

**Genomweite Transkriptionsanalyse von
Methanosarcina mazei Gö1**

Dissertation
zum Erlangen des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Raymond L. Hovey
aus Göttingen

Göttingen 2003

D7

Referent:

Priv. Doz. Dr. U. Deppenmeier

Korreferent:

Prof. Dr. G. Gottschalk

Tag der mündlichen Prüfung:

Inhaltsverzeichnis

Abkürzungsverzeichnis	IV
1. Einleitung	1
2. Material und Methoden	5
2.1 Organismen und Plasmide	5
2.2 Nährmedien, Puffer und Stammlösungen	10
2.2.1 Medien für <i>E. coli</i>	10
2.2.1.1 LB (Luria-Bertani)-Medium	10
2.2.1.2 SOB-Medium	10
2.2.1.3 SOC-Medium	11
2.2.2 Medien für <i>Ms. mazei</i>	11
2.2.2.1 Standard Medium für <i>Ms. mazei</i> Gö1	11
2.2.2.2 Vitaminlösung	13
2.2.2.3 Spurenelementlösung	13
2.2.2.4 Medium zur Expressionsanalyse in <i>Ms. mazei</i> Gö1	14
2.2.3 Medienzusätze	15
2.2.4 Stamm- und Pufferlösungen	16
2.2.4.1 Tris-Puffer für die Agarose Gelelektrophorese	16
2.3 Zellanzucht und Stammhaltung	16
2.3.1 Anaerobe und aerobe Zellanzucht	16
2.3.1.1 Zellanzucht anaerober Organismen in Flüssigmedium	16
2.3.1.2 Aerobe Anzucht von <i>E. coli</i>	17
2.3.2 Stammhaltung	17
2.3.2.1 Anaerobe Kulturen von <i>Ms. mazei</i> Gö1	17
2.3.2.2 Aerobe Kulturen	17
2.3.2.3 Blau-weiß-Test zur Selektion rekombinanter Klone	17
2.3.3 Reinheitskontrolle	18
2.3.4 Messung des Bakterienwachstums	18
2.4 Standardtechniken für das Arbeiten mit Nukleinsäuren	19
2.4.1 Analyse von Sequenzdaten	19
2.4.2 Vorbehandlung von Geräten und Lösungen	19
2.4.3 Isolierung chromosomaler DNA mit CTAB	19
2.4.4 Isolierung von Plasmid-DNA	21
2.4.4.1 Präparative Plasmidisolierung	21
2.4.5 Reinigung und Konzentrierung von Nukleinsäuren	21
2.4.5.1 Reinigung von PCR-Produkten	21
2.4.5.1.1 Phenol/Chloroform-Extraktion	21
2.4.5.1.2 PCR-Produkt Aufreinigung mit dem Qiaquick PCR Purification Kit	22
2.4.5.1.3 Konzentrationsbestimmung und Reinheitskontrolle von DNA	23
2.4.5.2 Isolation von Gesamt-RNA aus <i>Ms. mazei</i> Gö1	24
2.4.6 Auftrennung von DNA-Fragmenten	24
2.4.6.1 Agarosegelelektrophorese	24
2.4.6.2 Größenbestimmung von Nukleinsäuren	26

2.5	Polymerasekettenreaktion (PCR)	27
2.5.1	Standard Polymerasekettenreaktion (PCR)	27
2.5.2	Kolonien-Screening mittels PCR	28
2.5.3	Real-Time-PCR	29
2.6	Enzymatische Modifikation von DNA	30
2.6.1	Herstellung Cy3/Cy5 markierter cDNA Sonden	30
2.7	Übertragung von DNA in <i>E. coli</i> durch TOPO-Transformation	31
2.7.1	Herstellung kompetenter Zellen zur Transformation	31
2.7.2	TOPO-TA Klonierung	32
2.8	Herstellung und Verwendung von cDNA Microarray Chips	33
2.8.1	Übertragung von PCR Produkten auf Aminosilan beschichtete Glas Slides	33
2.8.2	Hybridisierung von DNA-Microarray Chips	33
2.8.3	Einlesen der DNA-Mikroarray Chips	36
2.9	Bioinformatische Materialien und Methoden	36
2.9.1	Standard Software	37
2.9.1.2	Ultraedit 32	37
2.9.1.3	Linux Redhat 2.1	37
2.9.1.4	Solaris 5	38
2.9.2	Implementation von Programmiersprachen in das Betriebssystem	38
2.9.2.1	ActiveState PERL	38
2.9.2.2	Java 2 Runtime Environment und die Java Virtual Machine	38
2.9.3	Primer3 und der Server der UC-Irvine	39
2.9.4	BLAST-Server	39
2.9.5	ERGO Data Suite	39
2.9.6	GenePix-Pro	40
2.9.7	Analyst Expressionist	40
2.9.8	Selbst erstellte Software	40
2.10	Chemikalien, Biochemikalien und Gase	41
3.	Experimente und Ergebnisse	43
3.1	Entwicklung von bioinformatischen Programmen zur Unterstützung der Annotation der Gene aus <i>Ms. mazei</i> Gö1	43
3.1.1	Extraktion der Sequenzdaten aller <i>Ms. mazei</i> Gö1 Gene aus der ERGO-Data-Suite	44
3.1.2	Sequenzabgleich zur Identifizierung von Bakterien-ähnlichen Genen	47
3.2	Genomweite Transkriptionsanalyse von <i>Ms. mazei</i> Gö1	50
3.2.1	Erstellung einer PCR-Produkt-Genbank des <i>Ms. mazei</i> -Genoms	50
3.2.2	Generierung der Genom umfassenden DNA-Microarray-Chips	55
3.2.3	Herstellung der fluoreszenzmarkierten Sonden	56
3.3	Vorversuche mit den Genom umfassenden DNA-Mikroarray-Chips	61

3.3.1	Genom-weite Transkriptionsanalyse von <i>Ms. mazei</i> Gö1 beim Wachstum auf Voll- bzw. Minimalmedium	61
3.3.2	Ergebnisse der DNA-Microarray Analysen von <i>Ms. mazei</i> Gö1 nach Wachstum auf Voll- bzw. Minimalmedium	68
3.3.3	Erkenntnisse aus den DNA-Microarray Daten von <i>Ms. mazei</i> Gö1 nach Wachstum auf Voll- bzw. Minimalmedium	73
3.4	Analyse der Gen-Expressionsmuster von <i>Ms. mazei</i> Gö1 beim Wachstum auf verschiedenen Kohlenstoffquellen	74
3.4.1	Expressionsanalyse von <i>Ms. mazei</i> Gö1 beim Wachstum auf Methanol bzw. auf Acetat	75
3.4.2	Quantitative Expressionsanalyse ausgewählter Gene durch Real-Time-PCR	88
3.5	Zusammenfassende Betrachtung zu den Experimenten und Ergebnissen dieser Arbeit	97
4.	Diskussion	98
4.1	Betrachtungen zum Stoffwechsel von <i>Methanosarcina mazei</i> Gö1	98
4.1.1	Eigenschaften von ungewöhnlichen methanogenen Kofaktoren	98
4.1.2	Methanbildung aus Methanol	100
4.1.3	Methanbildung aus Acetat	104
4.2	Anmerkungen zur Genom-weiten Transkriptionsanalyse in <i>Ms. mazei</i> Gö1 durch die DNA-Microarray Chip-Technologie	106
4.3	Anmerkungen zu bioinformatischen Ansätzen zur Genom-weiten Transkriptionsanalyse in <i>Ms. mazei</i> Gö1	108
4.3.1	Anmerkungen zur Programmiersprache PERL	108
4.3.2	Anmerkungen zur Computer-gestützten Analyse der Genom-weiten Transkriptionsdaten (=Data Mining)	111
4.3.3	Anmerkungen zur Gruppierung von Daten durch Cluster-Algorithmen im Prozeß des Data Minings	114
4.3.3.1	Hierarchische Algorithmen	115
4.3.3.2	Partitionelle Algorithmen	116
4.4	Untersuchungen der unterschiedlichen Expressionsmuster beim Wachstum von Zellen auf Methanol oder Acetat	118
4.4.1	Enzyme des oxidativen Zweigs der Methanbildung aus Methanol	120
4.4.2	F ₄₂₀ H ₂ Dehydrogenase	121
4.4.3	Ech-Hydrogenase	122
4.4.4	Methyltransferasen	124
4.4.5	Enzyme des aceticlastischen Stoffwechselwegs der Methanogenese	126
4.4.6	Biosynthese von Aminosäuren und Dipeptid/Oligopeptid Aufnahmesysteme	127
4.4.7	Detoxifizierung im aceticlastischen Stoffwechselweg	128
4.4.8	Verwertung von Methylthiolen	128
4.4.9	Kobalt Aufnahme	129
4.4.10	Eisen (III) Dicitrat Aufnahme	130
4.4.11	Vermutliche Regulatoren	131

4.5	Gene mit gleichem Expressionsmuster beim Wachstum von Zellen auf Methanol oder Acetat	133
4.5.1	Methyl-CoM Reduktase und Methyl-H ₄ MPT:HS-CoM Methyltransferase	133
4.5.2	F ₄₂₀ H ₂ -nicht-reduzierende Hydrogenase	134
5.	Zusammenfassung	136
6.	Literaturverzeichnis	139
A.	Anhang A – Quellcode und Dokumentation der selbst erstellten Computerprogramme	165
	Anhang B – Daten der Real-Time-PCR	CD
	Anhang C – Daten der DNA-Mikroarray-Chips Voll-/Minimalmedium	CD
	Anhang D – Daten der DNA-Mikroarray-Chips Acetat-/Methanol	CD

Abkürzungsverzeichnis

Abb.	Abbildung
Amp	Ampicillin
AS	Aminosäure(n)
ATP	Adenosintriphosphat
ATPase	ATP-Synthase
AU	Absorptionseinheiten
bp	Basenpaare
BSA	Rinderserumalbumin
bzw.	beziehungsweise
c-	centi
ca.	circa
CH ₄	Methan
CoM-S-S-CoB	gemischtes Disulfid aus HS-CoM und HS-CoB
CTAB	Cetyltrimethylammoniumbromid
Cyt.	Cytochrom
Da	Dalton
d. h.	dass heisst
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure
DNase	Desoxyribonuklease
ds	doppelsträngig
DSMZ	Deutsche Sammlung für Mikroorganismen
<i>E.</i>	<i>Escherichia</i>
E	Extinktion
ε	Extinktionskoeffizient
e ⁻	Elektron
EDTA	Ethylendiamintetraessigsäure
eds.	Editor: Herausgeber
et al.	und andere
EtOH	Ethanol
F ₄₂₀	(N-L-Lactyl-γ-L-GlutamyI)-L-Glutaminsäure-Phosphodiester des 7,8-Didemethyl-8-Hydroxy-5-Desazariboflavin-5'-Phosphates
Fa.	Firma
g	Gramm
h	Stunde
H ₂ O _{bidest.}	zweifach destilliertes Wasser
H ₂ O _{demin.}	demineralisiertes Wasser
H ₄ MPT	Tetrahydromethanopterin
HS-CoM	2-Mercaptoethansulfonat
HS-CoB	7-Mercaptoheptanoylthreoninphosphat
Hz	Hertz
i. d. R.	in der Regel
k-	Kilo
Kan	Kanamycin
kb	Kilobasenpaare
Konz.	Konzentration
l	Liter
log	Logarithmus (dekadisch)

M	molar
m	Meter
M	Mega
max.	maximal
MeOH	Methanol
mod.	modifiziert
μ-	mikro
m-	milli
MCS	multiple-cloning-site
MF	Methanofuran
min	Minute
<i>Ms.</i>	<i>Methanosarcina</i>
MT	Methyltransferase
MW	Molekulargewicht
n-	nano
NAD(P)	Nicotin-Adenin Dinucleotid (Phosphat)
NMR	Kernmagnetische Resonanzspektroskopie
Nr.	Nummer
°C	Grad Celsius
OH-	Hydroxy-
OD	Optische Dichte
PCR	Polymerase Kettenreaktion
ppm	Parts per million
RBS	Ribosomenbindestelle
RF	relative Mobilität
RNA	Ribonukleinsäure
rRNA	Ribosomale Ribonukleinsäure
rpm	Umdrehungen pro Minute
RT	Raumtemperatur
s	Sekunde
SDS	Natriumdodecylsulfat
SL	Spurenelemente-Lösung
Tab.	Tabelle
Tris	Tris (Hydroxymethyl) Aminomethan
U	Unit, Enzymeinheit (1 U = 1 μmol Substrat/min)
UE	Untereinheit
UV	Ultraviolett
v/v	Volumen pro Volumen
V_{\max}	Maximale Reaktionsgeschwindigkeit
w/v	Gewicht pro Volumen
x g	x-fache Erdbeschleunigung

1. Einleitung

Aufgrund ihrer 16 S rRNA-Sequenzen sind die Archaea im phylogenetischen Stammbaum von den Bacteria und Eukarya getrennt. Die Archaea unterteilen sich in die Crenarchaeota (extrem thermophile und thermoacidophile Archaea), die Euryarchaeota (insbesondere extrem halophile und methanogene Archaea) (Woese et al., 1990; Barns et al., 1996) und die durch ihre sequenzierten rRNA-Gene neugebildete Gruppe der Korarchaeota (Brown und Doolittle, 1997). Die Vertreter der letzten Gruppe ließen sich bisher nicht im Labor anziehen. Seit neuestem bilden die Nanoarchaeota eine eigene vierte Gruppe, von denen bisher nur *Nanoarchaeum equitans*, ein als Symbiont lebendes Archaeon mit einem nur 0,5 Mb grossen Genom, isoliert wurde (Huber et al., 2002).

Die Gruppierung der Archaea in ein eigenes Reich wird durch besondere morphologische und biochemische Eigenschaften bestätigt (Wolfe, 1985; Kates et al., 1993; Keeling et al., 1994; Klenk und Doolittle, 1994). Archaeale Organismen besitzen eine besondere Zellwandstruktur, die nicht wie die der Eubakterien aus einem Peptidoglykangerüst (β -1,4-glykosidische Mureinsäureverknüpfung) besteht, sondern aus Pseudomurein (N-Acetyl-Glucosamin β -1,3-glykosidisch mit Talosaminmuronsäure verknüpft), Polysacchariden oder Proteinen (Kandler und Hippe, 1977; Kandler und König, 1978) aufgebaut ist. Die Zellwände der Archaea sind deshalb gegen Lysozym, Penicillin und andere Zellwand-Antibiotika resistent. Die Cytoplasmamembran enthält Glycerinether mit C₂₀ (Phytanyl)- und C₄₀ (Biphytanyl)-Isoprenoidalkylen an Stelle von Fettsäureglycerinestern (Tornabene und Langworthy, 1978; Jones et al., 1987; De Rosa und Gambacorta, 1988; Kates, 1993). Der Translationsapparat ist unempfindlich gegenüber Chloramphenicol, wird aber durch Diphtherietoxin gehemmt, was auf den Proteinbiosyntheseapparat rückschließen lässt. Bei Archaeen ist wie bei Bakterien nur eine DNA-abhängige RNA-Polymerase für die Proteinbiosynthese vorhanden, deren Untereinheitenstruktur aber zu den korrespondierenden Enzymen aus Eukaryoten homolog ist (Zillig et al., 1982). Die Größe der archaeellen Ribosomen liegt zwischen derjenigen von Bacteria und Eukarya (Schmidt und Böck, 1984). Ferner wird während der Translation, Methionin statt Formyl-Methionin als Startaminosäure benutzt (Jones et al., 1987), und in der tRNA konnte kein Ribothymidin

nachgewiesen werden (Fox et al., 1980). Aus *Saccharomyces cerevisiae* oder *Homo sapiens* isolierte TATA-Box-Bindeproteine können archaeelle Transkriptionsfaktoren in zellfreien Transkriptionssystemen aus *Methanococcus voltae* ersetzen (Wettach et al., 1995), und Sequenzen von archaeellen Histonen ähneln denen aus Eukaryoten (Sandman et al., 1990). Außerdem wurden von Di Ruggiero et al. (1997) DNA-Reparaturproteine identifiziert, die homolog zu den korrespondierenden Enzymen der Eukarya sind und in den Bakterien nicht vorkommen.

Die Gruppe der Archaea steht also bezüglich der Transkription, der Translation und der Replikation den Eukarya näher als den Bacteria, denen sie aber wiederum im Zellstoffwechsel und in der Energiegewinnung ähneln. Diese frühen Befunde finden Bestätigung in Sequenzvergleichen mit Genomen von Eukaryoten und Bakterien, die durch die archaeellen Genomprojekte der vergangenen Jahre möglich sind (Bult et al., 1996; Klenk et al., 1997; Smith et al., 1997; Deppenmeier et al., 2002).

Der in dieser Arbeit untersuchte methanogene Organismus *Methanosarcina (Ms.) mazei* ist ein methylothermer Organismus der Familie *Methanosarcinaceae* und ist befähigt, einfache C₁-Verbindungen wie Methanol und Methylamine zu verwerten. Ferner kann *Ms. mazei* Göl als einer der wenigen Vertreter dieser Familie auch Acetat und H₂/CO₂ als Energie- und C-Quelle nutzen (Jarrel und Kalmokoff, 1988; Whitman et al., 1991; Boone et al., 1993; Deppenmeier et al., 2002a). Durch die Umsetzung der genannten Substrate zu Methan tragen die methanogenen Archaea als letztes Glied der anaeroben Nahrungskette zur Aufrechterhaltung des Kohlenstoffkreislaufes bei (Klass et al., 1984; Garcia et al., 2000). Als natürliche Standorte sind Reisfelder, Sümpfe, Feuchtgebiete, Faulschlamm (auch der Stamm Göl wurde aus einem Faulturm in Göttingen isoliert), Sedimente von Seen, Flüssen und anderen Gewässern sowie der Pansen von Wiederkäuern und der Verdauungstrakt von Termiten zu nennen. Besonders durch verstärkte Rinderzucht und intensiv betriebenen Reisanbau in den letzten Jahrzehnten wurde die Methanogenese stark erhöht, so dass methanogene Archaea jährlich etwa 10⁹ t Methan produzieren (Friedmann et al., 1990; Weiss und Thauer, 1993). Methanotrophe Bakterien reoxidieren gut die Hälfte davon, während die andere Hälfte in die Atmosphäre gelangt und zum Treibhauseffekt beiträgt. So lassen sich

inzwischen etwa 70 % des Methangehaltes der Atmosphäre, der nach Cammack (1997) bei 1,8 ppm liegt, auf biologische Methanogenese-Prozesse zurückführen.

Ms. mazei besitzt die speziellen Cofaktoren Methanofuran und Tetrahydro-methanopterin, die als C₁-Gruppenüberträger fungieren (Van Beelen et al., 1984; Leigh et al., 1985; Möller-Zinkhan et al., 1989). Der zentrale Elektronenüberträger ist das 5-Deazaflavin-Derivat F₄₂₀, das in chemischer Hinsicht dem monozyklischen Nikotinamid der Eukarya und Bacteria gleicht. Die Regeneration des reduzierten F₄₂₀ erfolgt bei allen methylo trophen Methanarchaea über eine F₄₂₀H₂-Dehydrogenase. Dieser Enzymkomplex wurde aus *Ms. mazei* Gö1 isoliert (Abken und Deppenmeier, 1997). Beim Wachstum auf Methanol konnte in dem methylo trophen, methanogenen Organismus *Ms. mazei* Gö1 gezeigt werden, dass die F₄₂₀H₂-Dehydrogenase Teil eines Energie konservierenden Elektronentransportsystems, der F₄₂₀H₂:Heterodisulfid-Oxidoreduktase, ist. Diese verbindet die Oxidation des reduzierten F₄₂₀ mit der Reduktion eines Heterodisulfids, zusammengesetzt aus den beiden Coenzymen M (2-Mercaptoethansulfonat) und B (7-Mercaptoheptanoyl-threoninphosphat) (Deppenmeier et al., 1990a, b, 1996). Die Elektronen des F₄₂₀H₂ werden durch die F₄₂₀H₂-Dehydrogenase auf den neu entdeckten membranintegralen Elektronenüberträger Methanophenazin übertragen, wobei ein zur ATP-Synthese nutzbarer Protonengradient aufgebaut wird (Abken, 1997; Abken et al., 1998; Bäumer et al., 1998; Ide et al., 1999; Bäumer et al., 2000). Abschließend hierzu sei erwähnt, daß das Genom von *Ms. mazei* Gö1 komplett sequenziert wurde und mit 3371 offenen Leserahmen und 4,1 Millionen Basenpaaren das z.Z. zweitgrößte Genom aller Archaeen besitzt (Deppenmeier et al., 2002). Die Kenntnis der DNA Sequenz und der offenen Leserahmen eines Organismus sagt jedoch noch nichts darüber aus, ob die entsprechenden Gene tatsächlich exprimiert werden und wie die Expression reguliert wird. Zum Beispiel werden die Proteine für die F₄₂₀H₂-Dehydrogenase in *Methanosarcina Mazei* Gö1 von zwei verschiedenen Operonen kodiert, die auf unterschiedlichen Positionen des Genoms lokalisiert sind (Deppenmeier et al., 2002). Ob diese Operone unterschiedlichen regulatorischen Mechanismen unterliegen, ist nicht bekannt.

Die Kenntnis über die Genregulation in Archaeen ist im allgemeinen zur Zeit sehr gering. Der Schwerpunkt der Forschung dieser Arbeit liegt daher in der relativen genomweiten Expressionsanalyse aller Gene des Archeons *Ms. mazei* Gö1 unter verschiedenen Wachstumsbedingungen. Hierbei soll von allen Genen das Expressionsverhalten über DNA-Mikroarray-Chips gemessen werden, um Rückschlüsse auf regulatorische Mechanismen zu ermöglichen und neue Verknüpfungen und Abhängigkeiten von Genen in Bezug auf die Stoffwechselwege zu entdecken.

2. Material und Methoden

2.1 Organismen und Plasmide

Die im Verlauf dieser Arbeit benutzten Archaea sind in Tabelle 1, die Bakterien in Tab. 2 und die verwendeten PCR- und Sequenzier-Oligonukleotide in Tab. 3 aufgelistet. Tab. 4 gibt einen Überblick über die verwendeten Klonierungsvektoren und rekombinanten Plasmide.

Tabelle 1: Archaea

<u>Bezeichnung</u>	<u>DSM-Stamm-Nr.</u>
<i>Methanosarcina mazei</i> Gö1	3647
<i>Methanosarcina mazei</i> Gö1 auf Acetat adaptiert	7222

Tabelle 2: Bakterien

<u>Stamm</u>	<u>relevanter Geno-/ Phänotyp^a</u>	<u>Referenz / Herkunft</u>
<i>Escherichia coli</i> DH5 α	F ⁻ , supE44, Δ lacU169, ϕ 80 lacZ Δ M15, hsdR17(r _K ⁻ K ⁺) thi-1, λ ⁻ , recA1, gyrA96, relA1, endA1	Hanahan, 1983

^aGenotyp-Abkürzungen für *E. coli*: siehe Bachmann (1987)

Tabelle 3: Oligonukleotide

Bezeichnung	Bemerkung/Erkennungsstelle/Verwendungszweck
Abi_for	Zur Sequenzierung und PCR-Amplifizierung aus pTZ19R 5'- ACGACGTTGTAAAACGACGGCCAG -3'
Abi_rev	Zur Sequenzierung und PCR-Amplifizierung aus pTZ19R 5'- GCCAAGCTCTAATACGACTCAC -3'
TOPO_for	Zur Sequenzierung und PCR-Amplifizierung aus pCR2.1 5'- GTAACGGCCGCCAGTGTGCTGGAATTC -3'
TOPO_rev	Zur Sequenzierung und PCR-Amplifizierung aus pCR2.1 5'- CAGTGTGATGGATATCTGCAGAATTCGC -3'
UCLA_for	Zur Sequenzierung und PCR-Amplifizierung aus pCR4.0 5'- CTAGTCCTGCAGGTTTAAACGAATTCGCCC -3'
UCLA_rev	Zur Sequenzierung und PCR-Amplifizierung aus pCR4.0 5'- AATTGAATTTAGCGGCCGCGAATTCGCCC -3'

Diese Tabelle zeigt nur diejenigen Primer, welche homologe Sequenzen zu ‚Multiple-Cloning-Sites‘ der verwendeten Vektoren haben. Die Sequenzen der ‚Multiple-Cloning-Sites‘ der Vektoren sind in den Abbildungen (1 und 2) dokumentiert. Ferner wurden für jeden ORF aus *Ms. mazei* ein 3'-terminaler und ein 5'-terminaler Primer generiert. Die Sequenzen dieser Primer wurden gegen die Sequenz-Datenbank des Göttinger Genomics Laboratory (Göttingen, Germany) abgeglichen, um eine mögliche PCR-Amplifizierungen mittels eines ORF-spezifischen Primer und des abi_for/abi_rev-Primers zu ermitteln. Ergab dieser Abgleich ein mögliches PCR Produkt, wurde nur ein spezifischer Primer bestellt (Fa. MWG, Ebersberg). Wenn es nicht möglich war ein PCR-Fragment mit einem Vektor der Genbank zu produzieren, wurden zwei spezifische Primer bestellt.

Die Primer wurden mittels des Programmes Primer3 (2.9.3) oder des „microarray tools“ auf dem Server der University of California, Irvine, generiert (2.9.3).

