basierend auf diesen Erkenntnissen sollte es möglich sein, die Produktion rekombinanter Proteine durch erhöhte Sekretion zu verbessern und die Produktivität zu erhöhen. Zur übersichtlichen Darstellung der teils sehr großen Datensätze sollte eine Datenbank aufgebaut werden.

Durchgeführte Arbeiten

1. Analyse der extrazellulären sowie periplasmatischen Proteinsekretion

Für diese Arbeiten wurde ein Arabinosemetabolismus-defizienter *E. coli* K12 Stamm (Ara 1655) benutzt, um die Metabolisierung des Induktors L-Arabinose zu umgehen. Anhand dieses Stammes erfolgten transkriptionelle Analysen mit variierenden Expressionsraten des LppBRPs. Um der sehr schnellen transkriptionellen Regulation und Adaption an Stimuli in Prokaryonten Rechnung zu tragen, wurden zudem zwei feingliedrige Kinetiken der differentiellen Genexpression aufgenommen, welche eine Probennahme schon 2 und 5 Minuten nach Induktion beinhalteten. Es konnte gezeigt werden, dass die BRP-Expression substantielle regulatorische Auswirkungen auf das Transkriptom der Zelle bewirkte. Um der Aufschlüsselung der zellulären Antwort auf den Induktionsstimulus und die damit

einhergehende BRP-Expression eine größere Systematik zu verleihen, wurden die regulierten Genen mit Hilfe der Multifun-Datenbank biologischen Funktionen zugeordnet und klassifiziert. Eine Evaluierung der verschiedenen Funktionsklassen und die Aufteilung der zugeordneten, regulierten Gene in die einzelnen Subklassen wurde dazu verwendet, besonders deutlich in den Induktionsstimulus involvierte biologische Pathways und Substanzklassen zu identifizieren. Neben verschiedenen Stress-induzierbaren Genen wurden besonders verschiedene Strukturgene und Operons von Enzymen, welche Schlüsselpositionen bei der Biosynthese des LPS der äußeren Membran aufweisen, differentiell exprimiert. Die Regulation dieser Gene war aufgrund der dramatischen Auswirkungen der BRP-Expression auf die Membranintegritäten und das Peptidoglykan nicht unerwartet. Eine evidente Beobachtung erfolgte für die Oberflächen-Antigene und der allgemeinen Zellantwort auf Membranstress. In diesem Zusammenhang sind besonders das Phagenschockoperon *pspABCDE*, das bei der Sphäroblastenbildung involvierte Gen *spy* und das *rfa*-Operon zu nennen. Eine zentrale Rolle kommt dabei im Falle der LPS-Biosynthese den beiden Proteinen *LolA* und *LolB* zu, diese haben maßgeblichen Anteil an der korrekten Translokation der Lipoproteinen von der inneren zur äußeren Membran. *LolA* ist dabei das periplasmatische Bindeprotein, und *LolB* ist in der äußeren Membran lokalisiert. Durch den BRP-induzierten semispezifischen Verlust des periplasmatischen Inhalts gelangt wahrscheinlich auch *LolA* in den extrazellulären Raum und steht der Zelle nicht mehr für die Translokation der Lipoproteine der äußeren Membran zur Verfügung. Im Kontext der BRP-induzierten partiellen Degradation der äußeren Membran durch die Aktivität von *PldA* wiegt dieses doppelt schwer. Insofern stellt das codierende *lolA*-Gen ein attraktives Ziel für eine Coexpressionstrategie dar, um es der Zelle zu ermöglichen die drastischen Auswirkungen der BRP-Expression zu kompensieren. Zudem wurden viele Gene als reguliert identifiziert, welche für Proteine codieren die in die Peptidoglykanbiosynthese und der Aufrechterhaltung der Membranintegritäten z.B. unter Bedingungen hoher Osmolarität eine zentrale Rolle spielen. An dieser Stelle können die Gene *osmB*, *mrdb*, *cspD* und *imp* als Beispiel fungieren. Besonders die differentielle Expression der *mlt*-Gene und die von *imp* gibt Hinweise auf eine mechanistische Beteiligung an der BRP-induzierten Freisetzung von Colicinen und periplasmatischen Proteinen. Deren Produkte bewirken eine lokale Öffnung der Peptidoglykanstruktur (*mlt*) bzw. eine erhöhte Membranpermeabilität (*imp*). Gleiches gilt für verschiedene Gene wie z.B. *tolA* und *tolQ* deren Produkte schon bekannter Weise für die Freisetzung und Aufnahme verschiedener Colicine fungieren. Zusätzlich wurden eine Reihe von weiteren periplasmatischen Proteinen und Chaperonen reguliert vorgefunden. Besonders die Chaperone *fkpA*, *DnaK* und *GroEL/GroES* stellen zudem interessante Kandidaten für Coexpressionstrategien mit dem Ziel erhöhter Sekretionseffizienzen dar. Interessant zu verfolgen war eine Regulationskaskade zur Initiation der zellulären Stressantwort. Das Gen *lexA* codiert für den Repressor der in einer SOS-Antwort involvierten Gene und stellt in etlichen transkriptionellen Vergleichen das am stärksten negativ regulierte Gen dar. Für die Inaktivierung von *LexA* fungiert das Produkt des Strukturgenes *recA*. Dieses Gen war in den angesprochenen transkriptionellen Vergleichen stets positiv reguliert. In der Zelle findet also eine substantielle Umstellung auf die Transkription der SOS-Gene statt, indem auf Genebene die Transkription des Repressors stark vermindert wird und auf Proteinebene *RecA* für dessen autokatalytische Degradation sorgt. Die darauf folgende Expression der SOS-Gene konnten experimentell ebenfalls nachvollzogen werden, *uvrCD* und *polB* gehörten sinngemäß zu den nach Induktion positiv regulierten Genen. *UvrD* ist in der DNA-Reparatur involviert und *PolB* stellt dabei die prozessierende Polymerase dar. Die Aktivität von *RecA* bewirkt, neben der Proteolyse von *LexA*, die Degradation zweier weiterer Proteine, des *Lambda*-Repressors und des *UmuD*-Proteins. Bedingt durch die Degradation des *Lambda*-Repressors findet die Induktion verschiedener Prophagengene statt, welches experimentell ebenfalls beobachtet wird. Von Bedeutung ist darüberhinaus die positive Regulation von *soxS*, da diese ubiquitär