Tabelle 4: Vektoren und rekombinante Plasmide

<u>Plasmid</u>	<u>Größe des Plasmids (kb)</u>	<u>relevante Merkmale</u>	<u>E. coli- Wirtsstamm</u>	<u>Referenz / Herkunft</u>
pCR2.1-TOPO	3,9	<i>lacZ</i> , T7-Promotor, SP3-Promotor, Gebundene TOPO-Isomerase, Amp ^r , Kan ^r	DH5 α oder TOP10	Fa. Invitrogen, Karlsruhe
pCR4-TOPO	3,9	<i>lacZ</i> , T7-Promotor, SP3-Promotor, Gebundene TOPO-Isomerase, Amp ^r , Kan ^r , ccdB	DH5 α oder TOP10	Fa. Invitrogen, Karlsruhe
pTZ-19R	3,2	Amp ^r , T7-Promotor, SP3-Promotor	DH5 α	Fa. Abi, St. Leon-Rot

In den folgenden Abbildungen sind die Vektorkarten der verwendeten Plasmide mit den Sequenzen der ‚Multiple-Cloning-Sites‘ abgebildet.

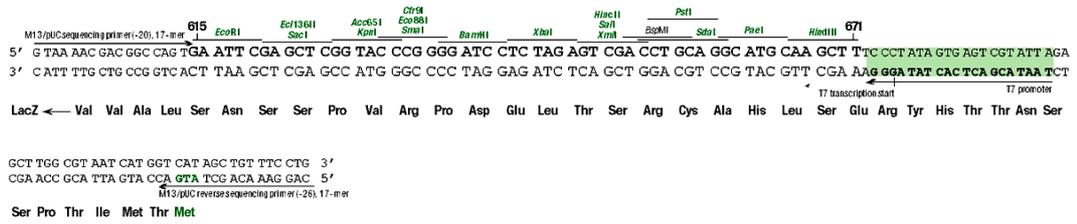
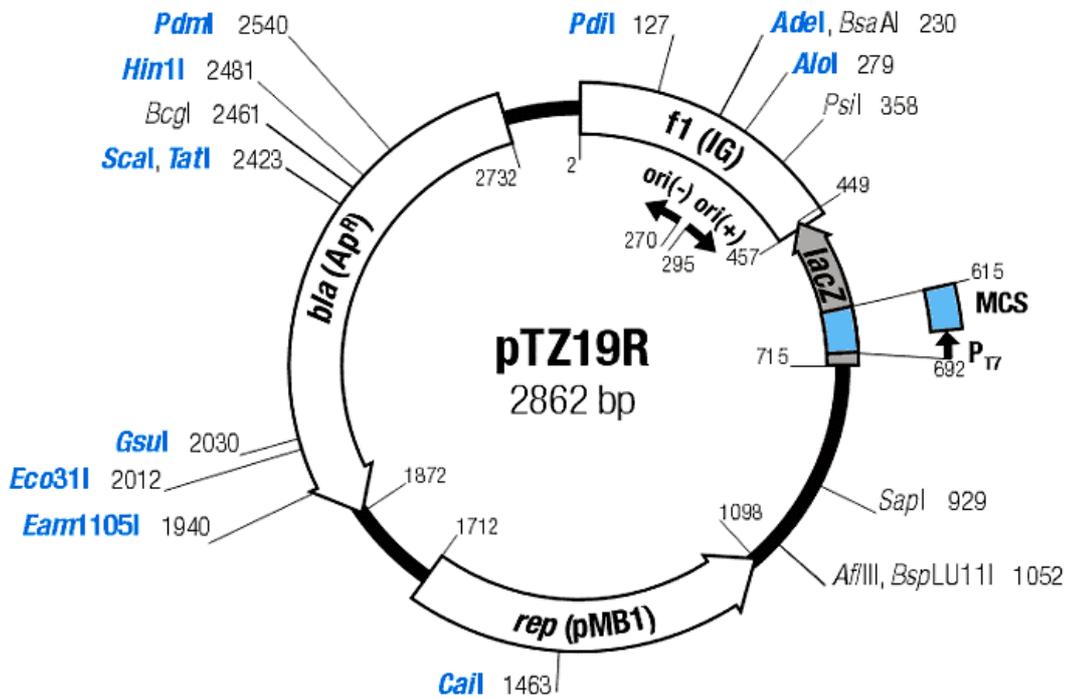


Abb. 1: Der Vektor pTZ19R

Dieser Vektor wurde vom Göttinger Genom Labor (G₂L) verwendet, um die Shotgun-Genbank von *Ms. mazei* zu klonieren. Die Plasmide dieser Shotgun-Genbank dienen zur Generierung von PCR-Produkten für die DNA-Microarray-Chips.

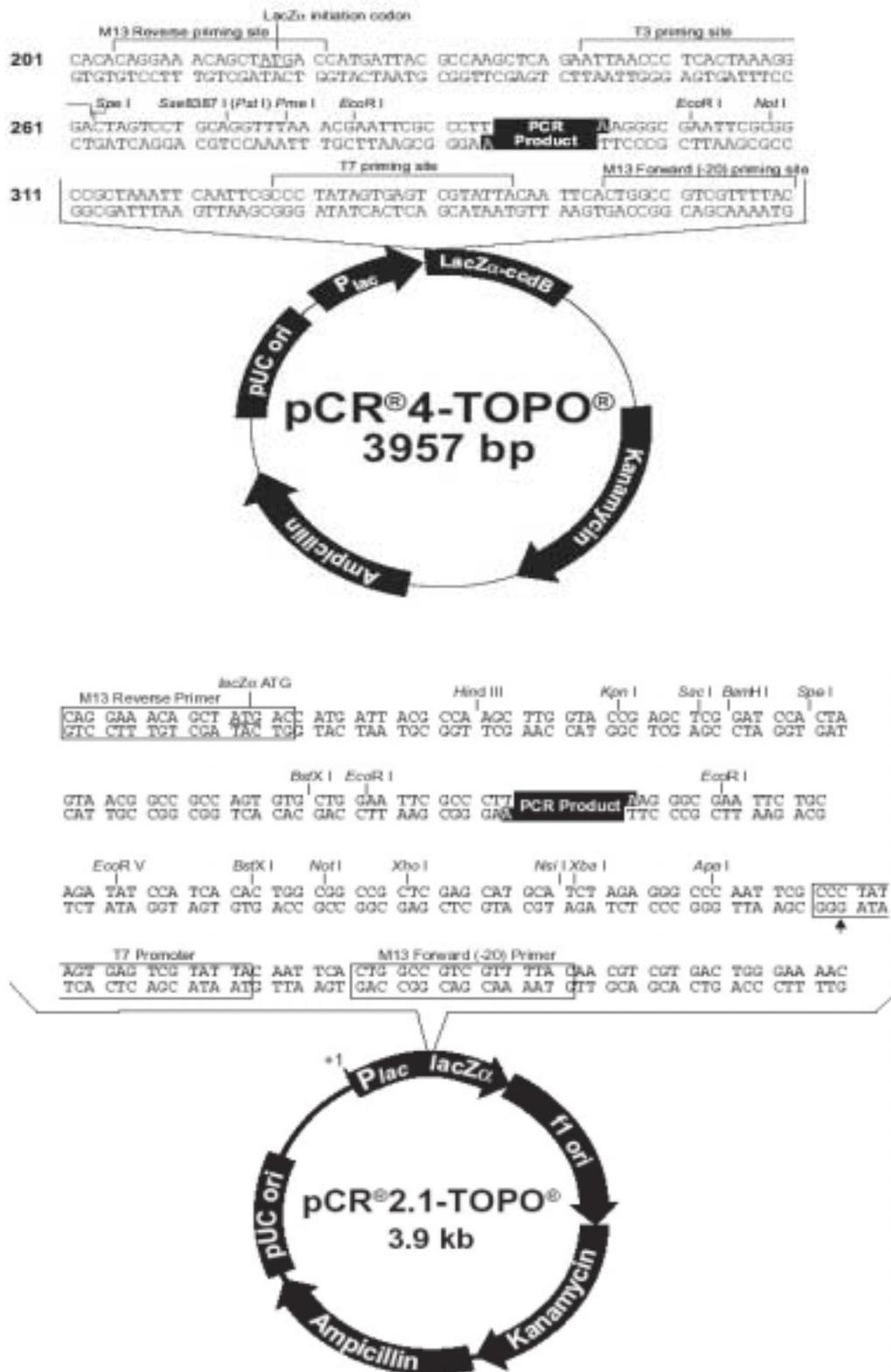


Abb. 2: Die Vektoren pCR2.1 und pCR4.0 der Firma Invitrogen, Karlsruhe

Diese Vektoren wurden benutzt, um die erhaltenen PCR Produkte zu klonieren. Die Firma Invitrogen stellte während des experimentellen Teils dieser Arbeit die Produktion um, sodass zuerst der Vektor pCR2.1 und später der Vektor pCR4 benutzt wurde.

2.2 Nährmedien, Puffer und Stammlösungen

2.2.1 Medien für *E. coli*

2.2.1.1 LB (Luria-Bertani)-Medium (Sambrook et al., 1989)

Bacto-Trypton	10,00 g
Hefeextrakt	5,00 g
NaCl	10,00 g
H ₂ O _{dest.}	ad 1000 g

Der pH-Wert des Mediums brauchte nicht eingestellt zu werden. Zur Herstellung von Agarplatten wurde Japanagar zu einer Endkonzentration von 1,5 % (w/v) zugesetzt.

2.2.1.2 SOB-Medium (Hanahan, 1983)

Trypton	20,00 g
Hefeextrakt	5,00 g
NaCl	0,58 g
KCl	0,19 g
H ₂ O _{dest.}	ad 990,00 ml

Nach dem Autoklavieren wurden 10 ml einer sterilfiltrierten Lösung aus 1 M MgCl₂ und 1 M MgSO₄ zugegeben.

2.2.1.3 SOC-Medium

Caseinpepton	2,0 %
Hefeextrakt	0,5 %
MgCl ₂	10,0 mM
MgSO ₄	10,0 mM
NaCl	10,0 mM
KCl	2,5 mM
Glucose	20,0 mM
H ₂ O	ad 100 ml
PH 7,0	

2.2.2 Medien für *Ms. mazei* Gö1**2.2.2.1 Standard Medium für *Ms. mazei* Gö1**

Die Anzucht des methanogenen Stammes *Ms. mazei* Gö1 erfolgte in modifiziertem Methanogenium-Medium (DSM-Nr. 141), das durch Kaliumphosphat und Natriumhydrogencarbonat gepuffert war.

K ₂ HPO ₄	0,348	g
KH ₂ PO ₄	0,227	g
NH ₄ Cl	0,500	g
Mg ₂ SO ₄ x 7 H ₂ O	0,500	g
CaCl ₂	0,250	g
NaCl	2,250	g
Na-Acetat	1,000	g
NaHCO ₃	0,850	g
Hefeextrakt	2,000	g
Casiton	2,000	g

FeSO ₄ x 7 H ₂ O	2,000 mg
Resazurinlösung (1mg/ml in H ₂ O _{bidest.})	1,0 ml
SL6 (PFENNIG und LIPPERT, 1966, modifiziert)	3,0 ml
Vitaminlösung (WOLIN et al., 1964, modifiziert)	10,0 ml
Methanol (99,5 %)	150,0 mM
oder	
Trimethylamin-HCl	25,0 mM
H ₂ O _{demin.}	ad 1000,0 ml

Die Nährmedien wurden nach der von Hungate (1969) und Bryant (1972) beschriebenen Technik anaerob hergestellt. Dazu wurde das Medium für 30 min mit einem N₂/CO₂ Gasmisch (80 % / 20 %, v/v) durchgast. Nach Ablauf dieser Zeit wurde die Lösung mit dem gleichen Gasmisch überschichtet und der letzte fehlende Zusatz, das NaHCO₃, hinzugefügt. Der pH-Wert wurde mit 6 N NaOH auf pH 6,9 eingestellt. Jeweils 5 ml dieser Lösung wurden in 16-ml-Hungate-Röhrchen bzw. 25 ml in Serumflaschen abgefüllt. Diese Gefäße wurden vor dem Befüllen mindestens 2 min mit dem genannten Gasmisch durchspült. Nach dem Befüllen der Kulturgefäße wurden diese mit Butylgummistopfen und Schraubverschlüssen bzw. Krampen verschlossen und bei 121 °C für 20 min autoklaviert. Na₂S und Cystein-HCl, die als Reduktionsmittel dienten, wurden separat in Serumflaschen als Stammlösungen (150 g/l) angesetzt und ebenso wie die 10 M Methanolstammlösung getrennt autoklaviert und dann dem sterilem Medium zugesetzt.

2.2.2.2 Vitaminlösung (Wolin et al., 1964, modifiziert)

Biotin	2,0	mg
Folsäure	2,0	mg
Pyridoxin-HCl	10,0	mg
Thiamin-HCl	5,0	mg
Riboflavin	5,0	mg
Nikotinsäure	5,0	mg
Ca-Pantothenat	5,0	mg
Vitamin B ₁₂	0,1	mg
p-Aminobenzoesäure	5,0	mg
α-Liponsäure	1,0	mg
H ₂ O _{bidest.}	ad 1000	ml

Die Lösung wurde bei 4 °C unter Lichtschutz gelagert.

2.2.2.3 Spurenelementlösung SL6 (Pfennig und Lippert, 1966, modifiziert)

10fach konzentrierte Stammlösung		
ZnSO ₄	1000	mg
MnCl ₂ x 4 H ₂ O	300	mg
H ₃ BO ₃	3000	mg
CoCl ₂ x 6 H ₂ O	2000	mg
CuCl ₂ x 2 H ₂ O	100	mg
NiCl ₂ x 6 H ₂ O	200	mg
Na ₂ MoO ₄	300	mg
H ₂ O _{bidest.}	ad 1000	ml

Die Lösung wurde bei 4 °C gelagert.

2.2.2.4 Medien zur Expressionsanalyse mit *Ms. mazei* Gö1

Die Anzucht der methanogenen Stämme *Ms. mazei* Gö1 (DSM-Nr. 3647) und *Ms. mazei* Gö1 (DSM-Nr. 7222; auf Acetat adaptiert) erfolgte in modifiziertem Methanogenium-Medium (DSM-Nr. 141), das durch Kaliumphosphat und Natriumhydrogencarbonat gepuffert war.

Für das Medium der beiden Stämme wurde eine gemeinsame Stammlösung erstellt.

K ₂ HPO ₄	0,348	g
KH ₂ PO ₄	0,227	g
NH ₄ Cl	0,500	g
Mg ₂ SO ₄ x 7 H ₂ O	0,500	g
FeSO ₄ x 7 H ₂ O	2,000	mg
CaCl ₂ x 2 H ₂ O	0,250	g
NaCl	2,250	g
Hefeextrakt	2,000	g
Casiton	2,000	g
NaHCO ₃	0,850	g
Resazurinlösung (1mg/ml in H ₂ O _{bidest.})	1,0	ml
SL6 (PFENNIG und LIPPERT, 1966, modifiziert)	3,0	ml
Vitaminlösung (WOLIN et al., 1964, modifiziert)	10,0	ml
H ₂ O _{demin.}	ad 1000	ml

Für Zellen, die auf Acetat angezogen wurden:

Pro Hungate-Röhrchen (5 ml) wurden 200 µl 3 M Na-Acetat Lösung (pH 6,9) zugesetzt.

Für Zellen, die auf Methanol angezogen wurden:

Pro Hungate-Röhrchen (5 m) wurden 200 µl Methanol-Lösung (3,75 Mol/l) zugesetzt.

Die Nährmedien wurden nach der von Hungate (1969) und Bryant (1972) beschriebenen Technik anaerob hergestellt. Dazu wurde das Medium für 30 min mit einem N₂/CO₂ Gasgemisch (80 % / 20 %, v/v) durchgast. Nach Ablauf dieser Zeit wurde die Lösung mit dem gleichen Gasgemisch überschichtet und der letzte fehlende Zusatz, das NaHCO₃, hinzugefügt. Der pH-Wert wurde mit 6 N NaOH auf pH 6,9 eingestellt. Jeweils 5 ml dieser Lösung wurden in 16-ml-Hungate-Röhrchen bzw. 25 ml in Serumflaschen abgefüllt. Diese Gefäße wurden vor dem Befüllen mindestens 2 min mit dem genannten Gasgemisch durchspült. Nach dem Befüllen der Kulturgefäße wurden diese mit Butylgummistopfen und Schraubverschlüssen bzw. Krampen verschlossen und bei 121 °C für 20 min autoklaviert. Na₂S und Cystein-HCl, die als Reduktionsmittel dienten, wurden separat in Serumflaschen als Stammlösungen (150g/l) angesetzt und ebenso wie die Methanol- und Na-Acetat-Lösungen getrennt autoklaviert und dann dem sterilen Medium zugesetzt.

2.2.3 Medienzusätze

Antibiotika-Stammlösungen wurden dem autoklavierten und auf mindestens 50 °C abgekühlten Medium zugesetzt und dienten zur Selektion entsprechender rekombinanter *E. coli*-Stämme.

Medienzusätze:

Zusatz	Stammlösung	Endkonzentration
Ampicillin	100 mg/ml in H ₂ O _{bidest.}	100 µg/ml
Kanamycin	50 mg/ml in H ₂ O _{bidest.}	50 µg/ml

Die Lösung wurde bei 4 °C gelagert

2.2.4 Stamm- und Pufferlösungen

2.2.4.1 Tris-Puffer für die Agarose-Gelelektrophorese

Für die Verwendung bei der Agarose-Gelelektrophorese wurden konzentrierte Stammlösungen bei 4 °C vorgehalten. Zur Einstellung geringerer Konzentrationen wurde mit H₂O_{bidest.} verdünnt:

20 x TAE-Puffer

Tris	96,9 g
Eisessig	22,8 ml
EDTA	7,5 g
H ₂ O _{bidest.}	ad 1000 ml
pH	8,5

2.3 Zellanzucht und Stammhaltung

2.3.1 Anaerobe und aerobe Zellanzucht

2.3.1.1 Zellanzucht anaerober Organismen in Flüssigmedium

Die Kultivierung der strikt anaeroben Bakterien erfolgte nach der Hungate-Technik (Hungate, 1969) und den modifizierten Methoden von Bryant (1972). Als Kulturgefäße dienten 16 ml Schraubdeckelröhrchen (Hungate-Röhrchen; Bellco Glass Inc., Vineland, New Jersey, USA) und Müller-Krempel-Serumflaschen (Fa. Ochs Glasgerätebau, Bovenden). Als Verschlüsse dieser Glasgefäße dienten gasdichte Stopfen aus Butylgummi bzw. Naturkautschuk. Zur Zugabe sowie Entnahme von Flüssigkeiten wurden sterile Plastik- oder Glasspritzen benutzt. Als Gasatmosphäre diente bei der Anzucht ein N₂/CO₂-Gemisch im Verhältnis von 80 % zu 20 %. Die Inkubation der Zellen erfolgte bei 37 °C für 2-4 Tage.

2.3.1.2 Aerobe Anzucht von *E. coli*

Aerobe Anzuchten von *E. coli* erfolgten in Kulturröhrchen oder Erlenmeyerkolben ohne Schikanen (Ochs, Bovenden). Das Kulturvolumen betrug ca. 20 % des Füllvolumens der Röhrchen oder Kolben. Die Kulturen wurden bei 130 rpm auf einem Rotationsschüttler bei 30 °C oder 37 °C inkubiert.

2.3.2 Stammhaltung

2.3.2.1 Anaerobe Kulturen von *Ms. mazei* Gö1

Die Stammhaltung von *Ms. mazei* Gö1 erfolgte bei 37 °C in 5 ml Komplexmedium. Alle drei bis vier Tage wurden die Zellen in frisches Medium überimpft. Eine Lagerung der Zellen konnte über wenige Tage bei 4 °C erfolgen. Wurden die Zellen nach dem Erreichen einer Optischen Dichte bei 600 nm (OD_{600nm}) = 0.5 - 0.6 bei 4°C gelagert, konnten sie auch für einen Zeitraum von 14 Tagen aufbewahrt werden.

2.3.2.2 Aerobe Kulturen

Für die kurzfristige Stammhaltung wurde *E. coli* auf LB-Agarplatten (plus entsprechende Antibiotika) bei 4 °C gelagert. Diese Kulturen wurden alle 4 Wochen überimpft. Zur längerfristigen Lagerung wurden 850 µl Zellsuspension einer Übernachtkultur mit 150 µl 87 %igem Glycerin versetzt und in flüssigem Stickstoff eingefroren. Die Lagerung erfolgte bei - 70 °C.

2.3.2.3 Blau-Weiß-Test zur Selektion rekombinanter Klone

Bei dieser Methode (auch bekannt unter dem Namen ‚Alpha-Komplementations-Test‘) handelt es sich um eine Variante der kurzfristigen Stammhaltung von *E.coli* auf Agarplatten nach einer Transformation.

Zur Selektion von Transformanten, die ein Insert im Plasmid enthielten, wurde dem Agar IPTG (48 $\mu\text{g/ml}$) und β -D-Galactosid (X-Gal; 40 $\mu\text{g/ml}$) zugesetzt (Horowitz et al., 1964). Der verwendete *E. coli* Klonierungsstamm DH5 α ist durch eine Deletion im *lacZ*-Gen gekennzeichnet und kann daher keine aktive β -Galactosidase mehr bilden. Durch Transformation mit Plasmiden wie pCR2.1 oder pCR4.0, die über das Gen der α -Untereinheit der β -Galactosidase verfügen, in dem sich wiederum die Multiklonierungsstelle befindet, kommt es zur Bildung eines aktiven Enzyms. Dies kann auf X-Gal und IPTG-haltigen Platten sichtbar gemacht werden, auf denen das *lacZ*-Gen durch IPTG induziert wird und die β -Galactosidase das Glukoseanalogon X-Gal spaltet, wonach durch Luftoxidation ein blauer Farbstoff entsteht. Soll in die Multiklonierungsstelle des Plasmides ein Insert ligiert werden, kann man den Erfolg der Insertion nach Transformation anhand der Kolonienfarbe sehen. Zellen, die Plasmide mit einer Insertion in der Multiklonierungsstelle und damit im *lacZ*-Gen beinhalten, können keine aktive β -Galactosidase bilden und erscheinen weiß (Blau/Weiß-Selektion).

2.3.3 Reinheitskontrolle

Kulturen von *E. coli* und *Ms. mazei* wurden im Phasenkontrastmikroskop (Fa. Carl Zeiss, Oberkochen) anhand ihrer Zellmorphologie auf Reinheit überprüft. Auch die Koloniemorphologie wurde als Nachweis einer Reinkultur verwendet, wozu von *E. coli* Vereinzelausstriche auf LB-Agarplatten angelegt wurden.

2.3.4 Messung des Bakterienwachstums

Der Wachstumsverlauf der Kulturen wurde über die optische Dichte bei einer Wellenlänge von 600 nm in einem Spektralphotometer UV-120.02 (Fa. Shimadzu Europa GmbH, Duisburg) verfolgt. Als Nullwert diente unbeimpftes Medium. Um Messungenauigkeiten zu vermeiden, wurden die Proben mit unbeimpftem Medium entsprechend verdünnt, sobald eine Extinktion von 0,3 überschritten wurde.

2.4 Standardtechniken für das Arbeiten mit Nukleinsäuren

2.4.1 Analyse von Sequenzdaten

Die DNA-Sequenzierungen der PCR-Produkte und der rekombinanten Plasmide wurde im Genomlabor des Instituts für Mikrobiologie und Genetik, Göttingen, durchgeführt. Die Analyse von Sequenzdaten erfolgte auf Macintosh-Computern (Fa. Apple Computers Inc., Cupertino, USA) mit Hilfe des Programms DNA-Strider (Marck, 1988). Sequenzabgleiche erfolgten mit der internen Datenbank des ERGO-Programms (Fa. Integrated Genomics, Chicago).

2.4.2 Vorbehandlung von Geräten und Lösungen

Zur Inaktivierung von Nukleasen wurden alle hitzestabilen Geräte und Lösungen durch Autoklavieren (20 min bei 121°C) sterilisiert. Nicht autoklavierbare Materialien wurden mit 70 %igem Ethanol (v/v) gespült oder mit 96 %igem Ethanol (v/v) abgeflammt. Hitzelabile Substanzen (z.B. Antibiotika) wurden sterilen Lösungen in pulverisierter Form zugegeben, oder entsprechende Lösungen wurden sterilfiltriert.

Für Materialien, die zum Arbeiten mit RNA verwendet werden sollten, wurden diese Schritte ein zweites Mal wiederholt. Geräte wurden kurz vor dem Arbeiten mit RNA mit einer ‚RNaseAWAY‘-Lösung sterilisiert (Fa. Molecular Bio Products, San Diego, Kalifornien, www.mbpinc.com).

2.4.3 Isolierung chromosomaler DNA mit CTAB (Ausubel et al., 1987)

Ms. mazei wurde in mit 500 ml Medium (siehe 2.2.2.1) gefüllten Müller-Krempel-Serumflaschen bis zur stationären Phase angezogen. Die Zellen wurden durch Zentrifugation (8000 rpm, 4 °C, 15 min im GS 3-Rotor in der Sorvall Zentrifuge RC-5B (Fa. Du Pont, Bad Homburg) sedimentiert und das Pellet in 9,5 ml TE-Puffer resuspendiert. Nach Zugabe von 0,5 ml 10 % (w/v) SDS und 50 µl Proteinase-K-Lösung (20 mg/ml) wurde 1 h bei 37 °C inkubiert. Anschließend wurden 1,8 ml einer

5 M NaCl und 1,5 ml einer CTAB/NaCl-Lösung zugegeben, geschwenkt und die Suspension für 20 min bei 65 °C im Wasserbad inkubiert. Zur Extraktion der Proteine wurden 13,3 ml (1 Vol.) Phenol/Chloroform 24:1 (v/v) zugegeben und gut gemischt. Nach der Zentrifugation (SS 34, 6000 rpm, 4 °C, 10 min) wurde die obere, wässrige Phase mit 0,5 Vol. Chloroform/Isoamylalkohol 1 : 1 (v/v) extrahiert. Zur resultierenden wässrigen Phase wurden zur Fällung der DNA 0,6 Vol. Isopropanol gegeben und erneut zentrifugiert (6000 rpm, 4 °C, 15 min). Das Pellet wurde mit 1 ml eiskalten 70 %igem (v/v) Ethanol gewaschen, für 10 min bei 60 °C getrocknet und danach in 2 ml TE-Puffer aufgenommen.