über nahezu alle transkriptionellen Vergleiche auftrat. SoxS fungiert als Regulator für die zelluläre Antwort auf oxydativen Stress. Das Gen *marB*, dessen Transkription von SoxS positiv reguliert wird, findet sich auch im Rahmen der evaluierten Analyse unter den positiv regulierten Genen und verifiziert das Ergebnis. In bakteriellen Biofilmen ist die Transkription von *soxS* stark erhöht, experimentell wird interessanterweise die Regulation von weiteren in die Biofilmbildung involvierten Genen beobachtet. An erster Stelle ist hier das Operon *ycdQP* zu nennen. Unter Bedingungen hoher Osmolaritäten lagert *E. coli* größere Mengen an Galaktose intrazellulär ein, dies kann experimentell anhand der Regulation des Operons *galKET* nachvollzogen werden. Diese Gene werden zudem für die Ausbildung eines Kapselpolysaccharids, des K-Antigens, benötigt. Extremophile Organismen lagern in höherem Maße sogenannte biokompatible Solute ein, dabei handelt es sich um niedermolekulare organische Osmolyte die vom Wirtsorganismus als Reaktion auf zellulären Stress produziert werden um sich vor den vorherrschenden, drastischen Umweltbedingungen zu schützen. Eine Adaption dieser Anpassungsmechanismen auch die vorliegende Problematik der BRP-induzierten Zelllyse könnte ebenfalls eine zielführende Strategie darstellen höhere Expressionsraten des BRPs zu realisieren. Das K-Antigen spielt als Oberflächen-Adhäsion zudem bei der bakteriellen Biofilmbildung die zentrale Rolle. Die Bedeutung bakterieller Biofilme als Schutzmechanismus vor harschen Umweltbedingungen mit ihren bedeutenden Implikationen für medizinische und wirtschaftliche Bereiche wurde eingehend diskutiert. Ein bisher nur aufgrund von Sequenzhomologien vorhergesagtes Transportsystem zur Zuckeraufnahme konnte aufgrund der konsistent positiven Regulation nach Induktion, welche auch im BRP-defizienten Stamm auftrat, experimentell bestätigt werden. Es handelt sich dabei um den Gencluster *ytfRQST*. Clusteranalysen wurden als wichtiges Werkzeug zur Evaluierung äquivalenter Expressionsprofile und biologischen Zusammenhängen zwischen Genen wurden in die Auswertungen implementiert. Anhand dieser konnten verschiedene signifikante Expressionsverläufe identifiziert werden und die in dem Cluster organisierten Gene in einen funktionellen Zusammenhang gestellt werden.