CTAB/NaCl-Lösung:

NaCl	700 mM
in 80 ml H ₂ O lösen, auf ca. 50 °C erwärmen	
CTAB	275 mM
langsam zugeben, lösen	
H ₂ O _{bidest.}	ad 100 ml

TE-Puffer:

Tris/HCl	10 mM
EDTA	1 mM
pH 8,0	

2.4.4 Isolierung von Plasmid-DNA

2.4.4.1 Präparative Plasmidisolierung

DNA-Sequenzierungen sowie Klonierungen erfordern einen hohen Reinheitsgrad der verwendeten DNA. Dies wurde durch die Benutzung des „Wizard DNA Purification Kits“ (Fa. Promega, Heidelberg) für die Isolierung von Plasmid-DNA sichergestellt. Die verwendeten Lösungen lagen dem Kit bei. Eine *E. coli*-Kultur wurde über Nacht in LB-Medium mit entsprechendem Antibiotikum (Ampicillin) bei 37 °C angezogen, in sterile 1,5ml Reaktionsgefäße überführt und abzentrifugiert (7000 rpm, 7 min; Biofuge 15, Fa. Heraeus GmbH, Osterode). Das Zellpellet wurde in 200 µl „Cell Resuspension Solution“ resuspendiert, mit 200 µl „Cell Lysis Solution“ versetzt und geschwenkt, bis die Lösung klar wurde. Die Neutralisierung erfolgte unter Zugabe von 200 µl „Neutralisation Solution“ und erneutem Umschwenken. Nach Zentrifugation (12000 rpm, 5 min) wurde der Überstand mit 1 ml „DNA Purification Resin“ gemischt und der Ansatz auf eine Minisäule gegeben, die anschließend mit 2 ml „Wash Solution“ gewaschen wurde. Die Elution der Plasmid-DNA von der Säule erfolgte mit 50 µl TE-Puffer (pH 8,0; 50 °C). Nach einminütiger Inkubation wurde abzentrifugiert (12000 rpm, 20 sec) und die DNA bei 4 °C gelagert.

2.4.5 Reinigung und Konzentrierung von Nukleinsäuren

2.4.5.1 Reinigung von PCR-Produkten

2.4.5.1.1 Phenol/Chloroform-Extraktion

Um PCR-Produkte im kleinen Maßstab zu reinigen, wurden 75 µl H₂O_{bidest.} zu 25 µl des Produkts pipettiert, und diese DNA-Suspension mit 1 Vol. Phenol-Lösung versetzt und kräftig gemischt. Das eingesetzte Phenol war mit 0,5 M Tris-HCl (pH8,0) gesättigt. Zur Phasentrennung wurde 3 min bei 13000 rpm zentrifugiert.

Danach wurde die obere, wässrige Phase in ein neues Eppendorf-Reaktionsgefäß überführt und mit 1 Vol. Chloroform/ Isoamylalkohol (24:1) extrahiert, bis keine weiße Interphase mehr sichtbar war und anschließend erneut zentrifugiert. Die obere Phase wurde mit Ethanol gefällt.

Zur Konzentrierung bzw. Entsalzung der Nukleinsäuren wurde eine Fällung mit 1/10 Vol. (v/v) Natriumacetat (2,5 M; pH 5,2) durchgeführt, um den Gehalt an monovalenten Kationen zu erhöhen. Die Präzipitation erfolgte durch Zugabe von 3 Vol. eiskaltem 99 %igen Ethanol (v/v). Nach vorsichtigem Durchmischen wurde die DNA bei der Ethanol-Fällung 30 min bei -20 °C inkubiert. Nach der Zentrifugation (13000 rpm, 4 °C, 25 min) wurde das DNA-Pellet einmal mit 1 ml eiskaltem 70 %igen Ethanol (v/v) gewaschen, getrocknet und in sterilem H₂O_{bidest.} aufgenommen.

2.4.5.1.2 PCR-Produkt Aufreinigung mit dem Qiaquick PCR Purification Kit

Alternativ kam für die Reinigung von PCR-Produkten der „Qiaquick PCR Purification“-Kit (Fa. Qiagen GmbH, Hilden) zum Einsatz. Bei dieser Methode wird die DNA bei hohen Salzkonzentrationen reversibel an eine Silica-Membran gebunden. Bei der Durchführung wurde nach Anweisung des Herstellers verfahren. Zur Elution der DNA wurde steriles H₂O genommen und die gereinigte DNA anschließend bei 4 °C oder -20 °C gelagert.

Für die Aufreinigung im 96er Maßstab diente eine automatisierte Variante dieses Kits, welche von einem Roboter vollzogen wurde (Qiagen Biorobot 2000, Fa. Qiagen GmbH, Hilden)

2.4.5.1.3 Konzentrationsbestimmung und Reinheitskontrolle von DNA

Die Konzentration von DNA-Lösungen wurde durch Absorption bei einer Wellenlänge von 260 nm in 1 ml Quarzküvetten mit einer Schichtdicke von 1 cm mit Hilfe eines Photometers (Uvikon 810 Photometer, Kontron Instruments, Eching) ermittelt.

Der Nullwert wurde mit Wasser oder TE-Puffer eingestellt, je nachdem, worin die DNA gelöst war. Für eine OD_{260} von 1,00 wurden folgende Konzentrationen angenommen (Sambrook et al., 1989):

doppelsträngige DNA	50 $\mu\text{g/ml}$
einzelsträngige DNA	40 $\mu\text{g/ml}$
Oligonukleotide	31 $\mu\text{g/ml}$

Zusätzlich konnte die Reinheit einer DNA-Lösung durch OD-Bestimmungen bei 230 nm und 280 nm überprüft werden. Für reine DNA gilt:

$$OD_{260} : OD_{280} = 1,8 \quad (\text{Sambrook et al., 1989})$$

$$OD_{230} : OD_{260} : OD_{280} = 0,45 : 1,0 : 0,515 \quad (\text{Marmur, 1961})$$

Verunreinigungen durch Proteine oder Phenol waren an einer deutlichen Steigerung der Absorption bei 280 nm erkennbar, Polysaccharide hingegen erhöhten die Absorption bei 230 nm. Konnte diese Methode der Konzentrationsbestimmung aufgrund der geringen Menge an vorhandener DNA nicht durchgeführt werden, wurde die Konzentration anhand der Intensität der Banden in Agarose-Minigelen geschätzt.

2.4.5.2 Isolation von Gesamt-RNA aus *Ms. mazei* Gö1

Zur Isolierung der Gesamt-RNA aus *Ms.mazei* Gö1 wurde das RNeasy-Kit (Firma Qiagen GmbH, Hilden) verwendet. Die Anzucht erfolgte, wie unter 2.3.2.1 angegeben, bis zu einer optischen Dichte bei 600 nm (OD_{600nm}) von 0,5. Es wurden je zwei Hungate-Röhrchen abzentrifugiert (10 min, 1700 rpm, 4°C) pro Aufreinigung. Der Überstand wurde verworfen, die Zellpellets in 250 µl 30 mM Natrium-Acetat (pH5,2) resuspendiert und in einem neuen Eppendorf Reaktionsgefäß vereinigt. Durch Zugabe von 100 µl 10% SDS wurden die Zellen anschließend bei Raumtemperatur und leichtem Schwenken lysiert. Hiernach wurden 350 µl Puffer RLT (990 µl RLT Kit Puffer + 10 µl Mercaptoethanol) hinzugegeben und sofort geschüttelt bis eine weißliche Färbung eintrat.

Bei den restlichen Schritten wurde nach den Angaben des Hersteller verfahren. Die noch enthaltene DNA wurde nach der Aufreinigung durch einen Verdau mit DNaseI (Fa. Boehringer Mannheim GmbH, Mannheim) entfernt (2 Stunden, 37 °C). RNA wurde aus dem Ansatz durch Phenol/Chloroform (2.4.5.1.1) extrahiert und ein Aliquot mittels Agarosegelelektrophorese auf Konzentration und DNA-Rückstände überprüft.

2.4.6 Auftrennung von Nukleinsäure-Fragmenten

2.4.6.1 Agarosegelelektrophorese

Sowohl zur analytischen als auch zur präparativen Auftrennung von DNA und DNA-Fragmenten wurden horizontale Gelelektrophoresen in Kammern der Größe 11,0 x 6,5 cm (kleine Gele) und 14 x 12 cm (große Gele) durchgeführt. Die Agarose wurde in Abhängigkeit der zu untersuchenden Fragmentgrößen in der Regel 0,6 %ig (w/v) bzw. 1,5 %ig (w/v) in TAE-Puffer angesetzt, welcher gleichzeitig auch als Elektrophoresepuffer diente. Für die Elektrophorese von RNA wurde die Agarose zweimal autoklaviert. Vor dem Auftragen in die Geltaschen wurden die DNA-Proben mit 0,2 Vol. Stop-Mix (= Beschwerer-Lösung) versetzt. Dadurch konnte man

während der Elektrophorese die Lauffront verfolgen, und eventuell noch vorhandene Proteine wurden denaturiert. Zur späteren Auswertung und Größenbestimmung wurden zusätzlich noch 4 µl (bei einem großen Gel 10 µl) DNA-Längenstandard aufgetragen (siehe 2.4.6.3). Nach der Elektrophorese (25-70 V, 0,5-2 h für Minigele; 40-120 V, 4-6 h für präparative Gele) wurden die Gele für 15 min im Ethidiumbromidbad (1 µg/ml H₂O) angefärbt, gewässert und die Nukleinsäurebanden mittels eines UV-Transilluminators in der Gel-Doc-Anlage (Gel-Doc-Anlage, Fa. Biorad, München) bei 254 nm sichtbar gemacht.

Bei Bedarf wurden die Gele zur Dokumentation in der Gel-Doc-Anlage (Fa. Biorad, München) digitalisiert. Wahlweise war anschließend eine Übertragung des Bandenmusters auf Thermopapier mittels einer Thermaldruckers möglich (Fa. Biorad, München).

Stop-Mix (Sambrook et al., 1989):

EDTA	19 mg	1,0 mM
Harnstoff	21 g	7,0 M
Saccharose	25 g	50,0 % (w/v)
Bromphenolblau	50 mg	0,1 % (w/v)
H ₂ O _{bidest.}	ad 50 ml	
pH 7,0		

Aliquots der autoklavierten Lösungen wurden bei - 20 °C gelagert.

Ethidiumbromid-Stammlösung:

Die Stammlösung wurde in einer Konzentration von 10 mg/ml in Wasser angesetzt und lichtgeschützt bei 4 °C gelagert.

2.4.6.2 Größenbestimmung von Nukleinsäuren

Um die Größe gelelektrophoretisch aufgetrennter, linearer DNA-Fragmente ermitteln zu können, wurde ein Längenstandard (MBI Fermentas GmbH, St.Leon-Rot) auf das Gel aufgetragen. Die Größe der erhaltenen Fragmente wurde durch den Vergleich der Laufstrecken mit denjenigen der bekannten Fragmente der 1 kb-Leiter bestimmt. In der folgenden Abb. 3 sind die zur Bestimmung herangezogenen Fragmentgrößen angegeben.

DNA-Längenstandard:

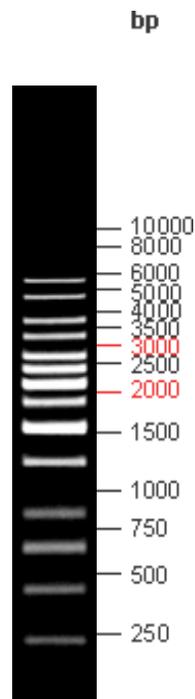


Abb. 3: 1 kb-Leiter auf einem 1 %igen Agarosegel

Zur Auswertung wurden die bekannten Fragmentgrößen halblogarithmisch gegen ihre Wanderungsstrecke aufgetragen. Diese Eichkurve wurde zur Größenbestimmung herangezogen.

2.5 Polymerasekettenreaktion (PCR)

2.5.1 Standard Polymerasekettenreaktion (PCR)

Diese Methode der *in vitro*-Amplifizierung von DNA diente zur Herstellung spezifischer DNA-Fragmente für Klonierungen. Als DNA-Polymerase wurde die Taq-Polymerase (aus *Thermus aquaticus*, Fa. MBI, St.Leon-Rot), die Amplifikate mit 3'-Überhängen synthetisiert, eingesetzt. Die verwendeten Starteroligonukleotide (Primer) hatten eine Länge von 25-28 Basen (Fa. MWG, Ebersberg). Als Matrize diente chromosomale DNA (10 ng) von *Ms. mazei* Gö1 oder eine mit einem Plasmid transformierte *E.coli*-Kultur. Eine optimale Amplifizierung war gewährleistet, wenn beide Primer den gleichen Schmelzpunkt (T_m) hatten und in gleicher Konzentration vorlagen. Die Berechnung des Schmelzpunktes erfolgte nach Chester und Marshak (1993):

$$T_m = 69,3 + 0,41 (\% \text{ G+C-Gehalt}) - 650/L$$

% G+C-Gehalt = prozentualer G+C-Gehalt des Primers

L = Länge des Primers

Die Standard-PCR-Reaktionen fanden in 100 µl-Ansätzen statt. Folgende Komponenten wurden - das Enzym erst zum Schluß - auf Eis zusammenpipettiert:

Matrize (DNA o. flüssig-Kultur)	1,0 µl o. 5µl
Primer 1 (100pmol)	0,5 µl
Primer 2 (100pmol)	0,5 µl
dNTP-Gemisch (100mM)	1,0 µl
10 x Reaktionspuffer	10,0 µl
H ₂ O _{bidest.}	ad 49,5 µl
DNA-Polymerase (5U/µl)	0,5 µl

Nach Durchmischung wurde der Ansatz im PCR-Thermoblock (Fa. MWG, Ebersberg) 5 min bei 95 °C denaturiert. Zur Amplifizierung schlossen sich folgende Temperaturzyklen in mehrfacher Wiederholung an:

Denaturierung	95 °C	30 s
Anlagerung	x °C	30 s
Polymerisierung	72 °C	y s

Die Anlagerungstemperatur (x) war abhängig von der Schmelztemperatur der Primer und lag idealerweise 5 °C tiefer als diese. Die Dauer y der Polymerisierung betrug 1 min pro 1 kb zu amplifizierender DNA. Zur Optimierung der PCR konnten einzelne Parameter variiert werden wie z.B. die Anlagerungstemperatur und die Anzahl der Zyklen oder auch die Konzentrationen der eingesetzten Primer sowie der Matrizen-DNA. Nach Abschluss der Amplifikation wurde ein Aliquot im Agarosegel überprüft und der gesamte Ansatz falls nötig, aufgereinigt (siehe 2.4.5.1.2).

Diese Methode wurde - soweit möglich - für jeden ORF von *Ms. mazei* im 96er Maßstab in Mikrotiterplatten durchgeführt, wobei die erhaltenen PCR Produkte eine Größe von 200 bis 800 Basenpaaren hatten. Die ORF-spezifischen Primer wurden mit einer Schmelztemperatur von 66-68°C generiert.

2.5.2 Kolonien-Screening mittels PCR

Dies ist eine Variante der PCR, in der keine gereinigte DNA als Template eingesetzt wird. Diese Methode wurde angewandt, um *E. coli*-Kolonien, welche ein Plasmid mit dem gewünschten Insert trugen, zu identifizieren. Hierfür wurden die Klone von den Agar-Platten in flüssiges LB-Medium mit Antibiotika überimpft und über Nacht bei 37 °C auf einem Schüttler (150 rpm) angezogen. Als Template in der PCR wurden 5 µl der frischen Kultur eingesetzt. Die PCR wurde mit 25 Zyklen durchgeführt und anschließend auf einem Agarosegel überprüft.

2.5.3 Real-Time-PCR

Bei der Real-Time-PCR handelt es sich um eine Variante der PCR bei der in Echtzeit (=Real-Time) die Konzentration des PCR-Produktes quantitativ gemessen werden kann. Dadurch ist es möglich die Ausgangsmenge des Templates zu errechnen. Der PCR-Ansatz enthält hierbei den Fluoreszenz-Farbstoff SYBR-Green, welcher unspezifisch an doppelsträngige DNA bindet. Der Farbstoff wird während der PCR mit Licht einer spezifischen Wellenlänge angeregt und die entstehende Fluoreszenz mittels eines CCD-Chips gemessen. Die gemessene Fluoreszenz des gebundenen Farbstoffes liegt 200-fach höher als die des freien Farbstoffes.

Als Matrize wurde Gesamt-RNA aus *Ms. mazei* Gö1 verwendet (2.4.5.2), welche zuerst in cDNA umgeschrieben worden war. Diese cDNA diente in der nachfolgenden PCR der hot-start-Taq-Polymerase als Template. Es wurde das ‚Quantitect SYBR Green RT-PCR‘- Kit der Fa. Qiagen verwendet, welches alle benötigten Reagenzien, außer Primern und Template, in vorgefertigten Lösungen enthält. Die für eine Real-Time-PCR benötigten Primer (Fa. MWG, Ebersberg) mußten folgende besondere Bedingungen erfüllen:

- Die Länge der Primer mußte 23-25 Basen betragen.
- Der GC-Gehalt der Primer mußte bei 50% liegen.
- Das zu bildende PCR Produkt sollte 145-150 Basen lang sein.
- Die Primer-Paare durften am 3'-Ende keine Homologie von mehr als einer Base haben.
- Es durften nicht mehr als 3 G/Cs am 3'-Terminus liegen.
- Die Primer durften am 3'-Ende nicht mit einem T beginnen.
- Die Primer durften keine komplementären Sequenzen mit sich selbst oder innerhalb des Primer Paares haben
- Die Sequenzhomologie zum Template mußte 100% betragen

Die Primer wurden mit dem Primer3 Programm (2.9.3) generiert. Alle Schritte der Real-Time-PCR wurden in einem iCycler der Firma Biorad (München) durchgeführt. Zuerst erfolgte eine reverse Transkription innerhalb von 20 Minuten bei 50 °C. Anschließend wurde für 15 Minuten bei 95 °C erhitzt, um die Reverse Transkriptase

zu deaktivieren und die Hot-Start-Taq-Polymerase zu aktivieren. Darauf erfolgte die eigentliche PCR über 40 Zyklen:

Denaturierung:	15s	94°C
Annealing:	30s	55°C
Elongation:	30s	72°C

Die Messung der Fluoreszenzen wurde am Ende der Elongationsschritte durchgeführt. Die Software errechnete den c_t -Wert der einzelnen PCR-Reaktionen automatisch nach Beendigung der PCR. Dieser c_t -Wert ist die errechnete Anzahl der Zyklen nach der eine signifikante Erhöhung des Fluoreszenzsignals vorliegt. Nach dem letzten Zyklus der PCR wurden die entstandenen PCR-Produkte geschmolzen und anhand der Fluoreszenzabnahme eine Schmelzkurve erstellt. Dieser Schritt ermöglichte eine Reinheitskontrolle der entstandenen PCR Produkte.

2.6 Enzymatische Modifizierung von Nukleinsäuren

2.6.1 Herstellung Cy3/Cy5 markierter cDNA Sonden

Die cDNA Sonden für die Microarray-Hybridisierung wurden mit den Farbstoffen Cy3 und Cy5 (Fa. Amersham Biosciences, Freiburg) markiert. Dies erfolgte in Rahmen zweier reversen Transkriptionen bei denen aufgereinigte, DNA freie, Gesamt-RNA aus *Ms. mazei* Gö1 als Template benutzt wurde. Die RNA wurde aus Zellen isoliert, welche entweder mit Methanol oder Acetat als Substrat angezogen wurden. Für jede dieser beiden RNAs wurden je zwei Markierungen parallel mit jeweils einem anderen Farbstoff durchgeführt. Jede einzelne Markierung erfolgte mit dem ‚cDNA first strand labeling kit‘ der Firma Amersham Biosciences nach Angaben des Herstellers mit folgenden Parametern:

- Es wurden 8 µl Gesamt-RNA (min. 5 µg, max. 20 µg) als Matrize eingesetzt.
- Für die „random nonamer“ Primer wurde empirisch eine ideale Menge von 2 µl pro Ansatz ermittelt.

- Zur Aufreinigung der Sonden wurde das GFX Protokoll des Kits benutzt.
- Es wurde zu dem Template am Anfang jeweils 1 µl Spike-Mix der Universal Scorecard (Amersham Biosciences, Freiburg) gegeben.
- Vor der Aufreinigung mit dem GFX Protokoll wurden die Sonden der Acetat-cDNA mit den anders markierten Sonden der Methanol-cDNA vereinigt.

Nach der Aufreinigung wurden die Sonden sofort für die Hybridisierung verwendet und nicht gelagert.

2.7 Übertragung von PCR-Produkten in *E. coli* durch TOPO-Transformation

2.7.1 Herstellung kompetenter Zellen zur Transformation (Inoue et al., 1990)

250 ml SOB-Medium wurden mit einer über Nacht gewachsenen *E. coli*-Vorkultur 3 %ig angeimpft und bei 18 °C im Schüttelwasserbad bei 150 rpm bis zu einer OD_{600nm} von 0,6 angezogen. Die Zellen wurden dann 10 min im Eisbad geschwenkt und im vorgekühlten GSA-Rotor abzentrifugiert (4000 Rpm, 4 °C, 10 min in Sorvall Zentrifuge RC-5B, Fa. Du Pont Instruments, Bad Homburg). Das Zellpellet wurde vorsichtig in 80 ml kalten Transformationspuffer (TB) resuspendiert, für 10 min auf Eis inkubiert und unter gleichen Bedingungen erneut abzentrifugiert. Nach schonender Suspendierung in 20 ml TB wurden 1,5 ml DMSO (7 % [v/v] Endkonzentration) langsam unter Schwenken zugesetzt und für 10 min auf Eis inkubiert. Aliquots von 200 µl wurden in vorgekühlte E-Cups transferiert und in flüssigem Stickstoff schockgefroren. Die Lagerung der Zellen erfolgte bei -70 °C über mehrere Monate ohne Verlust der Kompetenz.

Transformationspuffer (TB):

PIPES	10 mM
CaCl ₂	15 mM
KCl	250 mM
MnCl ₂	55 mM
pH 6,7	

Alle Komponenten mit Ausnahme des Mangansalzes wurden zusammengegeben, der pH-Wert mit KOH eingestellt, eine Stunde gerührt und nachjustiert. Dann wurde MnCl₂ zugesetzt, die Lösung sterilfiltriert und bei 4 °C gelagert.

2.7.2 TOPO-TA Klonierung

Die Methode der TOPO-Klonierung ermöglicht es, in einem Schritt ein PCR Produkt in einen Vektor zu klonieren und diesen Vektor in *E. coli* zu transformieren. Für diese Methode muß das PCR Produkt von einer Polymerase erzeugt worden sein, welche T/A-Überhänge an den Enden des PCR Produktes produziert (z.B. Taq-Polymerase). Der in dem Kit enthaltene Vektor hat ebenfalls diese Überhänge und ist an eine TOPO-Isomerase gekoppelt. Dieses Enzym ligiert das PCR Produkt in den Vektor. Es wurde das TOPO-Klonierungs-Kit der Firma Invitrogen, Karlsruhe benutzt. Der Vektor des Kits wurde verdünnt eingesetzt (1:2) und es wurden selbst hergestellte kompetente DH5 α Zellen (2.7.1) verwendet. Ansonsten wurde nach den Angaben des Herstellers verfahren.

Die erhaltenen *E. coli* Transformanten wurden wie in 2.3.2.3 beschrieben ausplattiert und die erhaltenen weißen Klone durch Kolonie-Screening-PCR (2.5.1) auf ihr Insert überprüft. Positive Klone wurden anschließend mit DMSO (10% Endkonzentration) in 96-Loch TPP-Platten (Fa. Schütt, Göttingen) bei -70°C gelagert (Gesamtvolumen 150 μ l).

2.8 Herstellung und Anwendung von cDNA Microarray-Chips

2.8.1 Übertragung von PCR Produkten auf Aminosilan beschichtete Glas Slides

10 µl der im 96er Maßstab aufgereinigten PCR Produkte wurden in 384-Loch-Platten umpipettiert und 1:1 mit DMSO verdünnt. Diese 384er Platten wurden mit einem Spotting-Roboter (Lucidea Array Spotter Generation V, Amersham Biosciences, Freiburg) auf mit Aminosilan beschichtete ‚Type7* Microarray-Slides‘, (Amersham Biosciences, Freiburg) gedruckt.

Das Drucksystem dieses Roboters ist ein sogenanntes ‚Kontakt-Druck‘-System. Die 24 Nadeln des Druckkopfes enthalten Mikrokapillaren, welche ca. 200 nl PCR-Lösung aufnehmen und (ähnlich wie ein Füllfederhalter) die aufgenommene Flüssigkeit durch Aufsetzen auf der Chipoberfläche abdrucken. Dieser Vorgang erfolgte über Nacht in einem staubfreien Raum bei einer konstanten 65%igen Luftfeuchtigkeit. Um die aufgedruckten PCR Produkte an die Oberfläche zu binden, wurden die Slides mit UV-Licht (50 mJ/cm²) bestrahlt. Die bedruckten und fixierten Slides wurden bis zur Hybridisierung im Dunkeln bei Raumtemperatur gelagert.