2. Microarray-Analyse mit sekretionsspezifischen *low-density* Microarrays

Im Rahmen der Analyse wurde die positive Regulation einer Reihe von potenziellen Kandidaten für Coexpressionstrategien detektiert. Die codierenden Gene für die Hitzeschock-Chaperone *DnaK* und *DnaJ* wurden nach Induktion signifikant positiv reguliert und De Lisa *et al.* (2007) konnten anhand der Coexpression dieses Genes eine signifikante Erhöhung der sekretierten Proteinmenge nachweisen. Die Coexpression von Chaperonen ist eine häufig angewandte und in vielen Fällen erfolgreiche Strategie für eine Optimierung der Sekretionseffizienzen. Chaperone assistieren fehlgefalteten und naszierenden Proteinen dabei die korrekte Konformation einzunehmen. Jüngste Ergebnisse anhand des Chaperonins GroEL weisen auf eine aktive Einflussnahme in den Faltungsprozess hin. Unter Ausnutzung der Abstandsabhängigkeit des FRET-Effektes wurde beim Substrat Maltosebindeprotein eine signifikante Streckung im Komplex mit dem Chaperonin nachgewiesen. Bei den Chaperoninen handelt es sich um eine Unterklasse der Chaperone, diese bestehen aus zwei großen Taschen in denen sich die naszierenden Polypeptidketten abgeschirmt von äußeren Einflüssen falten können. Diese von den Chaperoninen gewährte Abschirmung könnte auch für Sekretionsstrategien genutzt werden, da akkumulierende, rekombinante Proteine im Cytoplasma vermehrt proteolytisch degradiert werden. Brüser *et al.* (2007) stellten das Postulat auf, dass jedes Sekretionssubstrat vor der Translokation ins Periplasma mit einem Chaperon wechselwirkt und dieser Komplex so vor proteolytischer Degradation der Signalpeptide geschützt ist. Mit *nuoA* wurde die Regulation eines Genes beobachtet welches für die Aufrechterhaltung der Triebkraft der Tat-Translokation, dem pH-Gradienten über der Membran, verantwortlich ist. Vorangehende Experimente haben gezeigt, dass die

Überexpression von Genen welche für Komponenten der Translokase codieren nur zu einem gewissen Grad zu einer Steigerung der sekretierten Stoffmenge rekombinanter Proteine anwendbar ist. Grund ist der gesteigerte Protonenfluss durch die Cytoplasmamembran bei einer vermehrten Membranpassage. Dieses Dilemma lässt sich durch Coexpression von *pspA* umgehen, da dieses Phagenschockprotein die Integrität der Cytoplasmamembran bewahrt und so den transmembranen Protonenfluss unter hohen Translokationsraten stark reduziert. Das *psp*-Operon in dem auch *pspA* lokalisiert ist wurde anhand der whole-genome Experimente als positiv reguliert identifiziert welches diese Gene, in Koinkidenz mit dem eben Angemerktem, für Coexpressionstrategien prädestiniert. Im Falle der BRP-induzierten Sekretion werden somit zwei Vorteile verbunden, zum einen wird der pH-Gradient über der Membran aufrechterhalten und zum anderen erfolgt ein Schutz der BRP-exprimierenden Zelle vor Lyse durch Wahrung der Membranintegrität. Da der pH-Gradient über der Membran die alleinige Triebfeder für den Tat-Export darstellt und einen wichtigen Teil der Energie für die Translokation mittels des Sec-Pfades liefert, ist dessen Aufrechterhaltung von immenser Bedeutung. Im Rahmen der Clusterergebnisse von whole-genome Experimenten wurde das *nuo*-Operon ebenfalls als reguliert identifiziert, was dessen Relevanz auch für den Sec-Pfad impliziert. Durch die detektierte Regulation von Genen welche unmittelbar in den Sekretionsprozess involviert sind, wie dies z.B. für das Gen der Signalpeptidase 2 der Fall war, wurde die unmittelbare biologische Sinnhaftigkeit der generierten Ergebnisse verifiziert. Die Tat-spezifische Translokation von Reporterproteinen ins Periplasma, hier exemplarisch an einem GFP-Konstrukt gezeigt, bewirkt die positive Regulation von verschiedenen Chaperonen, Teilen der bakteriellen Translokase und von Komponenten, die an der Aufrechterhaltung des pH-Gradienten über der Cytoplasmamembran verantwortlich sind. All diese Gene stellen potenzielle, vielversprechende Ziele für Coexpressionstrategien dar.

3. Bioinformatische Entwicklungen, Aufbau einer Datenbank

Die Reproduzierbarkeit von Experimenten bezüglich der Planung, Ausführung und Datenauswertung stellt eine Notwendigkeit in wissenschaftlichen Arbeiten dar. Diese Bedingung kann bei der Konzipierung eines geeigneten Auswerteprogramms von Microarray-Experimenten durch die Integration elektronischer Datenblätter realisiert werden. In den Datenblättern sollten detaillierte Informationen zum Design und der Durchführung der Versuche gespeichert werden. Die Eingabe dieser versuchsbezogenen Informationen sollte nach dem experimentellen Teil der Microarray-Versuche erfolgen und an die Aufforderung zur Auswertung geknüpft sein. Die entsprechenden Ergebnisse zu diesen Experimenten sollten nach Vollendung der Daten-Analyse elektronisch abrufbar an die Datenblätter gekoppelt in einer internen Datenbank gespeichert sein. Auf diese Weise können differenzierte Folgeversuche unter gleichen Bedingungen geplant werden und Ergebnisse analoger Versuche im Nachhinein miteinander verglichen werden. Sowohl die Dateneingabe wie auch der spätere Abruf der Informationen zu vorangegangenen Experimenten sollten benutzerfreundlich erfolgen, so dass auch Experimentatoren ohne weitere Computer- und Programmierkenntnisse das Programm einfach bedienen können. Die notwendigen Erläuterungen zum Experiment, die in den Datenblättern festgehalten werden sollen, beinhalten sowohl formelle Daten wie Angaben zur Person (Experimentator) und zum Datum der Durchführung wie auch versuchsspezifische Daten beispielsweise bezüglich der verwendeten Organismen, des Extraktions-, Transkriptions- und Hybridisierungsprotokolls oder auch zu den Materialeigenschaften der Microarrays und dem Scanprotokoll. Sowohl die technischen als auch die Organismus-spezifischen Kriterien sollten in einer einheitlichen Struktur gespeichert werden, so dass sie mit einem Algorithmus benannt und gesucht werden können. Des Weiteren sollten die dafür notwendigen technischen Angaben unmittelbar der statistischen Auswertung übergeben werden. Abgesehen von diesen für die Auswertung