2.8.2 Hybridisierung von DNA-Microarray Chips

Die Fluoreszenz-markierten cDNA Sonden (2.6.1) wurden in einem ‚ASP‘ (Automatic Sample Processor, Fa. Amersham Biosciences, Freiburg) über Nacht mit DNA-Microarray-Chips (2.8.1) hybridisiert. 50 µl des cDNA-Sonden Gemisches wurden kurz vor der Hybridisierung für 2 Minuten bei 95 °C aufgeköcht und auf Eis abgekühlt. Anschließend wurden die Sonden mit 100 µl Formamid (100%) und 50 µl 4fach-Hybridisierungspuffer (First Strand cDNA labeling Kit, Amersham Biosciences, Freiburg) verdünnt. Die erhaltenen 200 µl wurden mit einer Hamilton Spritze in die vorbereitete Hybridisierungsstation (=ASP) injiziert.

Hybridisierungs-Protokoll des ASP:

Flush: Spülen der Hybridisierungskammern und Schläuche

Prime : Füllen der Schläuche bis zur Hybridisierungskammer

Mix: Mixen der Lösung innerhalb der Kammer durch Vor- und Zurückpumpen

Heat: Heizen der Kammer auf eine bestimmte Temperatur

Air Pump: Trocknen der Kammer durch Luft

Wait: Warten bis zu einer Bestätigung vom User

Lösung 1: Wasser (bidest) (Min. Volumen, 200 ml)

Lösung 2: Wash 3 (0.1 x SSC)

Lösung 3: Wash 1 (1 x SSC/0.2% SDS)

Lösung 4: Wash 2 (0.1 x SSC/0.2% SDS)

Lösung 5: Isopropanol

"Wait", "Check dummy slides", "Überprüfen, ob in jeder Kammer ein dummy Slide ist"

"Flush", "Leeren der Schlauche und Kammern von alten Lösungen"

"Heat", "Kammern auf 38 °C aufheizen"

"Flush", "Kammern mit Isopropanol spülen und leeren"

"Flush", "Kammern mit Wasser spülen und leeren"

"AirPump", "Kammern mit Luft spülen für 60 s"

"Heat", "Switch off, Heizung ausschalten"

"Prime", "Schläuche mit Wash1 (Lösung 3) füllen"

"Wait", "INSERT TEST SLIDES", "test slides einlegen"

"Heat", "Kammern auf 42 °C aufheizen"

"Wait", "Inject probe solution", "Proben Lösung injizieren"

"Mix", "Mischen der Lösungen in den Kammern für 12 h"

"Flush", "Mit Lösung Wash 1 spülen"

"Heat", "Kammern auf 45 °C aufheizen"

"Mix", "Mischen mit Wash 1 für 10 min"

"Prime", "Schläuche mit Wash 2 (Lösung 4) füllen"

"Flush", "Mit Wash 2 spülen"

"Mix", "Mischen mit Wash 2 für 10 min"

"Flush", "Mit Lösung Wash 2 spülen"

"Mix", "Mischen mit Wash 2 für 4 min"

"Prime", "Schläuche mit Wash 3 (Lösung 2) füllen"

"Flush", "Mit Lösung Wash 3 spülen"

"Prime", "Schläuche mit Isopropanol füllen"

"Flush", "Mit Isopropanol spülen"

"Flush", "Kammern mit Luft spülen"

"AirPump", "Kammern mit Luft trocknen"

"Heat", "Heizung ausschalten"

"Wait", "REMOVE TEST SLIDES", "Test slides entfernen und mit dummy slides ersetzen"

"Heat", "Heizen auf 48 °C"

"Flush", "Mit Lösung Wash 2 spülen"

"Mix", "Mischen mit Wash 2 für 5 min"

"Prime", "Schläuche mit Wasser füllen"

"Flush", "Mit 6 ml Wasser spülen"

"Mix", "Mischen mit Wasser für 5 min"

"Flush", "Mit 6 ml Wasser spülen"

"Flush", "Mit Luft spülen"

"Prime", "Schläuche mit Isopropanol füllen"

"Flush", "Mit Isopropanol spülen"

"Flush", "Mit Luft spülen"

"AirPump", "Kammern mit Luft trocknen"

[Ende]

2.8.3 Einlesen der DNA-Microarray Chips

Nach der Hybridisierung (2.8.2) wurden die DNA-Chips mit einem Chip-Reader der Firma Axon (Union City, Kalifornien USA, <http://www.axon.com>) eingelesen. Dieser Scanner enthält zwei Laser mit spezifischen Wellenlängen (565 nm und 635 nm) für die verwendeten Fluoreszenzfarbstoffe Cy3 und Cy5 (Amersham Biosciences, Freiburg) mit einer Auflösung bis zu 5 µm. Die durch die Laser-Anregung entstandene Fluoreszenz wurde mittels eines Photomultipliers im Scanner gemessen. Aus den Meßdaten erzeugte die Betriebssoftware des Scanners (siehe 2.9.6) zwei Bilder des Chips, für jeden Fluoreszenzkanal eins. Nach dem Scannen wurden die Chips bei Raumtemperatur unter Lichtausschluß gelagert.

2.9 Bioinformatische Materialien und Methoden

Die Generierung und Auswertung eines Genom-umfassenden DNA-Microarray-Chips erforderte die Handhabung von großen Datenmengen. Dies machte das Gebiet der Bioinformatik zu einem essentiellen Bestandteil dieser Arbeit. Für viele Schritte gab es keine verfügbare Software, weshalb diese vom Autor dieser Arbeit eigenständig programmiert wurde.

Soweit nicht anders angegeben, wurden die bioinformatischen Schritte an einem PC mit 2 Gigabyte RAM, einem 2,5 Ghz Intel Pentium 4 Prozessor und Windows 2000 (Microsoft Corp., Seattle) als Betriebssystem, entwickelt und durchgeführt.

2.9.1 Standard Software

2.9.1.1 Microsoft Produkte

Es wurden folgende Software Programme der Microsoft Corporation in Seattle verwendet:

Microsoft Windows 2000: Als Betriebssystem auf den Computern welche in dieser Arbeit verwendet wurden, soweit nichts anderes angegeben ist.

Microsoft Word: Zur Erstellung von tabellarischen Schemata zur Handhabung und Protokollierung der 96er Platten und deren Belegung.

Microsoft Excel : Zur Bearbeitung von Tabellen.

In der Regel waren dies Tabellen, welche als tab-getrennte .txt Files von anderen Programmen generiert wurden.

2.9.1.2 Ultraedit 32

Eine Shareware Version des Programms Ultraedit32 (www.tucows.de) wurde als Editor zum erstellen der Programme und Skripte benutzt. Ferner wurde seine FTP (File-Transfer-Protokoll) Unterstützung für ‚remote-access‘ Programmierung auf den Großrechnern des Institutes für Mikrobiologie und Genetik und der GWDG eingesetzt.

2.9.1.3 Linux Redhat 2.1

Dieses Betriebssystem steuert den Großrechner des Institutes für Mikrobiologie und Genetik. Der Rechner ist ein Cluster aus 4 Alpha Prozessoren. Die CPU Zeit wird über ein ‚batch system‘ dynamisch verteilt. Dieser Server wurde für selbst geschriebenen Programmen benutzt, welche eine hohe Rechenleistung erforderten.

2.9.1.4 Solaris 5

Dieser Dual-Prozessor Server der SUN Corporation enthält die Instituts interne Datenbank der ERGO-Data-Suite (Integrated Genomics, Chicago). Selbstgeschriebene Programme, die auf die ERGO-Datenbank manuell zugegriffen, benutzen diesen Server.

2.9.2 Implementation von Programmiersprachen in das Betriebssystem

Windows 2000 bietet keine Unterstützung für Java und PERL. Die Möglichkeit des Betriebssystems die beiden Programmiersprachen zu verstehen ist allerdings unabdingbar gewesen, weshalb folgende Software installiert wurde.

2.9.2.1 ActiveState PERL

Die Implementation von PERL in das Windows 2000 Betriebssystem wurde durch eine freie PERL 5.0 Distribution von der Firma ActiveState durchgeführt (<http://www.activestate.com>). Die Programme, im Rahmen dieser Arbeit erstellt wurden, sind in PERL geschrieben.

2.9.2.2 Java 2 Runtime Environment und die Java Virtual Machine

Die Java Unterstützung wurde durch freie Software der SUN Corporation (<http://www.sun.com>) gewährleistet. Ein ‚Java 2 Runtime Environment‘ und eine ‚Java Virtual Machine‘ waren notwendig, um die ‚Analyst Expressionist Software‘ (Fa. Genedata, Wien) zu benutzen.

2.9.3 Primer3 und der Server der UC-Irvine

Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) ist ein Programm, welches es ermöglicht, Primer mit bestimmten Spezifikationen zu generieren. Primer3 unterliegt einer ‚open-source‘ Distribution des Programmierers. Die University of California, Irvine stellt ein Web-Interface (<http://promoter.ics.uci.edu/Primers/ORF.htm>) zur Verfügung, welches erlaubt, mehrere Primer auf einmal zu generieren. Dies erfolgt über Primer3 Routinen. Beide Programme wurden benutzt, um die Primer zu generieren, die für die Projekte dieser Arbeit benötigt wurden.

2.9.4 BLAST-Server

Der BLAST-(Biological Local Alignment Search Tool)-Algorithmus ist der zur Zeit am häufigsten verwendete Algorithmus um Sequenzhomologien zu finden. Der BLAST- Algorithmus ist für eine hohe Durchsatzmenge an Daten und hohe Geschwindigkeit optimiert. Die NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) stellt eine freie Distribution zur Verfügung, welche von einigen selbst geschriebenen Programmen in dieser Arbeit benutzt wurde.

2.9.5 ERGO Data suite

Die ‚ERGO Data Suite‘ ist ein kommerziell erhältliches Server-basiertes Datenbank- und Annotationsprogramm der Firma Integrated Genomics (Chicago, USA). Der ERGO Server des Instituts für Mikrobiologie und Genetik liegt auf dem unter 2.9.1.4 beschriebenen Rechner. Die Ansteuerung dieser Software erfolgte mit dem Microsoft Internet Explorer oder Netscape 4.7. Die Datenbank enthielt zum Zeitpunkt dieser Arbeit 394 Genome, darunter das Genom von *Ms. mazei* Gö1.

Die ‚ERGO Data Suite‘ wurde zur Annotation des Genoms von *Ms. mazei* Gö1 verwendet. Diese Annotationsdaten waren die Grundlage für die Auswertung und Generierung der in dieser Arbeit hergestellten DNA-Microarrays.

2.9.6 GenePix-Pro

GenePix-Pro ist eine kommerziell erhältliche Software der Firma Axon (Union City, Kalifornien USA, <http://www.axon.com>), die den Microarray-Scanner (2.8.3) steuert und für die Vorauswertungen und Normalisierungen der Expressionsdaten verwendet wurde. Die Normalisierungs- und Auswertungs-Protokolle dieser Software sind jedoch nicht sehr umfangreich und lassen sich nur auf DNA-Chips anwenden, bei denen keine starke Änderungen der Gen-Expressionen vorliegt. Die in dieser Arbeit verwendeten Chips wurden deshalb mit der Analyst Expressionist Software der Firma Genedata (Wien) ausgewertet.

2.9.7 Analyst Expressionist

Der Analyst Expressionist ist Bestandteil des kommerziell erhältlichen Analyst Software Paketes (Fa. Genedata, Wien). Die Expressionist Software lädt Expressionsdaten aus Files oder aus einer Oracle Datenbank (= CoBi, Bestandteil des Analyst Paketes). In dieser Arbeit wurden die Files mit den Expressionsdaten direkt in den Analyst geladen. Die Software führt Normalisierungen der Expressionsdaten über mehrere Experimente aus und stellt diverse Möglichkeiten für Korrelationsberechnungen und Gruppierungs-Algorithmen (Cluster-Algorithmen) der einzelnen Expressionsdaten zur Verfügung.

2.9.10. Selbst erstellte Software

Sämtliche Software, die im Rahmen dieser Arbeit selbst programmiert wurde, ist mit offenem Quellcode in dieser Arbeit dokumentiert. Als Programmiersprache wurde ausschließlich PERL verwendet. Zur Implementation dieser Programme in ein Server basiertes System wurde das Common-Gateway-Interface (CGI) benutzt.

2.10 Chemikalien, Biochemikalien und Gase

Standardchemikalien wurden, soweit nicht anders im Text angegeben, mit dem Reinheitsgrad „pro analysis“ oder „reinst“ von den Firmen Boehringer (Mannheim), Fluka (Buchs, Schweiz), Merck (Darmstadt), Riedel de Haen (Hannover), Serva (Heidelberg) und Sigma (Taufkirchen) verwendet. Die Bezugsquellen weiterer Chemikalien, Enzyme, Verbrauchsmaterialien und Kits sind nachfolgend aufgeführt:

ABgene, Hamburg

Verschlussfolien für *E. coli* Kulturen in 96-Loch Platten

Amersham Biosciences, Freiburg:

„First Strand cDNA labeling“-Kit, Type7*-Microarray-Slides, Cy3 und Cy5 Fluoreszenz Farbstoffe, Random Nonamer Primer, „Lucidea Universal Score Card“ - Kit

Biomol, Hamburg

IPTG, Lysozym, Phenol (äquilibriert mit Tris-Puffer), X-Gal

Biorad, München

PCR-Platten für die Real-Time PCR, optisch durchlässige Verschlussfolien für die Real-Time PCR

Boehringer Mannheim GmbH, Mannheim

DNaseI,

Difco Laboratories, Detroit, USA

Agar

Gerbu Trading GmbH, Gaiberg

Ampicillin, Kanamycin

Gibco-BRL, Eggenstein

Hefeextrakt, Trypton

Invitrogen, Karlsruhe:

„TOPO-TA-Cloning-Kit“ Version C and J

MBI, St.Leon-Rot:

DNA-Längenstandards, Taq-Polymerase

Millipore, Eschborn:

Sterilfilter

Oxoid, Basingstone, England:

Agar, Trypton, Hefeextrakt

Peq-Lab, Erlangen

Agarose

Promega Corporation, Madison, USA

„Wizard DNA Purification Kit“

Qiagen GmbH, Hilden:

„Qiaquick PCR Purification“-Kit, „Quantitect SYBR Green RT-PCR“-Kit

Sartorius AG, Göttingen:

Sterilfilter

Serva, Heidelberg

EDTA, Tris

Alle Gase sowie flüssiger Stickstoff wurden von der Fa. Messer Griesheim, Kassel, bezogen. Die Gase waren von folgender Reinheit:

N₂: 99,99 %

N₂/CO₂-Gemisch (80 / 20), Reinheit nicht angegeben

H₂: 99,9 %

3. Experimente und Ergebnisse

Die Grund legende Fragestellung dieser Arbeit war die Untersuchung der Regulation des Kohlenstoff- und Energiestoffwechsels von *Methanosarcina mazei* Gö1. Dieser Organismus ist im Stande, H₂+CO₂, Methanol, Methylamine und Acetat als Substrate zu verwerten. Abhängig von der Kohlenstoffquelle benutzt der Stamm Gö1 den CO₂-reduzierenden, den methylo trophen oder den aceticlastischen Weg der Methanogenese. Basierend auf der Genomsequenz sollte untersucht werden, welche Gene in den jeweiligen Wegen der Methanogenese aktiv sind, wie diese Gene reguliert werden und in welche regulatorischen Netzwerken sie eingebunden sind.

Durch die Experimente dieser Arbeit sollten essenzielle Grundlagen für das Verständnis der Genexpression und der Genregulation in Archaeen und speziell in *Ms. mazei* Gö1 geschaffen werden. Die Methode der Wahl hierfür war in dieser Arbeit die genomweite Transkriptionsanalyse aller Gene beim Wachstum mit verschiedenen Kohlenstoffquellen durch DNA-Microarrays. Hierbei wurden Gene klassifiziert, die einen Unterschied in der Stärke ihrer Expression unter den jeweiligen Wachstumsbedingungen der Zellen zeigten. Ferner wurden spezielle Gene einer absoluten Transkriptions-Analyse durch Real-Time-PCR unterzogen. Diese Experimente sollten zum Verständnis der zellulären Prozesse, welche direkt und indirekt mit der Methanogenese verbunden sind, beitragen.

3.1 Entwicklung von bioinformatischen Programmen zur Unterstützung der Annotation der Gene aus *Methanosarcina mazei* Gö1

Zu Beginn des experimentellen Teiles dieser Arbeit war das Genom-Sequenzierungs-Projekt von *Ms. mazei* der Göttinger Genomics Laboratories (G₂L) in der letzten Phase. Die Anzahl der DNA-Contigs des Genoms betrug ca. 60 und die Identifizierung der ORFs auf den Contigs wurde gerade durchgeführt. Es kam schließlich heraus, dass das Chromosom ca. 3300 offene Leserrahmen enthält. Diese Zahl überstieg die vorhergesagten Gene in anderen sequenzierten Archaeen um mehr als den Faktor zwei. Erste Annotationsvorarbeiten zeigten zudem, dass die

abgeleiteten Aminosäuresequenzen zahlreicher Gene hoch homolog zu bakteriellen Proteinen waren. Diese Erkenntnis ließ darauf schließen, dass es im Zuge der Evolution von *Ms. mazei* zu zahlreichen lateralen Gentransfer-Ereignissen gekommen war. Um nachzuweisen, dass chromosomale Gene von *Ms. mazei* ihren Ursprung in einem anderen Organismus hatten, wurde ein Abgleich der DNA-Sequenzen aller bis dahin definierten ORFs gegen eine biologische Datenbank ausgeführt. Dies erfolgte auf bioinformatischer Ebene durch PERL-Programme, welche im Rahmen dieser Arbeit programmiert wurden.

Es wurden alle Gene von *Ms. mazei* und deren abgeleitete Aminosäuren mit einer Datenbank der GWDG (Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen) verglichen. Diese Datenbank wurde von Hr. Dr. Rainer Bohrer zur Verfügung gestellt und enthielt die Gensequenzen aus allen öffentlichen Datenbanken (z.B. NCBI, Swissprot, etc.). Die Datenbank wurde täglich aktualisiert und unmittelbar vor einem Abgleich in die BLAST-Datenbank des bearbeitenden Computers integriert. Die Sequenzen aller ORFs von *Ms. mazei* Gö1 wurden benötigt, um einen Abgleich gegen die GWDG-Datenbank durchzuführen. Diese Sequenzen waren in der Datenbank der ERGO-Data-Suite gespeichert. Die ERGO-Data-Suite (siehe 2.9.5) war zu diesem Zeitpunkt noch in der Beta-Testphase und bot keine Möglichkeit, die Sequenzen aller ORFs eines Organismus auf einmal auszugeben. Aus diesem Grund mussten PERL-Skripte programmiert werden, um die Daten aus der ERGO-Datenbank zu extrahieren.

3.1.1 Extraktion der Sequenzdaten aller *Ms. mazei* Gö1 Gene aus der ERGO-Data-Suite

Da die ERGO-Data-Suite die Sequenzdaten in ‚Flat-files‘ abspeicherte, für deren Decodierung keine Schnittstellen definiert waren, musste das Skript einen Umweg über das Hyper-terminal-transfer-Protokoll (http) gehen und die Daten im html-Format abfragen. Der Quell-Code dieses Skriptes ist in Anhang-A Skript-1 dokumentiert und kommentiert. Die erhaltenen Daten lagen im html-Format vor (Abb. 4).

```

<TITLE>Checked AA Sequences</TITLE>
<HTML>
<BODY>

<TABLE BORDER=0 BGCOLOR="#000000" WIDTH=100% CELLPADDING=2
CELLSPACING=0>

<TR><TD WIDTH="100%">
  <TABLE BGCOLOR="#CCBB22" BORDER=0 WIDTH=100% CELLSPACING=0
  CELLPADDING=5>
    <TR>
      <TD WIDTH="40%" ALIGN="left" NOWRAP>
        <FONT FACE="Arial,Helvetica" COLOR="#FFFFFF">
          <STRONG><BIG><BIG><BIG>&nbsp;&nbsp;&nbsp;ERGO&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;</BIG></BIG>AA
          Sequences</BIG></STRONG>
          </FONT></TD>
      <TD ALIGN="right" NOWRAP>
        <FONT FACE="Arial,Helvetica" COLOR="#FFFFFF">
          <STRONG><BIG>Checked Set</BIG></STRONG></FONT></TD>
    </TR>
  </TABLE>
</TD>
</TR>
</TABLE>
<pre>
>RMMZ00004 [Methanosarcina mazei] gap_closure.21.34469_3380960_3380316
ANTHRANILATE SYNTHASE COMPONENT II (EC 4.1.3.27)
MKQKTSKSSRRDFQMKIVFINNKDSFVWNLVDYISYFEKDTLVLPNTVTLEELKEIKP
DALVISPGPGNPSDPRDIGNCLEIIREMGREIPLLGVCLGHQAINVAFGGPVRRCVKVGPV
HGKSSRIRHAESALFTTLEEKFEAGRYHSLEIGDPAPGIKVTALAEDGTIMAVEHSEFPI
YGLQFHPESVLTDPGLKIIERFLEISRDFKKRQPAV
</pre>
</BODY>
</HTML>

```

Abb. 4: Beispielausgabe von Aminosäure-Sequenzen im html-Format

Die im html-Code eingebetteten Daten konnten nicht direkt für einen Sequenzabgleich verwendet werden, sondern mussten in ein multi-FASTA-Format konvertiert werden. Dies erfolgte über das in Anhang-A Skript-2 dokumentierte PERL-Skript, welches ein multi-FASTA-File generiert. Es wurden die DNA- und die Proteinsequenzen aller ORFs von *Ms. mazei* Gö1 extrahiert. Ein Beispiel eines multi-FASTA-Files mit Proteindaten ist in Abb. 5 dargestellt.

```
>RMMZ00001          Contig08396_240_713
MTFHIWTAYSPFTTWVQIVKDWMKTKGDTGKRKTFVNTTLGETWEA
KIGERPDAEVMAERKEHYSAPVPDRVAYLTAGIDSQLDRYEMRVWG
WGPGEESWLIDRQIIMGRHDDEALLHSVPGDPTGIPRGVKCISMQLHQ
CHDRDQFRRIQNHGIRLR
>RMMZ00002          Contig08396_1987_965
MSMYTTAQLLAANEQKFKFDPLFLRLFFRESYPFTTEKVYLSQIPGLVN
MALYVSPIVSGEVIRSRGGSTSEFTPGYVVKPKHEVNPQMTLRRLPDEDP
QNLADPAYRRRRRIIMQNMREELAIQAQEEMQAVSAVLKKGKYTMTGE
AFDPVEVDMGRSEENNITQSGGTEWSKRDKSTYDPTDDIEAYALNASG
VVNIIVFDPKGWALFRSFKAVKEKLDTRRGSNSELETAVKDLGKAVSY
KGTYGDVAIVVYSGQYVENGVKKNFLPDNTMVLGNTQARGLRITYGCI
QDADAQREGINASARYPKNWVTTGDPAREFTMIQSAPLMLLADPDEF
VSVQLA
```

Abb. 5: Beispiel eines multiFASTA-Files mit Proteindaten

Ein solches multi-FASTA-File konnte entweder Protein- oder DNA- Sequenzen enthalten. Dieses File-Format wurde von den BLAST-Programmen der NCBI unterstützt.

3.1.2 Sequenzabgleich zur Identifizierung von Bakterien-ähnlichen Genen

Die Sequenzen der ORFs des generierten multi-FASTA-Files dienten als ‚queries‘ für einen BLAST-Abgleich gegen die Datenbank der GWDG. Dies wurde von einem PERL-Programm ausgeführt, dessen Quell-Code im Anhang-A Skript-3 dokumentiert und kommentiert ist. Dieses Programm glich jeden ORF aus dem multi-FASTA-file durch den BLAST-Algorithmus gegen die Datenbank ab. Die Ausgabe des Programms war wahlweise eine Tabelle mit den höchsten Homologien oder nur mit den besten annotierten Homologien. Die Tabelle wurde von dem Skript als tab-getrenntes .txt File abgespeichert und konnte in dem Excel-Programm weiterverarbeitet werden. Abb. 6 stellt eine Beispielausgabe eines Teils eines erzeugten .txt Files dar:

```
>RMMZ00057 Contig29629_62389_60626ref|NP_069604.1| signal-transducing histidine kinase [Archae... 75 2e-012
>RMMZ00066 Contig15945_126_1256gb|AAG18694.1| (AE004974) Vng0056h [Halobacterium sp. ... 105 1e-021
>RMMZ00068 Contig15945_3015_3542pir|E71028 probable transcription regulator - Pyrococcus horiko... 50 1e-005
>RMMZ00069 Contig15945_3567_4022ref|NP_071274.1| heme biosynthesis protein (nirH) [Archaeoglo... 65 5e-010
```

Abb. 6: Beispielausgabe eines Teils des erzeugten .txt files

Die im Rahmen der Vorversuche erstellten Programme erzeugten die gewünschten Daten einwandfrei. Einige Programme wurden daher später in anderen Projekten und Arbeitsgruppen ebenfalls eingesetzt (z.B. im *Gluconobacter oxydans* Sequenzierungs-Projekt des Kompetenz-Netzwerkes Göttingen).

Wie erwartet, stellte sich heraus, dass ca. 30 % der Proteine von *Ms. mazei* Gö1 ihre höchsten Homologien zu Proteinen von Bakterien hatten. Unter diesen ca. 1000 ORFs befanden sich etwa 500, die kein Gegenstück in irgendeinem vollständig sequenzierten Archaeon aufwiesen und nur Ähnlichkeiten zu bakteriellen Proteinen zeigten. Eine Auswahl dieser Proteine und deren Funktion ist in Tabelle 5 dargestellt.