unerlässlichen technischen Informationen, werden in den Datenblättern auch technische Details zu den Experimenten abgefragt, die zwar für Wiederholungsexperimente unerlässlich, für die Auswertung selbst jedoch nicht erforderlich sind (Angaben zur Hybridisierungsdauer, Reduktionsbedingungen, Waschpuffern, usw.). Zusammenfassend bedeutet das, dass die experimentellen Details inhaltlich in Organismus-spezifische und technische Merkmale unterteilt werden können. Von den technischen Details muss ein Teil wiederum an die statistische Auswertung übergeben werden. Dieser Analyse-orientierten Unterscheidung zufolge gliedert sich die Datenblatt-Abfrage in zwei Gruppen, von denen die eine Informationen zum Microarray-Experiment beinhaltet, die sowohl technischer als auch Organismus-spezifischer Natur sind und für die Auswertung keine Rolle spielen. Diese Gruppe wird durch die ersten beiden der insgesamt drei Datenblättern repräsentiert. In die Gen-Annotationen fließen aber auch Informationen ein, die erst nach Durchlauf der Auswertung zur Verfügung stehen. So werden im Laufe der Analyse der Microarrays frei verfügbare Datenbanken wie z.B. *KEGG* oder *Multifun* mit den Accessionnummern der Gene auf die Existenz entsprechender Einträge abgesucht und die betreffenden regulatorischen Pfade, in denen das Gen erwiesenermaßen beteiligt ist, angegeben. Die Gen-Annotationen sind über eine Schnittstelle an das experimentelle Daten-System gekoppelt. Eine derartige Verknüpfung ist sinnvoll, da aus dem experimentellen Design des Chips alle weiteren Einzelinformationen der Experimente hervorgehen. So bedingt die über das experimentelle Design definierte Auswahl des Chips einer Microarray-Familie die auf den Microarrays vorkommenden Gene. Die Suche nach einem bestimmten Gen setzt daher in der Regel die Kenntnis der zugehörigen Chip-Familie voraus, so dass Gen-Annotationen über die Auswahl der entsprechenden Microarray-Familie abgerufen werden können. Die Suche liefert dann nicht nur alle der Microarray-Familie angehörigen Einzelexperimente der multikonditionellen Experimente, sondern über ein separates Excel-Datenblatt auch alle verfügbaren Informationen zu den Genen der Microarray-Familie. Um die in den Excel-Dateien vorliegenden Experiment-Informationen zu einem späteren Zeitpunkt nutzen zu können, muss der Benutzer jedoch nicht nur den Speicherort und -namen der Datei kennen, sondern auch wissen, ob eine entsprechende Datei überhaupt existiert. Da aber nicht vorausgesetzt werden kann, dass jeder Experimentator über alle in seinem Interessensbereich liegenden, bereits durchgeführten Microarray-Experiment im Bilde ist, muss eine Suchfunktion für die angelegten Excel-Dateien zur Verfügung gestellt werden. Abgesehen von der Suche nach konkreten Dateien, kann die Nachforschung nach definierten Experimenten (und den dazugehörigen Ergebnissen) basierend auf einzelnen in den Dateien enthaltenen Informationen zum Experiment anhand verschiedenster Suchkriterien (und damit verbunden: aus unterschiedlicher Motivation heraus) vorgenommen werden. So kann ein Benutzer beispielsweise alle bisherigen Versuche eines auf seinem Gebiet arbeitenden Kollegen abfragen wollen (*username*). Es kann nach allen mit einem bestimmten Organismus (*source*) oder einem spezifischen Chip (*chip*) gesucht werden. Ein Anwender kann auch die in einem bestimmten Zeitraum (*date*) durchgeführten Experimente erfragen wollen. Aus diesem Grund werden die zu einem Experiment gemachten Angaben nicht nur in Form von Excel-Dateien gespeichert, sondern auch in einer übergeordneten internen Datenbank angelegt, die alle Einzelinformationen kategorisch speichert. Der Funktionsweise dieser übergeordneten Datenbank liegen unterschiedliche Konzepte zugrunde, die den Anforderungen an die interne Datenbank gehorchen. Neben dem Anspruch der multiplen Datenspeicherung, ermöglicht die interne Datenbank auch sinnvolle Querverknüpfungen innerhalb zweckmäßiger Kategorien. Vor dem Hintergrund der Notwendigkeit einer detaillierten Dokumentation des experimentellen Designs und der Durchführung von Microarray-Experimenten wurden also über die Auswertung der Daten hinaus elektronische Datenblätter entwickelt, die mithilfe automatisierter Abfragemasken die Eingabe aller Einzelheiten zu den Chipversuchen erleichtern. Die auf das multivariate Anforderungsprofil der Microarrays ausgerichteten

Datenblätter werden einer internen Datenbank übergeben. Die Datenbank verwaltet somit alle durchgeführten Experimente und kann jederzeit anhand verschiedener Suchkriterien auf bereits durchgeführte Microarray-Experimente durchsucht werden.

Literatur:

Brüser T. (2007) The twin-arginine translocation system and its capability for protein secretion in biotechnological protein production. *Appl Microbial Biotechnol* 76, 35-45

Pérez-Rodríguez R, Fisher AC, Perlmutter JD, Hicks MG, Chanal A, Santini CL, Wu LF, Palmer T, DeLisa MP. (2007). An essential role for the DnaK molecular chaperone in stabilizing over-expressed substrate proteins of the bacterial twin-arginine translocation pathway. *J Mol Biol.* 367, 715-30