Tabelle 5: Bakterien-ähnliche Proteine und deren Funktion in dem methanogenen Archaeon *Ms. mazei*:

ORF-Nr.	Annotation	Nächster Verwandter	Funktion des Proteins
1133	Serine-Hydroxymethyltransferase	<i>Thermotoga maritima</i>	Aminosäure-Biosynthese
1266	Glutamin-Synthetase	<i>Clostridium acetobutylicum</i>	Aminosäure-Biosynthese
2367	Acetolactat-Decarboxylase	<i>Bacillus subtilis</i>	Aminosäure-Biosynthese
0125	GroES	<i>Thermotoga maritima</i>	Bakterielles Chaperon
0126	GroEL	<i>Bacillus stearothermophilus</i>	Bakterielles Chaperon
0100	Pyruvate, Orthophosphate Dikinase	<i>Thermotoga maritima</i>	Gluconeogenese
0484	Fructose-1,6-Bisphosphatase	<i>Synechocystis sp.</i>	Gluconeogenese
2283	Phosphoribosylaminoimidazole-carboxamide-Formyltransferase	<i>Bacillus</i>	Purin-Synthese
0472	F ₄₂₀ -abhängige Glucose-6-Phosphat-Dehydrogenase	<i>Mycobacterium phlei</i>	Zuckermetabolismus
0984	dTDP-6-Deoxy-D-Glucose-3,5 Epimerase	<i>Bacteroides fragilis</i>	Zellwandsynthese
2087	Dolichol-Phosphat-Mannosyltransferase	<i>Aquifex aeolicus</i>	Zellwandsynthese
2354	Mannosyltransferase B	<i>Synechocystis sp.</i>	Zellwandsynthese
0447	Thioredoxin-Reduktase	<i>Thermotoga maritima</i>	Zentralmetabolismus
2228	Succinate-Semialdehyde-Dehydrogenase	<i>Bacillus thermoleovorans</i>	Zentralmetabolismus
1771	Gluconat-Permease	<i>Bacillus halodurans</i>	Transport
0316	Type I Restriktionsenzym EcoKI M	<i>E. coli K12</i>	DNA-Modifizierung
0328	DNA-Mismatch-Reparatursystem Protein	<i>Staphylococcus aureus</i>	DNA-Modifizierung
0941	ATP-abhängige DNA helicase	<i>Escherichia coli O157:H7</i>	DNA-Modifizierung
1496	Endonuclease R	<i>Escherichia coli</i>	DNA-Modifizierung
2016	Type I Restriktionsenzym	<i>Helicobacter pylori</i>	DNA-Modifizierung
2704	Natrium/Alanine-Symporter	<i>Vibrio cholerae</i>	DNA-Modifizierung
0825	Hydrogenase-4 Membranuntereinheit	<i>Escherichia coli</i>	Elektronentransport
1011	Flavoredoxin	<i>Desulfovibrio gigas</i>	Elektronentransport
2145	Aminoacyl-Histidin-Dipeptidase	<i>Vibrio cholerae</i>	Peptidmetabolismus
0039	Membran-Alanyl-Aminopeptidase	<i>Zymomonas mobilis</i>	Peptidmetabolismus
0304	Serin/Threonine-Protein-Phosphatase	<i>Campylobacter jejuni</i>	Regulation
0760	Serin/Threonin-Protein-Kinase	<i>Deinococcus radiodurans</i>	Regulation
1251	Sensor-Histidin-Kinase	<i>Mesorhizobium loti</i>	Regulation
1759	Transkriptionsregulator	<i>Bacillus halodurans</i>	Regulation
1953	Response-Regulator	<i>Mesorhizobium loti</i>	Regulation
1381	Katalase	<i>Bacillus halodurans</i>	Stress-Antwort
1383	Oxidoreduktase, Aldo/Keto-Reduktase Familie	<i>Thermotoga maritima</i>	Zuckermetabolismus
1481	Oxidoreduktase, Aldo/Keto-Reduktase Familie	<i>Thermotoga maritima</i>	Zuckermetabolismus
1237	L-Lysin-2,3-Aminomutase	<i>Clostridium subterminale</i>	Stress-Antwort
2814	H ⁺ -translozierende Pyrophosphatase	<i>Rhodospirillum rubrum</i>	Transport

Bakterien-ähnliche Proteine von *Ms. mazei* sind den folgenden zellulären Funktionen involviert: 1) Aminosäure und Purinbiosynthese 2) Anaplerotische Reaktionen und Synthese von Aminosäure Vorstufen. 3) Tetrahydrofolat-Metabolismus. 4) Synthese und Verwertung von Speicherstoffen (Polysaccharide, Polyphosphate). 5) Transporter (ABC Transporter und diverse Permeasen). 6) Sensoren und Regulatoren der Transkription. (Über 20 Gene kodieren regulatorische Zwei-Komponenten-Systeme und andere Transkriptionsregulatoren, die sonst nur in Bakterien vorkommen; viele hiervon haben ihre nächsten Verwandten in der Familie der Cyanobakterien (Deppenmeier et al., 2002)). 7) RNA/DNA Modifikation und Reparatur. 8) Zellwand Biosynthese. 9) Stress-Antwort. 10) Energiekonservierung (herausragende Beispiele sind die $F_{420}H_2$ -Dehydrogenase (Baeumer et al., 2000) und die Ech-Hydrogenase (Meuer et al., 2003), die eng verwandt mit dem Komplex I der bakteriellen und mitochondrialen Atmungskette sind und die F_{420} -nichtreduzierende Hydrogenase (Deppenmeier et al., 1995), die hoch homolog zu bakteriellen Membran gebundenen Hydrogenasen ist) .

Ferner besitzt *Ms. mazei* neben den archaeellen Thermosomen auch das bakterielle Chaperonin System (GroEL/ES), das in anderen Archaea mit Ausnahme vom *Ms. acetivorans* nicht vorkommt (Klunker et al., 2003; Galagan et al., 2002). Das GroEL/ES System ist offenbar notwendig, um die hohe Anzahl an bakteriellen Proteinen korrekt zu falten. Zusammen mit Genen die für Mobilität, Chemotaxis und Transkriptionsregulation kodieren, eröffnen diese Gene dem Organismus unerwartete Möglichkeiten, sich an die unterschiedlichen metabolischen und äußeren Einflüsse anzupassen.

3.2 Genomweite Transkriptionsanalyse von *Ms. mazei*

Die aus den Vorversuchen (und den daraus weitergehenden Experimenten) erhaltenen Daten legten nahe, dass viele Gene auf dem Chromosom von *Ms. mazei* ihren Ursprung außerhalb der Gruppe der Archaea haben. Es blieb jedoch die Frage offen, ob diese Gene auch tatsächlich exprimiert werden.

Um diese Frage zu untersuchen und um einen Einblick in die regulatorischen Netzwerke von *Ms. mazei* zu gewinnen, wurde ein DNA-Microarray-Chip generiert, welcher es ermöglichte, die Expression fast aller 3371 Gene des Organismus in einem Experiment zu analysieren. Die im Rahmen dieser Arbeit etablierte Methode erlaubte es, Unterschiede in der Transkription des Genoms zwischen zwei unterschiedlichen Wachstums-Bedingungen zu untersuchen.

3.2.1 Erstellung einer PCR-Produkt-Genbank des *Ms. mazei*-Genoms

Erstes Ziel der Experimente war es, eine Genbank anzulegen, welche es ermöglichen sollte, PCR Produkte für die DNA-Microarray-Chips zu generieren. Ausgangspunkt hierfür war die *E.coli* DH5 α Plasmid-Genbank der Göttinger Genomics Laboratories (G₂L), die zur Genomsequenzierung von *Ms. mazei* angelegt wurde. Diese Genbank trug das Genom von *Ms. mazei* Gö1 als ‚shotgun-library‘ in dem pTZ19R Vektor (siehe 2.1 Abb.: 1). Die Plasmide beinhalteten chromosomale Fragmente des Genoms von einer durchschnittlichen Größe von 2000 bp und wurden jeweils auf beiden Strängen terminal sequenziert. Diese Inserts dienten als Templates in einer PCR, um die DNA-Microarray-Chips herzustellen. Der Vorteil dieses Ansatzes war, dass nur ein genspezifischer Primer benötigt wurde, um die PCR-Produkte zu generieren (siehe Abb. 7). Ferner lag ein großer Vorteil dieser Genbank darin, dass die Plasmide bekannte Inserts enthielten und damit das Risiko der Erzeugung von falschen PCR-Produkten minimiert wurde.

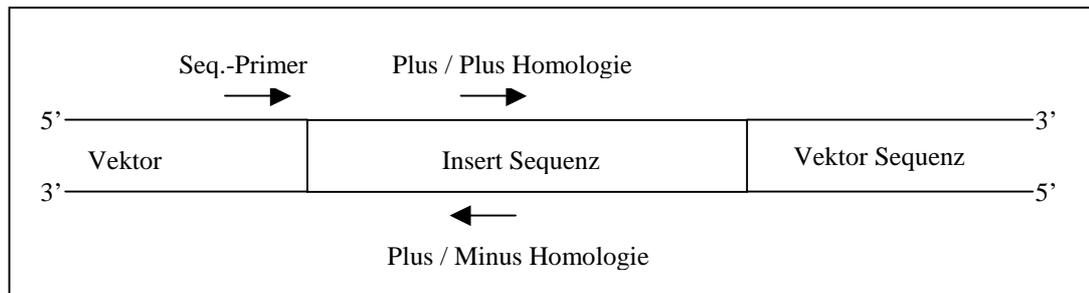


Abbildung 7: Schema der PCR-Produkt-Erzeugung ausgehend von Genbank-Plasmiden

Diese Abbildung ist eine schematische Darstellung von möglichen Homologien der genspezifischen Primer zu den Sequenzen der Genbankplasmide. Der Sequenzierungsprimer (= Seq-Primer) liegt auf dem Strang der Sequenz, welche in der BLAST-Datenbank vorlag. Genspezifische Primer mit einer Plus / Minus-Orientierung der Homologie zu den Sequenzdaten wurden zusammen mit dem Sequenzierprimer verwendet, um mittels PCR die Genfragmente (PCR-Produkte) für die TOPO-Genbank zu produzieren.

Da für die Erstellung von Microarrays große Mengen an PCR-Produkten und Primern benötigt wurden, und weil Kooperationen mit anderen Arbeitsgruppen bestanden, wurden diese PCR-Produkte in einen TOPO-Vektor kloniert. Diese TOPO-Genbank ermöglichte es, mit zwei universellen Primern alle PCR Produkte in großen Mengen nach zu produzieren.

Für die Erzeugung der PCR-Produkte mit Hilfe der Genbank der G₂L (Göttinger Genomics Laboratories) war es notwendig, geeignete Plasmide in der Genbank (ca. 50.000 Plasmide) zu identifizieren. Hierzu wurde ein multi-FASTA-File aus den einzelnen FASTA-Files der Sequenzierläufe mittels eines selbst erstellten PERL-Programms erstellt. Der Quellcode dieses Programms ist in Anhang A Skript-4 dokumentiert. Die Sequenzierläufe des Göttinger Genom Labors enthielten keinen Verweis auf die zugehörigen Klone. Diese waren in einem anderen File abgespeichert und mussten durch das Skript-5, dokumentiert in Anhang A, dem multi-FASTA-File zugewiesen werden. Das erhaltene neue multi-FASTA-File wurde durch die ‚formatdb‘ Funktion des lokalen BLAST-Servers in eine BLAST-Datenbank konvertiert. Diese Datenbank wurde gegen Primerpaare abgeglichen, welche in einer PCR den ganzen ORF amplifizieren würden. Die 3'-terminalen und 5'-terminalen Primer wurden mit Hilfe des Webinterfaces der UC-Irvine (siehe

2.9.3) für jeden ORF von *Ms. mazei* generiert. Hierfür musste mit Microsoft Excel für jeden Contig eine Tabelle folgenden Formats erstellt werden.

Tabelle 6: Beispiel einer Eingabetabelle für den Server der UC-Irvine

A	B	C	D	E	F	G	H	
b0001	thrL	190	255	>		1	190	189
b0002	thrA	337	2799	>		255	337	82
b0003	thrB	2801	3733	>		2799	2801	2
b0004	thrC	3734	5020	>		3733	3734	1
b0005	b0005	5234	5530	>		5020	5234	214
b0006	yaaA	5683	6459	<		6459	6529	70
b0007	yaaJ	6529	7959	<		7959	9928	1969
b0008	talB	8238	9191	>		5530	8238	2708
b0009	mog	9306	9893	>		9191	9306	115
b0010	yaaH	9928	10494	<		10494	10643	149
b0011	b0011	10643	11356	<		11356	11382	26
b0012	htgA	10725	11315	>		9893	10725	832
b0013	yaal	11382	11786	<		11786	15869	4083

Die vom Webserver generierten Primer wurden per E-mail an den Benutzer versandt.

In den Spalten sind folgende Parameter anzugeben gewesen:

A: Nummer des ORFs

B: Name des ORFs

C: Start des ORFs (5') auf dem (+)-Strang des Contigs

D: Ende des ORFs (3') auf dem (+)-Strang des Contigs

E: Orientierung des ORFs auf dem Contig (> = Vorwärts, < = Rückwärts)

F: Start der intergenen Region (5') auf dem (+)-Strang des Contigs

G: Ende der intergenen Region (3') auf dem (+)-Strang des Contigs

H: Größe der intergenen Region

Für detaillierte Anweisungen über die Bedingungen, welche von der Tabelle erfüllt werden müssen und wie sie am schnellsten generiert werden, wird auf die Homepage des Servers verwiesen: http://promoter.ics.uci.edu/Primers/manual.htm#io_table.

Diese Primer wurden über ein selbst geschriebenes PERL-Skript gegen die angelegte BLAST-Datenbank der Sequenzierläufe abgeglichen. Der Quellcode dieses Skriptes ist im Anhang A Skript-6 dokumentiert.

Tabelle 7: Beispielausgabe des Skriptes

Primer Orientierung	ORF	ORF Größe	Klon Name in der Genbank	Orientierung der Homologie	Start, Ende und Sequenz der Homologie mit der Insert Sequenz
for primer	2635	957	>G2L_53-1-4-A7-LL.exp	plus/minus	220 gctacagacgaccgaaaaggact 197
for primer	2635	957	>G2L_53-1-1-H5-LR.exp	plus/minus	311 gctacagacgaccgaaaaggact 288
for primer	2637	1026	>G2L_53-2-74-B12-13.exp	plus/minus	292 tatccatcagccgtttctttccc 269
for primer	2637	1026	>G2L_53-2-292-F10-13.exp	plus/plus	262 tatccatcagccgtttctttccc 285

Diese Ausgabe enthielt die Orientierung und Sequenz des Primers, die Länge und Nummer des ORFs, sowie die Position der Homologie innerhalb des Gens. Ferner wurden durch Plus/Plus und Plus/Minus die Orientierungen der Homologien in Relation zur Eingabesequenz definiert (siehe Abb.: 7).

Diese Ergebnisse wurden in einem Tabellenkalkulationsprogramm (Microsoft Excel) sortiert und nach folgenden Kriterien selektiert (hierbei war die Position der Homologie gleichzusetzen mit der Größe des zu erwartenden PCR-Produktes):

- Die Länge des PCR Produktes musste größer als 160 bp sein
- Die Länge des PCR Produktes musste kleiner sein als die Länge des ORFs
- Die Orientierung der Homologie musste Plus/Minus (=Gegenläufig) sein, damit nur ein spezifischer Primer für die PCR benötigt wurde.

War kein Plasmid mit gegenläufiger Orientierung vorhanden, wurden Plasmide verwendet, die eine Plus/Plus-Orientierung zum Primer aufwiesen. Hierbei mussten allerdings zwei Gen-spezifische Primer eingesetzt werden. In diesem Fall wurden neue Primer generiert (Primer3 Programm, siehe 2.9.3), die nicht den ganzen ORF

amplifizieren. Dies war notwendig, um eine eventuelle toxische Auswirkung der Gene bzw. ihrer Proteine aus *Ms. mazei* auf *E. coli* zu vermeiden. Die gefundenen Klone wurden in LB-Medium über Nacht kultiviert und 5 µl hiervon in eine PCR (2.5.1) eingesetzt. In den Fällen, in denen es kein geeignetes Plasmid gab, wurde die PCR mit genomischer DNA (2.4.5.2) von *Ms. mazei* Gö1 als Template durchgeführt (ca. 20 % der gesamten Anzahl von offenen Leserahmen). Für die Primer-Generierung galten die gleichen Bedingungen wie für Plus/Plus-orientierte Plasmide.

Die entstandenen PCR Produkte wurden nicht für die DNA-Chip Produktion verwendet, sondern unaufgereinigt in den pCR2.1 oder pCR4.0 TOPO-Vektor kloniert (2.7.2) und in *E.coli* DH5α (2.7.1) transformiert. Diese erneute Klonierung ermöglichte es, mit Plasmid-spezifischen Primern PCR-Produkte in großen Mengen zu generieren und später auch nachzuproduzieren. Diese Methode wurde gewählt, da die PCR mit den TOPO-Vektoren in allen Reaktionen unter den gleichen Bedingungen durchgeführt werden konnten. Dadurch war die Ausbeute der PCR-Produkte wesentlich höher und gleichmäßiger als in den PCR-Reaktionen mit chromosomaler DNA als Template (siehe Abb. 8). Aufgrund der Tatsache, dass für die Chip-Produktion 5-10 µg PCR-Produkte benötigt wurden und die Ausbeute bei vielen PCR-Reaktionen mit chromosomaler DNA als Template niedriger war, hätten ansonsten mehrere Reaktionen und Aufreinigungen durchgeführt werden müssen. Ferner ermöglichte diese TOPO-Genbank den Kooperationspartnern die Nachproduktion von PCR-Produkten mit nur zwei Primern.

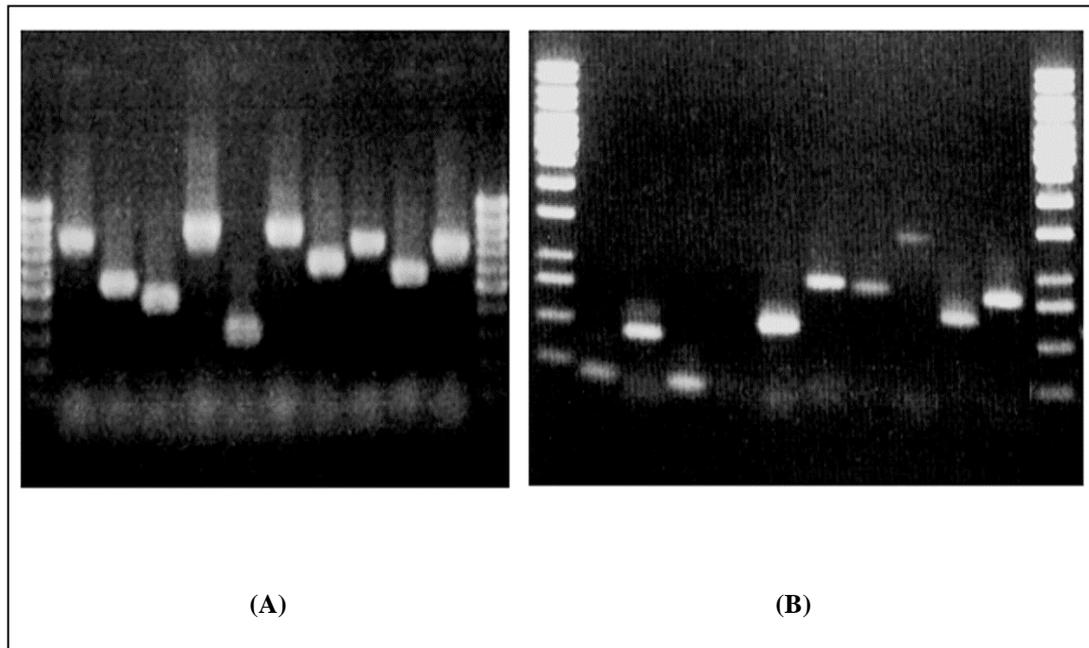


Abb. 8: Generierte PCR Produkte mit universellen Primern und mit Topo-Klonen als Template (A) bzw. mit Gen-spezifischen Primern und chromosomaler DNA als Template (B).

Für die Amplifizierung mit chromosomaler DNA wurden zwei spezifische Primer benutzt (B). Die PCR mit TOPO-Klonen als Template wurde mit zwei universellen Primern durchgeführt (A). In beiden Fällen betrug die Annealing-Temperatur 60 °C und die PCR durchlief 30 Zyklen in einem Volumen von 100 µl. 10 µl der PCR-Reaktionen wurden jeweils auf einem 1,7%igen Agarose Gel aufgetrennt. Wie aus den Abbildungen zu ersehen ist, war die Ausbeute der PCR-Reaktionen mit chromosomaler DNA als Template (B) niedriger und ungleichmäßiger als in den Reaktionen, bei denen TOPO-Klone als Template eingesetzt wurden (A).

3.2.2 Generierung der Genom umfassenden DNA-Microarray-Chips

Mit den TOPO-Klonen aus 3.2.1 wurde, wie in 2.5.1 beschrieben, im 96er-Maßstab eine PCR über 40 Zyklen durchgeführt. Als Template wurden 5 µl einer neu angezogenen über Nacht Kultur verwendet. Die PCR Produkte wurden anschließend mittels Agarosegelelektrophorese (2.4.5.1) überprüft und im 96er-Maßstab aufgereinigt (2.4.5.1.2). Die Elution der PCR-Produkte in der Aufreinigung erfolgte mit 60 µl H₂O (dest.). Diese 60 µl wurden empirisch ermittelt um in jedem Fall eine Konzentration von über 60 µg/µl DNA zu erhalten. Die Konzentrationsmessungen erfolgten in einem Plattenphotometer. Die aufgereinigten PCR-Produkte wurden in 96er-TPP Platten (Fa. ABgene, Hamburg) bei -20 °C gelagert. Als leere Chips wurden Aminosilan beschichtete Type7*-Slides (Fa. Amersham Biosciences,

Freiburg) benutzt. 10 µl der gewonnenen PCR Produkte wurden in 384 Loch Platten (Fa. Amersham Biosciences, Freiburg) pipettiert. In die Wells dieser Platten waren 10 µl DMSO vorgelegt worden. Die PCR-Produkte aus diesen 384 Loch Platten wurden, wie in 2.8.1 beschrieben, auf die Slides gedruckt. Nach Bestrahlung mit ultraviolettem Licht zur Fixierung der PCR-Produkte (2.8.1), wurden diese bis zur Hybridisierung unter Lichtausschluss im Exikator gelagert. Diese Fixierung ist in Abbildung 9 schematisch dargestellt.

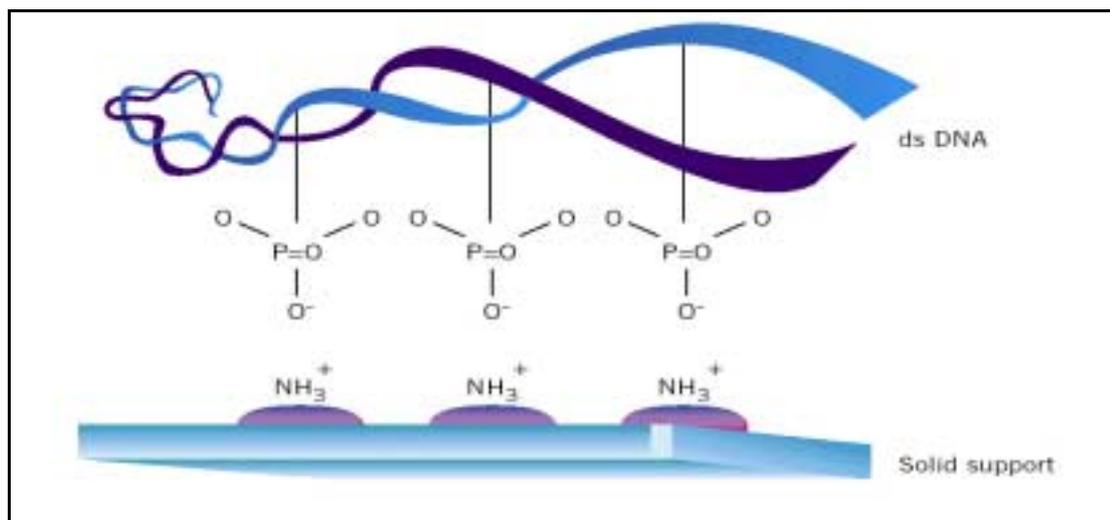


Abb. 9: Aminosilan Kopplung

Dies Abbildung zeigt ein Schema der Kopplung unmodifizierter DNA an eine Amino-Silan beschichtete Objektträger-Oberfläche. Die Kopplung erfolgt durch eine elektrostatische Wechselwirkung. Hierbei interagieren die positiv geladenen Aminogruppen der Slide-Beschichtung mit dem negativ geladenen Phosphat-Rückgrat der DNA. (Quelle: Amersham Biosciences, Freiburg)

3.2.3 Herstellung der fluoreszenzmarkierten Sonden

Für die in dieser Arbeit dargestellten Ergebnisse wurde Gesamt-RNA aus Stämmen von *Ms. mazei* Gö1 isoliert (2.4.5.2) und mittels reverser Transkription in cDNA umgeschrieben. Die RNA-Lösung durfte keine DNA mehr enthalten, da die verwendete CyScript-Reverse-Transkriptase (Fa. Amersham Biosciences, Freiburg) nach Angaben des Herstellers ebenfalls DNA als Template verwenden kann.

Wie aus den Abbildung 10a und 10b zu ersehen ist, lag keine Kontamination mit chromosomaler DNA vor, da keine Nukleinsäuren einer höheren Molekülgröße zu sehen waren.

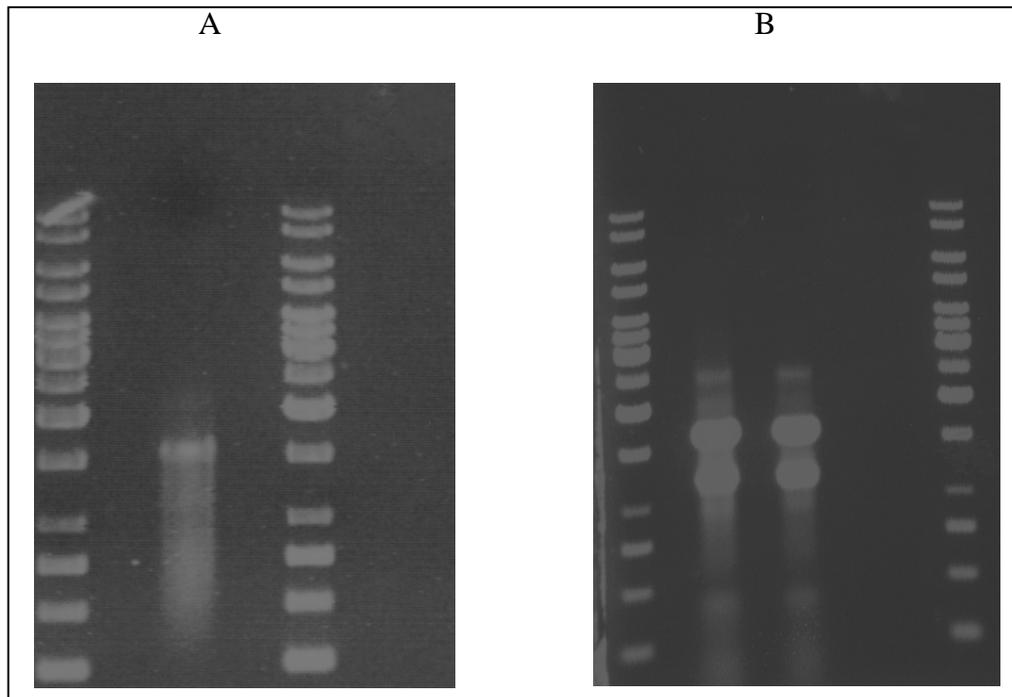


Abb. 10: Agarosegelelektrophorese von aufgereinigter RNA aus Acetat- (A) und Methanol- (B) gewachsenen Zellen

In diesen Abbildungen sind die Agarosegelelektrophoresen (2.4.6.1) von aufgereinigter RNA mit einer 100 bp Leiter (Fa. MBI) als Größenstandard dargestellt. Die RNA stammte aus Zellen, welche mit Methanol (B) oder Acetat (A) gewachsen waren. Es wurden jeweils 2 μ l der aufgereinigten RNA-Lösung nach DNaseI-Verdau aufgetragen.

Ferner war zu erkennen, dass bei Aufreinigungen von RNA aus Acetat-gewachsenen Zellen eine geringere Ausbeute erzielt wurde, als aus Kulturen, die auf Methanol gezogen wurden. Zudem war der Anteil an ribosomaler RNA bei Methanol-gewachsenen Zellen höher war als bei Acetat-gewachsenen Zellen. Die Beobachtung, dass unterschiedliche Mengen an rRNA beim Wachstum auf verschiedenen Substraten zu finden sind, wurde allerdings auch schon im Rahmen von Transkriptionsanalysen mit *E. coli* gemacht (Tao et al, 1999).

Nachdem die RNA-Präparationen mittels Agarosegel-Elektrophorese (2.4.6.1) überprüft worden war (siehe Abb. 10a und b), erfolgte eine getrennte Umschreibung der RNAs in cDNA. Hierbei wurden mit unterschiedlichen Farbstoffen markierte dCTPs verwendet, um fluoreszenzmarkierte cDNA zu erhalten. Im Zuge der reversen

Transkription wurden Fluoreszenz-markierte Desoxy-Nukleotide (dCTP) verwendet, so dass die entstandene cDNA mit den Farbstoffen markiert war (2.6.1). Als Fluoreszenzfarbstoffe wurden Cy3 und Cy5 (Fa. Amersham Biosciences, Freiburg) benutzt. Diese Fluoreszenz-markierte cDNA wurde aufgereinigt (2.6.1) und mit Hilfe eines Photometers auf ihre Reinheit und den Einbau der Farbstoffe überprüft (Abb. 11 a und b).

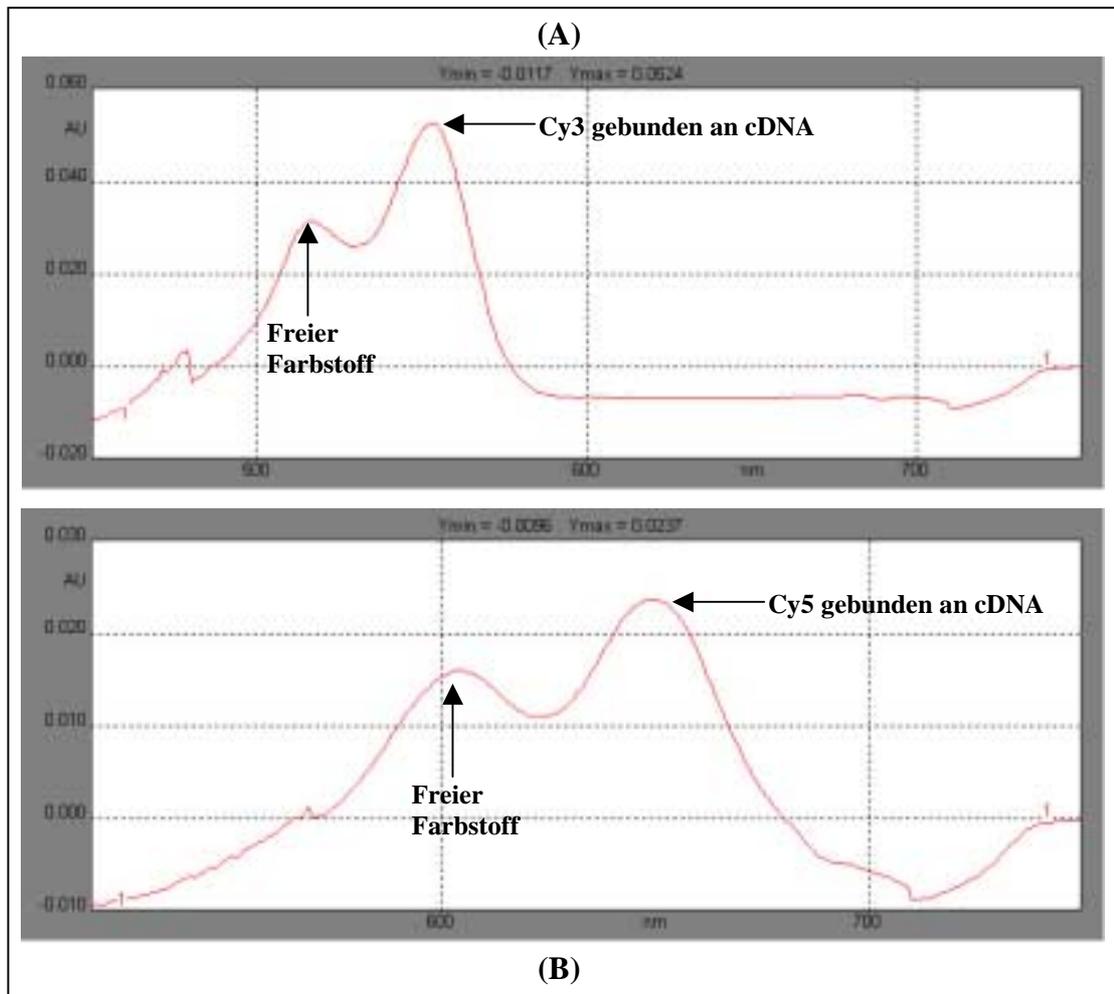


Abb. 11: Absorptionsspektren aufgereinigter Cy3/Cy5 markierter cDNA

Die Y_{\max} Werte definieren den Absorptionswert in AU (Absorptionseinheiten) des höchsten Peaks. Für die Messungen wurden 5 μ l der aufgereinigten cDNA mit Elutionspuffer (GFX Purification Kit, Amersham Biosciences, Freiburg) um den Faktor 20 verdünnt.

Abbildung 11 zeigt die Absorptionsspektren von Cy3- (A) bzw. Cy5- (B) markierter cDNA. Diese beiden cDNA-Lösungen wurden aus derselben Gesamt-RNA-Präparation hergestellt. Die Spektren wurden nach einer 1:20 Verdünnung aufgenommen. Die Spektren verdeutlichen, dass ein Rest des ungebundenen, freien Farbstoffs (dCTPs) nach der Aufreinigung beider Markierungsreaktionen verblieben

war. Dies war daran zu erkennen, dass links neben dem Peak mit der spezifischen Absorptionswellenlänge des in cDNA inkorporierten Farbstoffes (532 bei A und 635 bei B) ein zweiter Peak zu sehen war. Bei einer fehlgeschlagenen Markierungsreaktion wäre nur der zweite Peak zu sehen gewesen.

Die Abweichung zwischen den Y_{\max} Werten der Absorptionen (in AU = Absorptionseinheiten) beruht auf den unterschiedlichen molaren Absorptionskoeffizienten (ϵ) der Farbstoffe (Amersham Biosciences, Freiburg).

Molare Absorptionskoeffizienten nach Angaben des Herstellers:

$$\epsilon(\text{Cy5}) = 250000 \text{ mol}^{-1} \text{ cm}^{-1}$$

$$\epsilon(\text{Cy3}) = 150000 \text{ mol}^{-1} \text{ cm}^{-1}$$

Absolute Werte $Y_{\max}(\text{Cy3})/Y_{\max}(\text{Cy5})$:

$$Y_{\max}(\text{Cy3}) = 0.0524$$

$$Y_{\max}(\text{Cy5}) = 0.0237$$

Relativierte Werte von $Y_{\max}(\text{Cy3})/Y_{\max}(\text{Cy5})$:

$$Y_{\max}(\text{Cy3}) = 0.0524 * 150000 \text{ AU}/(\text{mol} * \text{cm}) = 7860 \text{ AU mol}^{-1} \text{ cm}^{-1}$$

$$Y_{\max}(\text{Cy5}) = 0.0237 * 250000 \text{ AU}/(\text{mol} * \text{cm}) = 5925 \text{ AU mol}^{-1} \text{ cm}^{-1}$$

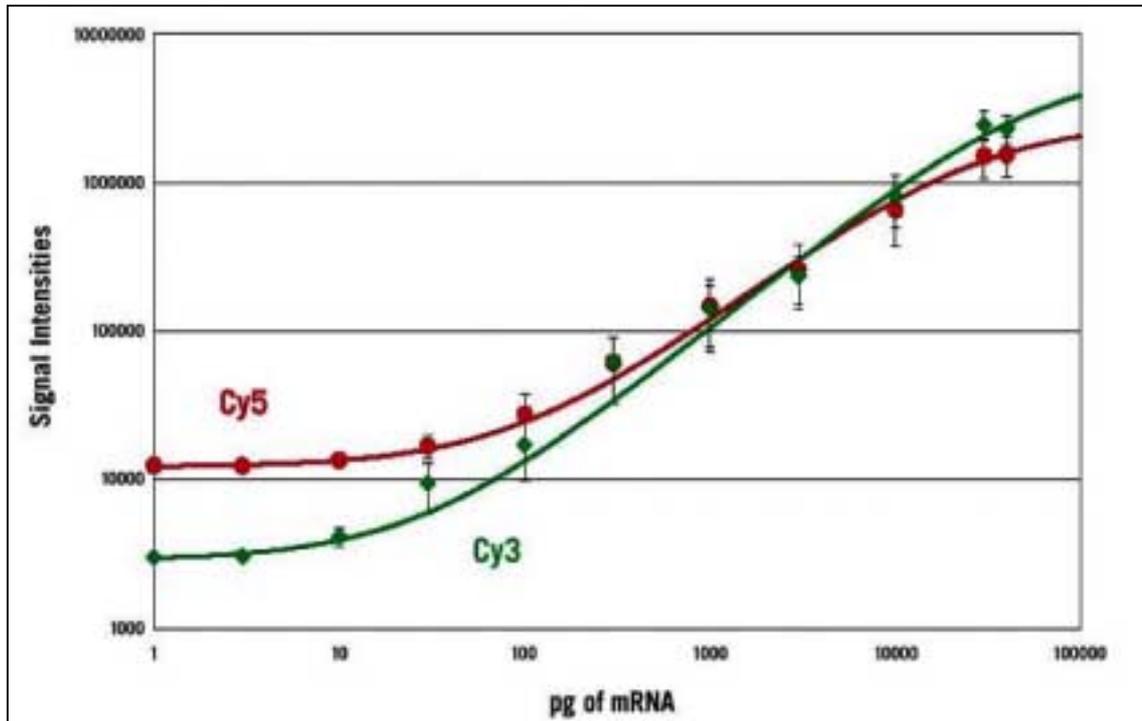
Daraus ergibt sich ein relativiertes Verhältnis der Inkorporationen von:

$$7860 / 5925 = 1.3:1.0$$

Dieser um den Faktor 1,3 schlechtere Einbau des roten Farbstoffes (Cy5) stimmte mit den Angaben des Herstellers (Amersham Biosciences, Freiburg) überein. Die verminderte Inkorporation beruht darauf, dass der Farbstoff Cy5 größer ist als Cy3 und daher schlechter von Reversen-Transkriptasen eingebaut wird.

Ferner besitzen die beiden Farbstoffe unterschiedliche physikalische Fluoreszenzeigenschaften. Der rote Farbstoff (Cy5) fluoresziert bei gleicher Anregungsenergie stärker als der grüne Farbstoff (Cy3). Dieser Effekt führt dazu, dass das Signal des roten Farbstoffes in Gegenwart geringeren Mengen an hybridisierter, markierter cDNA früher in den Bereich der Sättigung gelangt als das

Signal des grünen Farbstoffs. In diesem Bereich überlagern sich die Fluoreszenzsignale, und eine Auswertung der Daten ist nicht mehr möglich (siehe Abb. 12).



3.3 Vorversuche mit den Genom-umfassenden DNA-Microarray-Chips

Um die Funktionsfähigkeit der Chips zu testen und erste Erfahrungen mit der Genom-weiten Transkriptionsanalyse zu sammeln, wurden Vorversuche unternommen. Die Experimente fanden parallel zu der Chip- und TOPO-Genbank-Produktion statt. Ziel war es, das Transkriptom unter zwei Bedingungen zu untersuchen, bei denen eine starke Änderung im Transkriptionsverhalten vieler Gene erwartet wurde.

3.3.1 Genom-weite Transkriptionsanalyse von *Ms. mazei* beim Wachstum auf Voll- bzw. Minimalmedium

Es wurde die Genom-weite Transkription von *Ms. mazei* Zellen untersucht, die entweder auf Vollmedium (2.2.2.1) oder auf Minimalmedium (2.2.2.1 ohne Hefeextrakt und Casiton) angezogen wurden. Analysen wurde viermal mit neu angezogen Zellen und neu angesetzten Lösungen wiederholt. Dieses Vorgehen war notwendig, um die Reproduzierbarkeit der Daten zu zeigen. In diesem Zusammenhang bleibt anzumerken, dass die Chip Technologie eine hochsensitive Methode ist und die kleinsten Abweichungen in der Zusammensetzung der verwendeten Lösungen, Medien oder des Zellmaterials zu starken Schwankungen in den Ergebnisse führt. Deshalb kann nur der Mittelwert der relativen Expressionslevel aus mehreren Experimenten eine signifikantere Aussage ergeben (Lee et al., 2000; Liang et al., 2003).

Die Zellen wurden wie unter 2.3.1 auf Methanol bzw. Acetat angezogen und Gesamt-RNA aufgereinigt sobald die optische Dichte der Kulturen einen Wert von 0.5 erreicht hatte (2.4.5.2). Diese Gesamt-RNA Präparationen (Kapitel 2.6.1) dienten als Template für die Erzeugung fluoreszenzmarkierter cDNA (mit Cy3 oder Cy5). Die pro Experiment erhaltenen vier cDNA-Präparationen wurden mit zwei DNA-Microarray-Chips hybridisiert: i) Chip1: cDNA aus Zellen die auf Vollmedium angezogen worden waren, wurde mit Cy3 und cDNA aus Minimalmedium-gewachsenen Zellen wurde mit Cy5 markiert. ii) Chip2: zur Markierung von cDNA aus Zellen, die mit Vollmedium angezogen worden waren, diente Cy5, wohingegen cDNA aus Minimalmedium-gewachsenen Zellen mit Cy3 markiert wurde. Der Mittelwert der Ergebnisse aus diesen beiden Chip-Experimenten ermöglichte es, die unterschiedlichen Einbauraten der fluoreszenzmarkierten dCTPs im Zuge der reversen Transkription zu kompensieren (Tseng et al, 2001). Die aufgereinigten cDNA-Sonden wurden über Nacht in einem ASP (Automated Sample Processor) der Firma Amersham Biosciences (Freiburg) mit bedruckten Type7* DNA-Microarray-Chips hybridisiert (2.8.2). Die Chips wurden anschließend mit einem Laser-Scanner (Fa. Axon, Union City, Kalifornien USA, <http://www.axon.com>) gescannt (2.8.3). Abbildung 13 zeigt einen Ausschnitt eines DNA-Microarray Chips.



Abb. 13: Expressionsprofiling von *Ms. mazei* mit Hilfe eines Microarray Chips

In dem gezeigten Experiment wurde die cDNA einerseits mit Cy5-dCTP (Zellen aus Komplexmedium) andererseits mit Cy3-dCTP (Zellen aus Minimalmedium) markiert. Für diesen Versuch wurden *Ms. mazei*-Kulturen bei 37 °C auf Voll- bzw. Minimalmedium angezogen und bei einer optischen Dichte von 0.5 geerntet. Das Vollmedium setzte sich wie folgt zusammen: Salze und Spurenelementlösung (Hippe et al., 1979), Vitaminlösung (Wolin et al., 1964), Acetat 1 g/l, Hefeextrakt 2 g/l, Casiton 2 g/l und Methanol 150 mM (Kapitel 2.2.2.1). Vor dem Animpfen wurde das Medium mit Na₂S (0,3 mg/ml) und Cystein (0,3 mg/ml) reduziert. Im Fall des Minimalmediums wurde kein Hefeextrakt und kein Casiton eingesetzt. Die Zellen wurden im Vorfeld viermal in Minimalmedium überimpft und erst dann für die Chip Experimente verwendet. Dieses Vorgehen führte dazu, dass keine Spuren von Aminosäuren und anderen organischen Substanzen im Medium vorhanden waren. Die PCR-Produkte wurden im Duplikat gedruckt, weshalb die oberen und unteren Blöcke jeweils identisch sind. Die Laser wurden bei diesem Bild bereits justiert (siehe 3.3.1.2). Die Lücken auf dem Array rührten daher, dass die Chips, die in diesen Vorversuchen verwendet wurden, noch nicht das komplette Genom repräsentierten.

Die acht entstandenen Bilder der vier Wiederholungsexperimente mit jeweils 2 Markierungsexperimenten wurden mit der GenePix Pro 3.0 Software (Fa. Axon, Union City, Kalifornien USA, <http://www.axon.com>) zuerst dokumentiert und anschließend weiter ausgewertet. Ferner wurde für die Auswertung vor dem Scannen ein GenePix-Array-List'-file in die GenePix Software geladen, welches zuvor mit Microsoft Excel generiert wurde. Dieses File enthielt die Positionen der Spots auf dem Chip mit den dazu gehörigen Annotationen (Abb. 14).

```

ATF 1.0
28 5
Type=GenePix ArrayList V1.0
BlockCount=24
BlockType=0
URL=http://genome-www4.stanford.edu/cgi-bin/SGD/locus.pl?locus=[ID]

Block1=2510, 2700, 130, 16, 240, 9, 240
Block2=16215, 2745, 130, 16, 240, 9, 240
Block3=2630, 7210, 130, 16, 240, 9, 240
Block4=16240, 7255, 130, 16, 240, 9, 240
Block5=2585, 11695, 130, 16, 240, 9, 240
Block6=16210, 11745, 130, 16, 240, 9, 240
Block7=2595, 16205, 130, 16, 240, 9, 240
Block8=16215, 16245, 130, 16, 240, 9, 240
Block9=2605, 20690, 130, 16, 240, 9, 239
Block10=16185, 20755, 130, 16, 240, 9, 240
Block11=2570, 25210, 130, 16, 240, 9, 240
Block12=16185, 25255, 130, 16, 240, 9, 240
Block13=2580, 29715, 130, 16, 240, 9, 240
Block14=16190, 29775, 130, 16, 240, 9, 240
Block15=2544, 34220, 130, 16, 240, 9, 239
Block16=16240, 34275, 130, 16, 240, 9, 240
Block17=2505, 38725, 130, 16, 240, 9, 240
Block18=16215, 38770, 130, 16, 240, 9, 240
Block19=2490, 43225, 130, 16, 240, 9, 240
Block20=16185, 43285, 130, 16, 240, 9, 240
Block21=2510, 47715, 130, 16, 240, 9, 240
Block22=16180, 47790, 130, 16, 240, 9, 240
Block23=2455, 52235, 130, 16, 240, 9, 240
Block24=16110, 52270, 130, 16, 240, 9, 240
Block      Row      Column      Name      ID
1      1      1      Conserved protein      2414
1      1      2      Conserved protein      2672
1      1      3      ALAO H+ ATPASE_ SUBUNIT I      4041
.      .      .      .      .
.      .      .      .      .
.      .      .      .      .

```

Abb. 14: Gene Pix Array List file (.gal)

Der obere Teil dieses Files enthält die Positionen der 24 Blöcke auf dem Chip. Weiter unten folgen die Annotationen und Nummern der Gene, welche an den angegebenen Positionen in Form von PCR-Produkten vorlagen. Die Reihenfolge und das Format sind von der Firma Axon vorgeschrieben, damit die Software die Daten einlesen kann. Für genauere Angaben zur Erstellung dieser Files, wird auf die Angaben des Herstellers verwiesen. (<http://www.axon.com>).

Durch die unterschiedlichen physikalischen Eigenschaften der beiden Farbstoffe war es notwendig, die erhaltenen Daten zu normalisieren (siehe Abb. 12). Der Vorgang der Normalisierung ist eine bioinformatische Notwendigkeit vor der Auswertung von Microarray Daten, um Faktoren wie die unterschiedliche Inkorporation in cDNA und die unterschiedlichen molaren Fluoreszenzeigenschaften der Farbstoffe zu

kompensieren (Quackenbush, 2003; Bilban et al, 2002; Tran et al, 2002). Dieses unterschiedliche Fluoreszenzverhalten war ferner der Grund weshalb ein maximaler Fluoreszenzwert (upper threshold) und ein minimaler Fluoreszenzwert (lower threshold) gesetzt wurde. Die ‚thresholds‘ wurden für beide Kanäle auf den gleichen Wert gesetzt, da ansonsten keine Berechnung des Verhältnisses zwischen den beiden Kanälen möglich gewesen wäre. Als ‚upper threshold‘ wurde der Wert 25000 definiert, da oberhalb dieses Wertes eine Sättigung des Rot-Kanals eintrat. Alle Werte oberhalb dieses Wertes wurden in beiden Kanälen auf 25000 gesetzt. Als ‚lower threshold‘ wurde die doppelte Standardabweichung des lokalen Hintergrunds gesetzt. In der Regel war es die des Grün-Kanals, da der Hintergrund hier schwächer war. Dieser ‚threshold‘ wurde für jeden Spot einzeln berechnet, da der Hintergrund innerhalb eines Chips starken Schwankungen unterliegen kann. Lag der Wert in einem der beiden Kanäle unterhalb dieses ‚thresholds‘, wurde der Wert auf den Wert des ‚thresholds‘ gesetzt. Dies war notwendig, da alle Werte unterhalb des ‚lower threshold‘ stark von Schwankungen des Hintergrunds abhingen und daher die Berechnung des Verhältnisses zwischen den beiden Kanälen verfälschten. Gene mit Werten unterhalb der doppelten Standardabweichung des Hintergrunds in einem der beiden Kanäle (meist im Rot-Kanal), wurden gesondert betrachtet. Sie unterlagen trotz des ‚lower-threshold‘ noch starken Schwankungen des Hintergrunds in diesem Kanal.

Die in diesen Vorexperimenten erzeugten Daten wurden nicht - wie in 3.4.2 - mit den Statistik-Routinen der Software GenePix Pro 3.0 (Fa. Axon, Union City, Kalifornien USA, <http://www.axon.com>) normalisiert. Die Unterschiede in der Gen-Expressionen unter den gegebenen Wachstumsbedingungen waren zu stark, um sie mit einem automatisierten Algorithmus gegeneinander zu normalisieren. Dies war darauf zurückzuführen, dass die Expression sehr vieler Gene hoch oder herunter gefahren wurde, wenn ein Vergleich der Expressionsmuster beim Wachstum auf Komplex- bzw. Minimalmedium analysiert wurde. Statt der Software GenePix Pro 3.0 wurde ein externes Kontrollsystem der Fa. Amersham Biosciences benutzt, die ‚Lucidea Universal Scorecard‘. Dieses System beruht auf PCR-Produkten definierter Mengen und Konzentrationen, die mit auf den Chip gedruckt werden. Diese PCR-Produkte entstammen intergenen Regionen von *S. cerevisiae* und hybridisieren mit keiner bekannten kodierenden Sequenz. Ferner wird bei dieser Methode eine Kontroll-

RNA-Lösung in die Markierungsreaktion (2.6.1) eingesetzt, um die Inkorporationen und Halbwertszeiten der Farbstoffe relativieren zu können. Die Score-Card (Abb. 15) wurde in 16facher Kopie auf den Chip gedruckt.

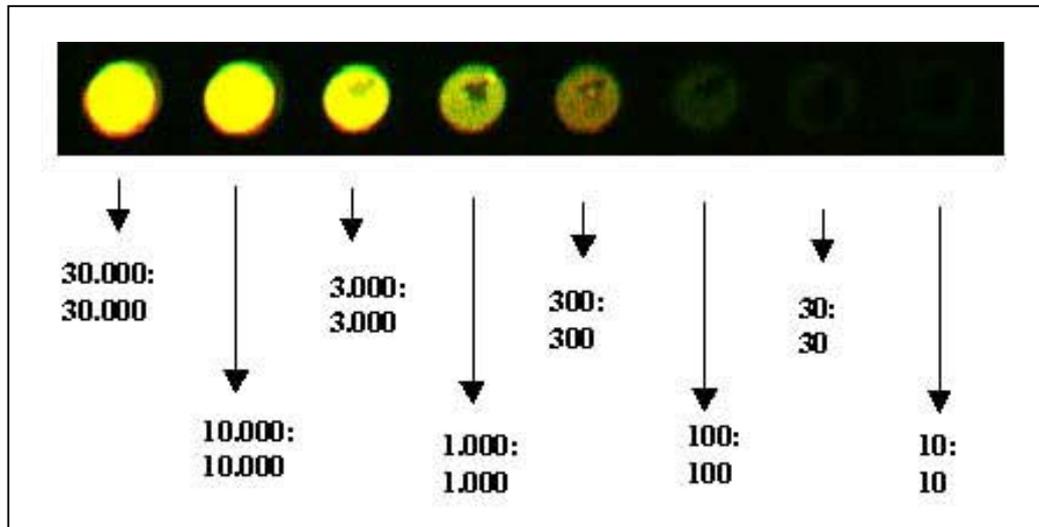


Abb. 15: Lucidea Universal Score Card

Diese Abbildung zeigt einen Scan einer Konzentrationsreihe der Lucidea Universal Score Card (Fa. Amersham Biosciences, Freiburg). Die Zahlen geben die in die Markierungsreaktion eingesetzte Menge an Template-Spike der Score-Card für die beiden Farbstoffe an (pg mRNA für die Cy5 Markierung : pg mRNA für die Cy3 Markierung). Für jedes Verhältnis wurde eine andere Template-Spike-mRNA in der Markierungsreaktionen eingesetzt. Ferner lässt sich erkennen, dass bei niedrigeren Mengen an Farbstoffen (300 pg : 300 pg und weniger) die Farbe des Spots nach Rot umschlug. Dieser Effekt war auf die unterschiedlichen physikalischen Eigenschaften der Farbstoffe zurückzuführen (siehe Abb. 12)

Im Verlauf des Scannens der Chips wurden die Laserintensitäten über diese Scorecard abgeglichen. Hierbei wurde zuerst der gesamte Chip gescannt und nur die Spots der 16 Scorecard Kopien im Plot analysiert. Eine idealer Plot, wie er erzielt werden sollte, ist in Abbildung 16 dargestellt. Falls die Verhältnisse der Fluoreszenzkanäle nicht gleich eins waren, wurde die Laserintensität des schwächeren Kanals erhöht, und es wurde erneut gescannt. Dies wurde so lange wiederholt, bis die Verhältnisse der gemessenen Intensitäten gleich eins war.

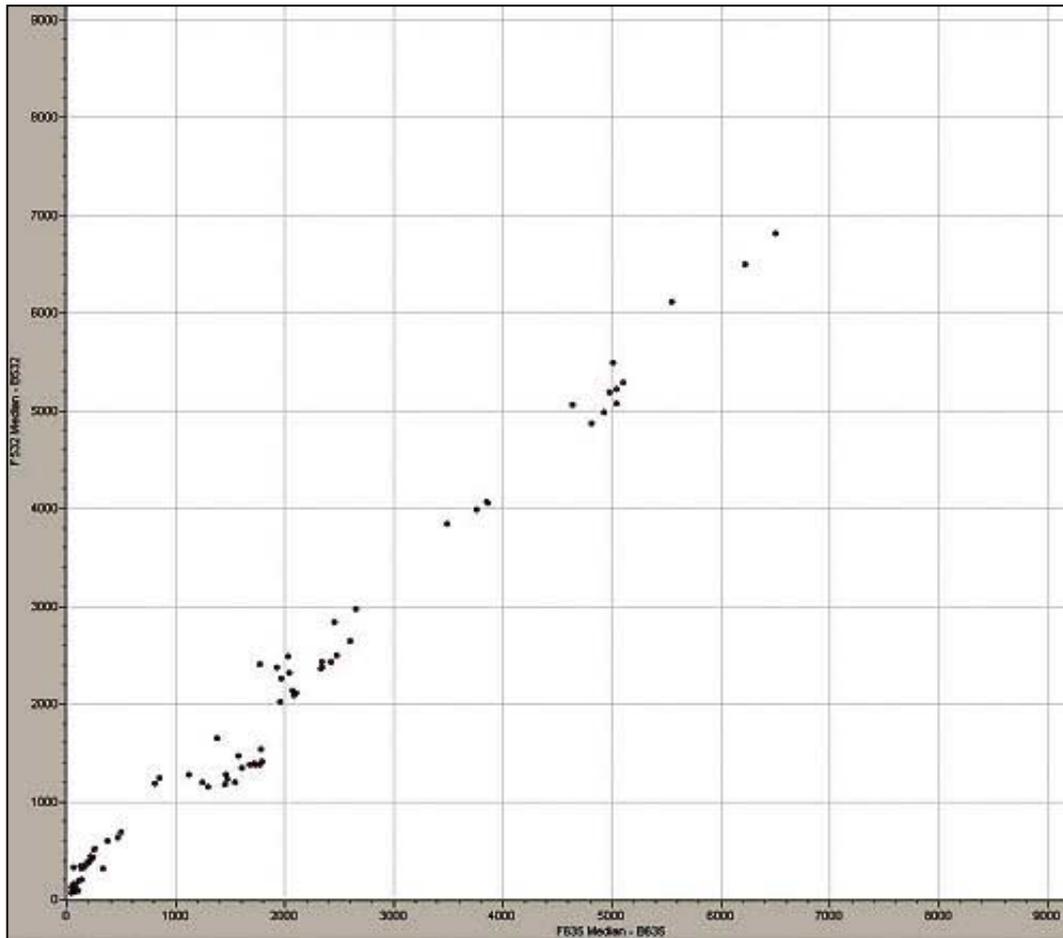


Abb. 16: Scatterplot der Score Card Daten nach Einstellung der Laser-Intensitäten

Aufgetragen sind die Intensitäten bei 532 nm minus des lokalen Hintergrunds bei 532 nm gegen die Intensitäten bei 635 nm minus des lokalen Hintergrunds bei 635 nm. Die Daten repräsentieren die 16 Kopien der Score-Card Verdünnungsreihe eines Chips. Jeder Punkt markiert das Verhältnis der beiden Fluoreszenzkanäle eines Spots. Die Punkte liegen alle annähernd auf der Achsen halbierenden, was bedeutet, dass das Verhältnis der Fluoreszenzkanäle gleich eins ist.

Erst nachdem die Laser so justiert waren, dass die Verhältnisse der Fluoreszenzkanäle aller Score-Card Spots gleich eins waren (wie in Abb. 16), wurde der gesamte Chip mit einer hohen Auflösung (5 μm) gescannt und die Daten aller Spots analysiert. Durch dieses Vorgehen war gewährleistet, dass die unterschiedlichen Fluoreszenz- und Inkorporationseigenschaften der beiden Farbstoffe relativiert waren. Dadurch wurden die Messwerte nicht mehr durch die unterschiedlichen Farbstoffe verfälscht. In Abbildung 17 sind die relativen Fluoreszenzdaten aller Spots in einem Scatter-Plot dargestellt.

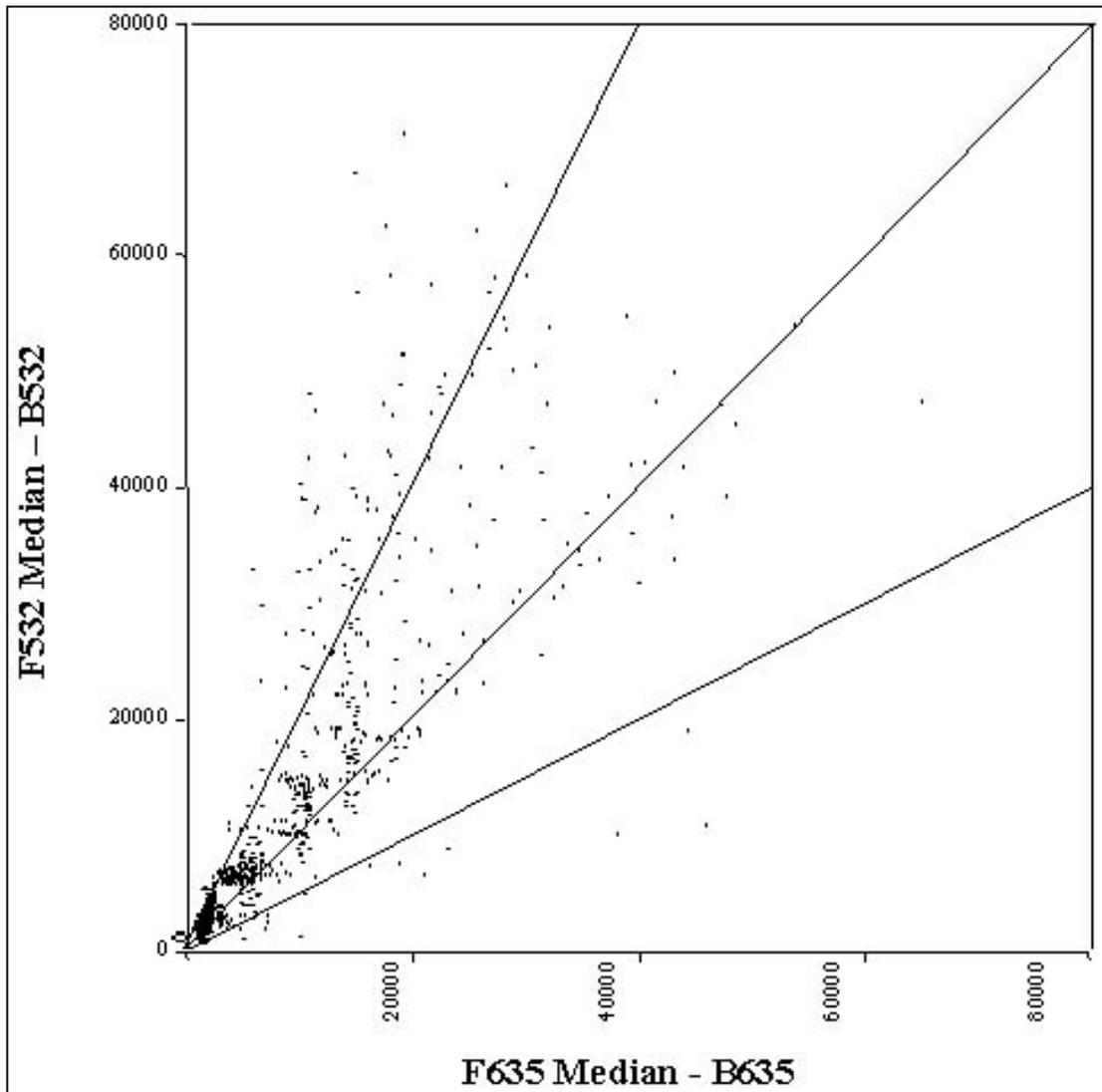


Abb. 17: Scatterplot der Transkriptomdaten von Zellen gewachsen auf Komplex- bzw. Minimalmedium

Aufgetragen sind die Intensitäten bei 532 nm minus des lokalen Hintergrunds bei 532 nm gegen die Intensitäten bei 635 nm minus des lokalen Hintergrunds bei 635 nm. Die Daten repräsentieren alle Gene, welche in beiden Kanälen eine positive Fluoreszenz aufwiesen. In diesem Experiment war die cDNA aus Vollmediums-Zellen mit Cy5-dCTP (635 nm) und die cDNA aus Minimalmediums-Zellen mit Cy3-dCTP (532 nm) markiert. Die Trennlinien verdeutlichen die Signalverhältnisse 1:1, 1:2 und 2:1.

Aus Abbildung 17 ist zu ersehen, dass eine breite Streuung der Verhältnisse der Fluoreszenzkanäle vorlag. Punkte oberhalb der oberen und unterhalb der unteren Trennlinie zeigten ein Signalverhältnis in den Fluoreszenzkanälen um mehr als den Faktor 2 (Lee et al, 2001). Hierbei wurde der lokale Background in den beiden Kanälen jedes Spots vorher von den Messwerten subtrahiert. Mehr als 10% der Spots

lagen signifikant oberhalb der oberen Trennlinie. Dies bedeutet, dass die durch diese Spots repräsentierten mRNAs in *Ms. mazei* beim Wachstum auf Minimalmedium mindestens dreimal häufiger vorkamen, als in *Ms. mazei*-Zellen, die mit Vollmedium angezogen wurden.

3.3.2 Ergebnisse der DNA-Microarray Analysen von *Ms. mazei* Gö1 nach Wachstum auf Voll- bzw. Minimalmedium

Für die Auswertung der Transkriptionsdaten wurde ein Konfidenzintervall von Faktor 3 gesetzt. Dies war wegen der starken Ungenauigkeiten, welche in der Natur der DNA-Chip-Technologie liegen, notwendig. Aus diesem Grund wurden die Experimente auch mehrfach wiederholt und der Mittelwert der einzelnen Experimente als Grundlage für Aussagen benutzt. Ferner wurde in jedem Experiment eine zweite umgekehrte Markierung vorgenommen, um das unterschiedliche Fluoreszenzverhalten und die unterschiedliche Inkorporationsrate der beiden Farbstoffe in die cDNA zu relativieren. Der Mittelwert der beiden Telexperimente wurde benutzt, um den Grad der relativen Expressionen zu definieren.

Wie aus Abbildung 17 zu ersehen ist, war die Konzentration von ca. 10 % der Gesamt-mRNA-Spezies beim Wachstum der Zellen auf Minimalmedium um mehr als den Faktor 3 erhöht (Tabelle 8)

Tabelle 8: Zusammenfassung von stark exprimierten Genen beim Wachstum auf Minimal-medium im Vergleich zum Komplexmedium (siehe auch Gesamttabelle Anhang C)

Gen-Nr.	Verhältnis der Mediane	Annotation	Funktion
555	0,103	Glycosyl-Transferase	Biosynthese
2199	0,119	Asparagin-Synthetase	Biosynthese
2899	0,136	Ribonucleosid-Diphosphat-Synthase	Biosynthese
1761	0,174	Pyruvat-Synthase, delta-Untereinheit	Biosynthese
2858	0,201	Na ⁺ /Prolin-Symporter	Transport
2428	0,21	Phosphat-Transportprotein	Transport
1760	0,212	Pyruvat-Synthase, gamma-Untereinheit	Biosynthese
3186	0,213	3-Isopropylmalat-Dehydratase	Biosynthese
2250	0,215	Phosphoribosylformylglycinamidin-Synthase	Biosynthese
3247	0,225	Transkriptionsregulator (PBSX Familie)	Regulation
1850	0,225	Oligopeptid-Transporter, Permease	Substrattransport
2435	0,243	ABC-Transporter, ATP-Bindeprotein	Transport
2063	0,244	Transkriptionsregulator (ArsR Familie)	Regulation
3829	0,245	2-Isopropylmalat-Synthase	Biosynthese
2520	0,249	Xanthin-Guaninphosphoribosyltransferase	Biosynthese
2165	0,249	Glycogen-"debranching"-Enzym	Speicherstoffnutzung
1366	0,256	Malat-Dehydrogenase	Biosynthese
1586	0,262	Citrat-Synthase	Biosynthese
1560	0,269	Transkriptionsregulator	Regulation
2927	0,271	Dipeptid/Oligopeptid-Transporter	Substrattransport
260	0,275	ABC-Transporter, Permease	Transport
1544	0,281	Transkriptionsregulator	Regulation
565	0,285	Transport-Protein	Transport
1774	0,286	Response Regulator	Regulation
1860	0,286	Transkriptionsregulator (AraC-Familie)	Regulation
2880	0,291	ABC-Transporter, Permease	Transport
1123	0,293	Transkriptionsregulator	Regulation
1751	0,294	5-Methyltetrahydropteroyltriglutamat-Homocystein-Methyltransferase	Biosynthese
2879	0,297	Transkriptionsregulator	Regulation
1713	0,299	Transkriptionsregulator	Regulation
188	0,302	Transkriptionsregulator	Regulation
2093	0,305	Polysaccharide- ABC-Transporter	Substrattransport
269	0,306	Aspartate-Semialdehyd-Dehydrogenase	Biosynthese
76	0,306	Dihydroorotat-Dehydrogenase	Biosynthese
2898	0,307	ABC-Transporter, Permease	Transport
1541	0,308	ABC-Transporter, ATP-Bindeprotein	Transport
77	0,313	Dihydroorotat-Dehydrogenase	Biosynthese
1848	0,314	Oligopeptid-Transporter	Substrattransport
1936	0,316	ABC-Transporter, Permease	Transport
94	0,332	Mevalonatkinase	Biosynthese
3197	0,365	Pyruvat-Synthase, alpha Untereinheit	Biosynthese
484	0,367	Fructose-1,6-bisphosphatase	Biosynthese
3196	0,371	Pyruvat-Synthase, beta Untereinheit	Biosynthese

Viele Gene mit einer höheren Expression beim Wachstum auf Minimalmedium kodieren für Enzyme, die für Biosynthesen von Aminosäuren und Nukleotiden verantwortlich sind. Diese höhere Menge an Transkripten bei Wachstum auf Minimalmedium war zu erwarten, da im Medium keine Aminosäuren und Nukleotide vorliegen, weshalb diese *de novo* von den Zellen synthetisiert werden mussten. Zudem wurde deutlich, dass zahlreiche andere Gene beim Wachstum auf Minimalmedium vermehrt exprimiert wurden, die für Proteine kodieren, die wahrscheinlich an der Speicherstoffnutzung, dem Transport oder an der Regulation des Stoffwechsels beteiligt sind.

Hingegen waren nur wenige Gene beim Wachstum auf Vollmedium stärker exprimiert. Tabelle 9 zeigt die Gene, welche nach Wachstum auf Vollmedium eine höhere Expression als beim Wachstum auf Minimalmedium aufwiesen. Mit Ausnahme des Feo-Systems (siehe unten) handelte es sich im Wesentlichen um konservierte- bzw. hypothetische Gene, denen bislang keine Funktion zugewiesen werden konnte. Angegeben sind die Nummern der ORFs, die aus den Fluoreszenzsignalen errechneten Verhältnisse der Transkripte (Vollmedium/Minimalmedium) und die Annotationen.

Tabelle 9: Stark exprimierte Gene beim Wachstum auf Vollmedium im Vergleich zum Minimalmedium

Gen-Nr.	Verhältnis der Mediane (635/532)	Annotation
3396	27.299	Konserviertes Protein
657	25.995	Hypothetisches Protein
656	9.457	Fe ³⁺ -Transport-Protein, FeoA
3131	8.520	Konserviertes Protein
3077	6.132	Konserviertes Protein
3228	5.414	Konserviertes Protein
2426	4.164	Phosphat-Bindeprotein
3076	3.729	Konserviertes Protein
627	3.274	Amidotransferase
3979	3.216	Hypothetisches Protein
3734	3.106	Hypothetisches Protein

Als Beispiel für eine regulierte Operon-Struktur ist in Abbildung 18 das Ergebnis für das *feoA/B*-Operon aufgeführt. Dieses System ist in zahlreichen Bakterien für die Aufnahme von zweiwertigen Eisenionen verantwortlich. Das *feo*-Operon aus *Ms. mazei* besteht wahrscheinlich aus vier Genen, wobei die Gene 0656 und 0657 für kleine Proteine kodieren (10 kDa und 9 kDa), die dem FeoA Protein aus *E. coli* sehr ähnlich sind. Das FeoB Protein (kodiert durch das Gen 0658) hat eine berechnete Größe von 73,4 kDa und stellt offensichtlich ein transmembranes Protein dar, das den eigentlichen Transport von Eisenionen über die Membran katalysiert (Marlovits et al., 2002). Wie in Abb. 18 zu erkennen, waren die Transkriptmengen der Gene *feoA1* und *feoA2* beim Wachstum auf Komplexmedium stark erhöht. Dies traf auch für den ORF 3396 zu, der für ein konserviertes Protein in *Methanosarcina*-Stämmen kodiert. Das Gen gehört somit offensichtlich zum Feo-System. Im Fall von *feoB* war nur eine leichte Erhöhung der mRNA-Konzentration zu beobachten. Die geringe Transkriptmenge von *feoB* im Vergleich zu *feoA1/2* ist verständlich, da das entsprechende Protein als Permease fungiert und eine übermäßige Bildung dieses Membranproteins von der Zelle vermieden werden muss. Das Gen 660 kodiert offensichtlich für ein eisenabhängiges Regulatorprotein. In diesem Fall wurde ersichtlich, dass das relative Transkriptionsniveau beim Wachstum auf Minimalmedium geringer war als bei der Kultivierung der Zellen auf Komplexmedium. Es ist daher wahrscheinlich, dass der Regulator als Repressor der Transkription des *feo*-Operons dient.

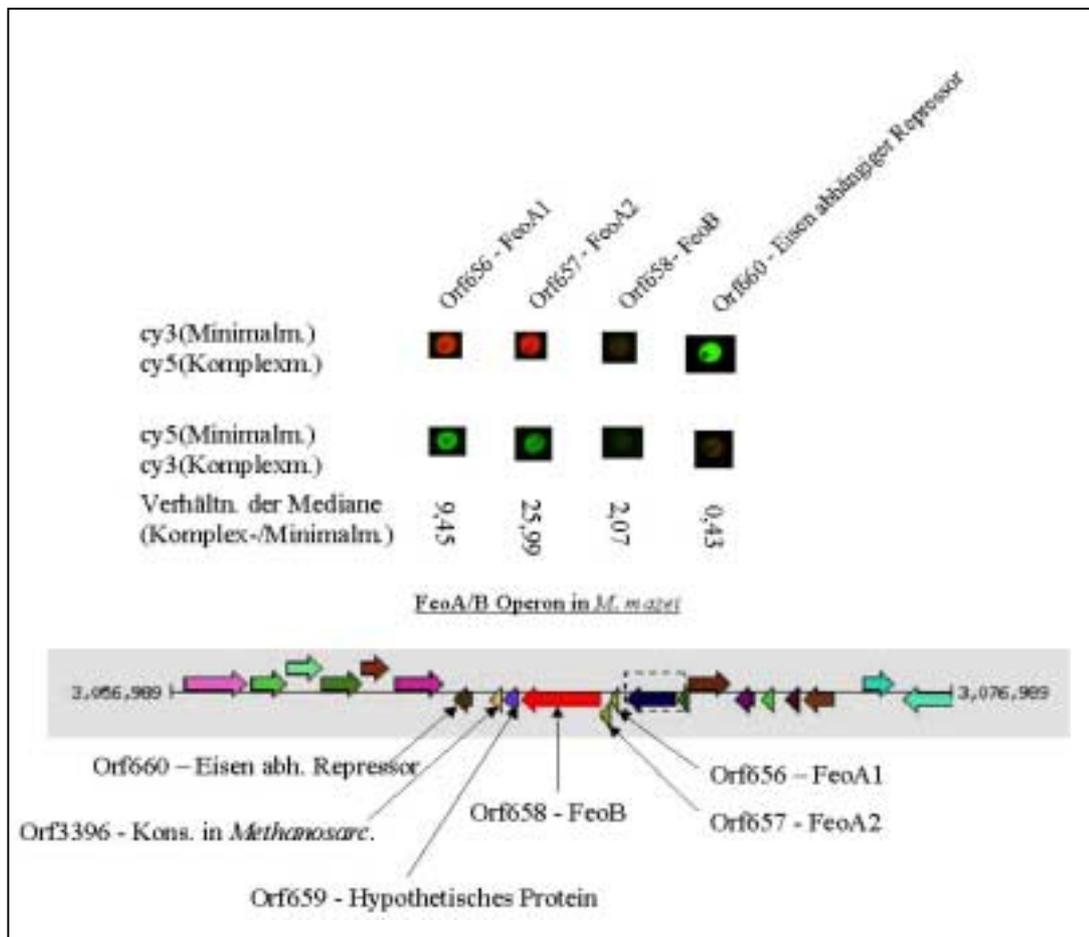


Abb. 18: Expressionsprofil des *feoA/B* Operons beim Wachstum auf Methanol bzw. Acetat

Diese Abbildung stellt einen Teil der Ergebnisse der Transkriptionsanalyse der Vorversuche beim Wachstum auf Voll- und Minimalmedium dar. Die Annotation der Gene entspricht der ERGO-Datensuite nach diesen Experimenten. Das Verhältnis der Mediane wurde über alle Experimente errechnet. Die dargestellten Spots entsprechen den Spots eines einzelnen Teilerperimentes. ORF 659 war zu diesem Zeitpunkt noch nicht auf dem Chip repräsentiert.

Das *feo*-System wurde intensiv in *E. coli* untersucht (Kammler, 1993). Wie oben erwähnt, dient es in diesem Organismus als Aufnahmesystem für Fe^{2+} -Ionen (Kammler, 1993). Das Operon wird durch das Fur-Protein (Ferric-Uptake-Regulation), welches mit gebundenem Fe^{2+} an die sogenannte ‚Iron-Box‘ nahe der Pribnow-Box bindet, reguliert. Es wurde nachgewiesen, dass das Fur-Protein mit gebundenem Fe^{2+} verschiedene Gene in *E. coli* reguliert, welche unabhängig von der Eisenaufnahme sind. Dies sind z.B. Gene, die für Proteine kodieren, die an der Chemotaxis, der Purin-Biosynthese, der Methionin-Biosynthese, dem TCA-Zyklus, der Respiration und der Redox-Stressantwort beteiligt sind. Daher wird das Fur-

Protein als globaler Regulator angesehen (McHugh et al., 2003; Touati, 1988; Vassinova und Kozyrev, 2000; Park und Gunsalus, 1995; Stojiljkovic et al., 1994). Ferner wurde gezeigt, dass die Repression von 52 und die Induktion von mindestens 46 Genen in *E. coli* durch Fur-Fe²⁺ stattfindet und dass der Einfluss auf die Purinbiosynthese (PurR-Repressor) nur beim Wachstum auf Minimalmedium vorliegt (McHugh et al., 2003).

3.3.3. Erkenntnisse aus den DNA-Microarray Daten von *Ms. mazei* Gö1 nach Wachstum auf Voll- bzw. Minimalmedium

Es wurden DNA-Arrays produziert die es ermöglichten, die Menge an Transkripten des kompletten Satzes an Genen von *Ms. mazei* zu untersuchen. Es wurden relative Expressionsprofile von Zellen erstellt, welche mit Voll- bzw. Minimalmedium wuchsen und in der späten logarithmischen Phase geerntet worden waren. Die Verhältnisse der Mengen an Transkripten aller Protein-kodierenden Gene (cds) wurden gemessen, und die Analyse dieser Verhältnisse ließen auf die physiologischen Unterschiede der beiden Zellkulturen unter den verschiedenen Wachstumsbedingungen schließen. Die Zellen wuchsen im Vollmedium schneller, und ein Großteil des Transkriptes lag in signifikant größerer Menge unter dieser Bedingung vor. Das Ergebnis entsprach den Daten vorhergehender Experimente mit *E. coli*- Zellen, die unterschiedliche Wachstumsraten und Geschwindigkeiten der Proteinbiosynthese zeigten (Tao et al., 1999). Im Fall von *Ms. mazei*-Kulturen, die auf Minimalmedium gezogen worden waren, wurde eine erhöhte Expression vieler Gene beobachtet, deren zugehörige Proteine an der Synthese von Zellbausteinen beteiligt sind. Hierbei seien besonders Gene und ihre zugehörigen Proteine erwähnt, die an der Synthese von Aminosäuren und Nukleotiden beteiligt sind.

Durch die in dieser Arbeit etablierten Methoden der Fluoreszenzsonden-Generierung und DNA-Microarray-Chip-Herstellung war somit eine verlässliche Möglichkeit einer Genom-weiten Transkriptionsanalyse von *Ms. mazei* Gö1 gegeben. Ferner wurden einige Daten dieser Experimente dazu verwendet, fragliche Annotationen in der ERGO-Datenbank zu verbessern. So zeigten die Analysen, dass die ORFs des *feo*-Genclusters in gleicher Weise unter den gegebenen Wachstumsbedingungen

exprimiert wurden. Diese Erkenntnis lies den Schluss zu, dass bestimmte Gene (MMZ656 - MMZ660) dem *feo*-Operon zugeordnet werden mussten, die vorher noch als ‚hypothetical Protein‘ annotiert gewesen waren.

3.4 Analyse der Gen-Expressionsmuster von *Ms. mazei* beim Wachstum auf verschiedenen Kohlenstoffquellen

Die Biochemie der Methanogenese wurde in den vergangenen Jahrzehnten im Detail untersucht, und viele wissenschaftliche Ergebnisse wurden hierzu veröffentlicht. Allerdings ist bis heute nur sehr wenig über die Regulation der Stoffwechselwege, welche zur Methanbildung aus verschiedenen Substraten führen, bekannt. Durch die Sequenzierung des Genoms von *Ms. mazei* Gö1 und durch die DNA-Chip Technologie ist es nun möglich, diese Wege im Detail zu analysieren. In diesem Teil des Projekts sollte die Expression aller Gene von *Ms. mazei* Gö1 beim Wachstum auf unterschiedlichen Kohlenstoffquellen (Methanol und Acetat) untersucht werden. Der Schwerpunkt sollte hierbei auf der Betrachtung der Expression bekannter Gene des methylotropen und des aceticlastischen Stoffwechselweges der Methanogenese liegen.

Von besonderem Interesse waren hierbei die Gene, die für folgende Proteine kodieren: 1) Energie-konservierende Enzyme (z. B. Membran gebundene Hydrogenasen, die $F_{420}H_2$ -Dehydrogenase, die Heterodisulfid-Reduktase, Membran gebundene Methyltransferasen und die ATP Synthase) 2) Enzyme des zentralen Stoffwechselweges des Methanogenese (z.B. die Formyl-Methanofuran-Dehydrogenase, die Methylen- H_4 MPT-Dehydrogenase und Reduktase sowie die Methyl-CoM-Reduktase) 3) Proteine, die spezifisch für einen bestimmten Weg der Methanogenese sind (z.B. lösliche Methyltransferasen, CO-Dehydrogenase, Acetat-aktivierende Enzyme und Hydrogenasen) 4) Enzyme, die an den Hauptstoffwechselwegen des Kohlenstoffmetabolismus beteiligt sind (z.B. Gluconeogenese, anaplerotische Reaktionen, Aminosäure- und Nukleotidbiosynthese).

Viele dieser Gene liegen in mehr als einer Kopie auf dem Chromosom von *Ms. mazei* Gö1 vor. Daher stellte sich auch die Frage, welches Gen bzw. Operon unter welcher Wachstumsbedingung aktiviert wird. Weiterhin wurde erwartet, dass durch den Vergleich mehrerer DNA-Array-Analysen ORFs anhand ihres Expressionsverhaltens klassifiziert werden können, denen bis heute keine Funktion zugewiesen werden konnte. Wenn Gene ein sehr ähnliches Expressionsverhalten mit bereits bekannten Genen zeigen, ist es naheliegend, dass diese ORFs in vergleichbare Stoffwechselwege involviert sind.

3.4.1 Expressionsanalyse von *Ms. mazei* Gö1 beim Wachstum auf Methanol bzw. auf Acetat

Abhängig von der Art der Kohlenstoff- und Energiequelle, benutzt *Ms. mazei* Gö1 entweder den CO₂ reduzierenden, den methyloptropen, oder den acetoclastischen Weg der Methanogenese (siehe 4.1.2; Deppenmeier et al., 2002b). Basierend auf der DNA-Microarray-Technologie sollte untersucht werden, welche Gene in dem jeweiligen Stoffwechselweg aktiv sind, wie diese Gene reguliert werden und welche regulatorischen Netzwerke daran beteiligt sind. Es bestand weiterhin die Möglichkeit, viele unterschiedlich exprimierte Gene zu erfassen, welche spezifisch für einen Stoffwechselweg der Methanogenese sind. Ferner sollten Gene identifiziert werden, die für akzessorische Proteine des Elektronentransfers und Enzyme zur Cofaktorbiosynthese und Metalleinbindung kodieren. Diese Informationen sollten zu einem besseren Verständnis des Metabolismus des methanogenen Organismus führen.

Zur Bestimmung der genomweiten Transkriptionsmuster von *Ms. mazei* beim Wachstum auf Methanol im Vergleich zu Acetat wurden die Zellen wie unter 2.3.1 beschrieben angezogen, die Gesamt-RNA isoliert (2.4.5.2) und markierte cDNA hergestellt (2.6.1). Mit diesen Fluoreszenzsonden wurde der DNA-Array-Chip hybridisiert (2.8.2) und die Expressionsdaten erfasst.

Auf eine Justierung der Laser über die Score-Card wurde in diesen Experimenten verzichtet, da die Übereinstimmung der Expressionen der Transkriptome in diesem Fall höher war, und da die Gesamtmenge an repräsentierten Genen mit 95% auf den Chips höher lag. Die Lasereinstellung wurden in diesen Experimenten über die Normalverteilung aller Signale, die über dem Hintergrund lagen, vorgenommen. Diese Methode ist exakter als ein Abgleich über die ScoreCard und liefert eine höhere Reproduzierbarkeit. Die Methode darauf beruht, dass die Signale in beiden Fluoreszenzkanälen einer berechenbaren Verteilung unterliegen (Hedge, 2000; Yeung et al, 2001), besonders wenn nur eine geringe Anzahl von Signalen sich verändert. Dadurch können die beiden Kanäle mathematisch exakt aufeinander abgeglichen werden und die Auswirkungen von unterschiedlichen Template-Konzentrationen wird zusätzlich mit relativiert. Dies war in den Vorversuchen durch die hohe Diskrepanz zwischen den beiden Kanälen nicht möglich (siehe 3.3.1).

Ein Beispiel eines gescannten Chips mit kalibrierten Lasern, ist in Abbildung 19 dokumentiert. Die Abbildung des Chips stellt einen Ausschnitt eines gescannten DNA-Microarray Chips eines Telexperiment dar. Hierbei wurde die cDNA, welche aus der Gesamt-RNA von Acetat-gewachsenen *Ms. mazei* Zellen stammt, mit Cy5 (rot) markiert, und die Methanol-cDNA wurde mit Cy3 (grün) markiert.

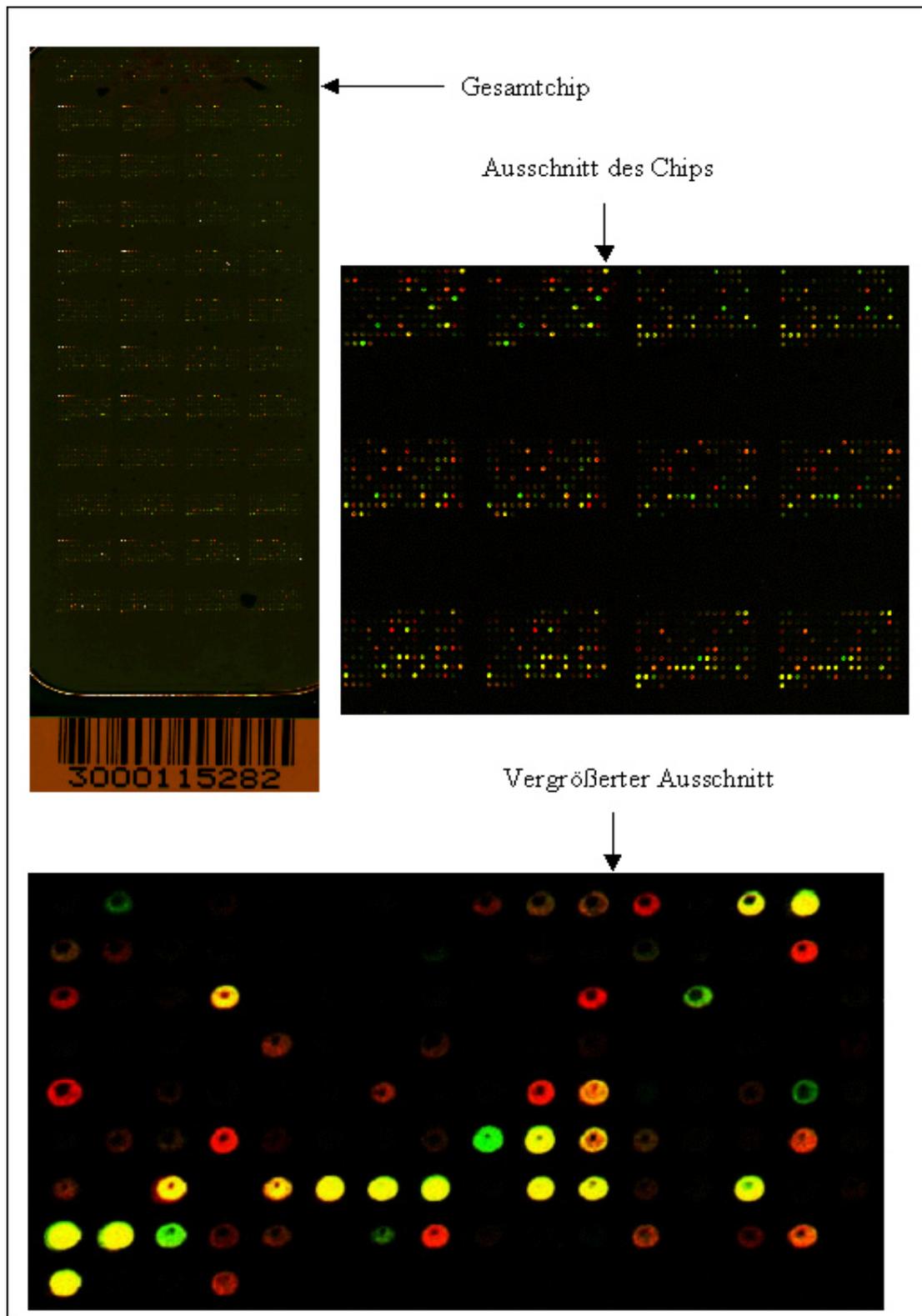


Abb. 19: DNA-Microarray Chip zum Vergleich der Expressionsprofile beim Wachstum von *Ms. mazei* auf Methanol bzw. Acetat

Diese Abbildung zeigt einen gesamten Chip und zwei Ausschnitte eines DNA-Microarray-Chips, bei dem die cDNA von Acetat-Zellen mit Cy5-dCTP(rot) und die cDNA aus Methanol-Zellen mit Cy3-dCTP(grün) markiert wurden. Das Bild wurde mit einer Auflösung von 5 μm pro Pixel gescannt. Die Laser wurden bei diesem Bild schon über die Normalverteilung der Signale justiert.

Die Abbildung verdeutlicht, dass in diesem Experiment die Verteilung der beiden Fluoreszenzkanäle besser und weniger gestreut war als bei der Analyse der Expressionsprofile nach Wachstum auf Komplex- bzw. Minimalmedium, wo die Score-Card als Normalisierungsbasis benutzt wurde (3.3.1). Die Optimierung der Fluoreszenzkanäle ließ sich daran erkennen, dass mehr gelbe Spots (gleiche Expression unter beiden Bedingungen) und ungefähr gleich viele rote (höhere Expression bei Wachstum mit Acetat) und grüne (höhere Expression bei Wachstum mit Methanol) Spots vorhanden waren (Abb. 19). Die erhaltenen Daten aus den Scans wurden anschließend durch die Auswertungssoftware GenePix Pro über die Normalverteilung der Signale in den beiden Fluoreszenzkanälen normalisiert. Hierbei wurde davon ausgegangen, dass der Mittelwert der Mediane über alle Signale gleich eins ist. (Angaben des Herstellers, Fa. Axon). Eine graphische Darstellung der Verteilung der Signale nach der Normalisierung ist in Abb. 20 als loglog-Plot gezeigt.

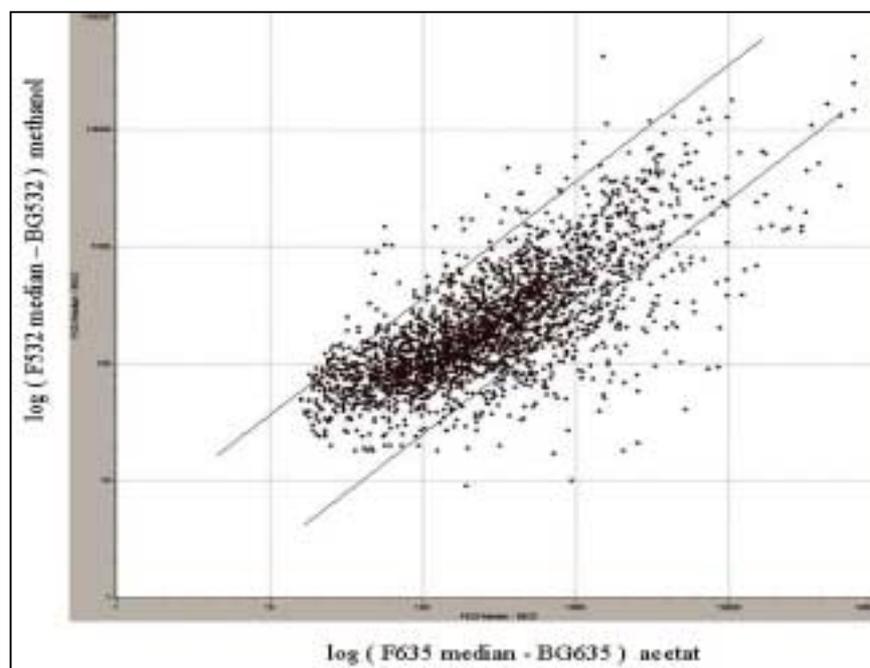


Abb. 20: Plot der Transkriptomdaten nach Einstellung der Laserintensitäten und Normalisierung

Aufgetragen sind die Intensitäten bei 532 nm minus des lokalen Hintergrunds bei 532 nm gegen die Intensitäten bei 635 nm minus des lokalen Hintergrunds bei 635 nm in logarithmischer Form. Die Daten repräsentieren alle Gene, die in beiden Kanälen eine positive Fluoreszenz aufwiesen. In diesem Experiment wurde die cDNA von Acetat-gewachsenen Zellen mit Cy5-dCTP(635 nm) und die cDNA aus Methanol-gewachsenen Zellen mit Cy3-dCTP(532 nm) markiert. Die Trennlinien markieren die Signalverhältnisse 1:3 und 3:1.

Aus dieser logarithmischen Plot-Darstellung lässt sich erkennen, dass sich durch die Normalisierung der Daten nicht alle unterschiedlichen Eigenschaften der beiden Farbstoffe kompensieren ließen. Dies liegt darin begründet, dass eine Normalisierung auf der Gesamtmenge aller Daten beruht. Somit wurde die Datenwolke auf eine Weise verschoben, dass die Achsen halbierende durch den Schwerpunkt der Wolke verlief. Hierbei können Krümmungseffekte, wie sie in dem Plot zu sehen sind, nicht kompensiert werden. Um diese Effekte zu kompensieren, gibt es zwei Möglichkeiten: 1) Eine mathematische Methode, die mit sogenannten mit sogenannten ‚smoothing Algorithmen‘ die Form der Datenwolke an die Achsen halbierende angleicht und 2) eine biologische Methode, die darauf beruht, das Experiment mit vertauschten Farbstoffen zu wiederholen. Bei der biologischen Methode wird der Mittelwert der beiden Experimente zur Auswertung der Daten verwendet, da bei einem zweiten Experiment mit vertauschten Farbstoffen die Krümmung der Datenwolke umgekehrt ist. In dieser Arbeit wurde die biologische Methode gewählt, da hierbei zugleich die Datenmenge erhöht wird und nicht genau vorhersehbar ist, ob einzelne Spots im Bereich der Krümmung nicht eventuell doch ein solches Expressionsverhalten aufweisen. Die mathematische Methode wird nur angewandt, wenn die Menge an mRNA ein limitierender Faktor ist, wie z.B. bei Gewebeproben in der medizinischen Diagnostik. Ferner wurden auch bei den Auswertungen dieser Experimente, ‚thresholds‘ wie in den Vorversuchen gesetzt (siehe 3.3.1). Ein Auszug aus der Ergebnistabelle mit signifikant veränderten Gen-Expressionen ist anfolgend aufgeführt. Die relativen Expressionswerte aller Gene sind im Anhang D dokumentiert.

Die Gene, die unterschiedliche Expressionslevel beim Wachstum auf Methanol bzw. Acetat zeigten, ließen sich in verschiedene Klassen einteilen (Tab. 10A/B). Bei der Methanogenese aus Acetat waren insgesamt 118 Gene um mehr als den Faktor drei höher exprimiert im Vergleich zum Wachstum auf Methanol. Die korrespondierenden Proteine erfüllen folgende Funktionen: Enzyme des acetoclastischen Stoffwechselweges der Methanogenese, Proteine für die aromatische Aminosäure-Biosynthese, Methylthiol:Coenzyme M-Methyltransferasen, Aldehyde:Ferredoxin-Oxidoreduktase und Aufnahmesysteme für Kobalt und Dipeptide/Oligopeptide sowie Transkriptionsregulatoren und viele hypothetische und konservierte Proteine.

Tab. 10A: Mittelwert der Expressionsverhältnisse von Acetat- bzw. Methanolgewachsenen Zellen aller Experimente (Tabellenauszug).

ORF-Nr.	Annotierung	Expressions Quotient Ac/Me	Funktion
2720	Aldehyd:Ferredoxin-Oxidoreduktase	10,84	Acetaldehyd-Oxidation
1441	Accessorisches Protein der CO-Dehydrogenase	19,31	Acetat-Verwertung
1180	Acetate-Kinase	11,49	Acetat-Verwertung
0567	CO-Dehydrogenase, alpha-Untereinheit	9,81	Acetat-Verwertung
4019	CO-Dehydrogenase, alpha-Untereinheit	15,40	Acetat-Verwertung
4021	CO-Dehydrogenase, beta-Untereinheit	22,79	Acetat-Verwertung
0571	CO-Dehydrogenase, delta-Untereinheit	13,24	Acetat-Verwertung
1442	CO-Dehydrogenase, delta-Untereinheit	8,90	Acetat-Verwertung
4020	CO-Dehydrogenase, epsilon-Untereinheit	15,63	Acetat-Verwertung
	CO-Dehydrogenase, epsilon-Untereinheit	15,63	Acetat-Verwertung
2802	CO-Dehydrogenase, gamma-Untereinheit	8,00	Acetat-Verwertung
	CO-Dehydrogenase, gamma-Untereinheit	5,05	Acetat-Verwertung
2838	Ech Hydrogenase, Untereinheit F	3,04	Acetat-Verwertung
2841	Ech Hydrogenase, Untereinheit C	4,26	Acetat-Verwertung
2842	Ech Hydrogenase, Untereinheit B	3,96	Acetat-Verwertung
2843	Ech Hydrogenase, Untereinheit A	4,31	Acetat-Verwertung
1181	Phosphotransacetylase	15,78	Acetat-Verwertung
3007	2-Dehydro-3-Desoxyphosphoheptanoat-Aldolase	15,03	Arom. AS-Synthese
3009	3-Dehydroquinate-Dehydratase	18,93	Arom. AS-Synthese
3008	3-Dehydroquinat-Synthase	21,83	Arom. AS-Synthese
2910	Anthranilat-Phosphoribosyltransferase	14,42	Arom. AS-Synthese
0003	Anthranilat-Synthase, Untereinheit I	6,76	Arom. AS-Synthese
0004	Anthranilat-Synthase, Untereinheit II	3,34	Arom. AS-Synthese
2906	Indol-3-Glycerinphosphat-Synthase	21,50	Arom. AS-Synthese
2911	N-5'-Phosphoribosyl)anthranilat-Isomerase	3,65	Arom. AS-Synthese
3011	Prephenat-Dehydrogenase	18,54	Arom. AS-Synthese
3010	Shikimat-5-Dehydrogenase	24,45	Arom. AS-Synthese
2907	Tryptophan-Synthase, beta Untereinheit	8,24	Arom. AS-Synthese
2278	CbiM protein	4,31	Cobalamin-Synthese und Co-Aufnahme
2279	CbiM protein	11,43	Cobalamin-Synthese und Co-Aufnahme
1336	Cobalamin-Biosynthese Protein	18,88	Cobalamin-Synthese und Co-Aufnahme
0414	Cobalt-Transport-System, ATP-Bindeprotein	6,97	Cobalamin-Synthese und Co-Aufnahme
1338	Cobalt-Transport-System, ATP-Bindeprotein	3,77	Cobalamin-Synthese und Co-Aufnahme
1339	Cobalt-Transport-System, Permease	4,99	Cobalamin-Synthese und Co-Aufnahme
0183	Dipeptid/ Oligopeptide ABC-Transporter	3,11	Dipeptid/Oligopeptid-Aufnahme
2930	Dipeptid/Oligopeptid-Transport-System, Bindeprotein	3,04	Dipeptid/Oligopeptid-Aufnahme
2928	Dipeptid/Oligopeptid-Transport-System, Permease	7,40	Dipeptid/Oligopeptid-Aufnahme
0589	Iron(III)-Dicitrate Transport-System, ATP-Bindeprotein	10,69	Eisenaufnahme
0588	Iron(III)-Dicitrate Transport-System, Permease	15,97	Eisenaufnahme

0587	Iron(III)-Dicitrate Transport-System, Substrat-Bindeprotein	3,96	Eisenaufnahme
4244	Methylthiol:Coenzyme M Methyltransferase	11,67	Methylthiol-Verwertung
4245	Methylthiol:Coenzyme M Methyltransferase	5,26	Methylthiol-Verwertung
0837	Transkriptionsregulator	12,20	regulatorische Funktion
1987	Transkriptionsregulator, MarR Familie	7,01	regulatorische Funktion
1334	Hypothetical protein	10,39	unbekannt
2522	Hypothetical protein	6,86	unbekannt
0402	Hypothetisches Protein	3,90	unbekannt
0502	Hypothetisches Protein	5,44	unbekannt
0541	Hypothetisches Protein	18,43	unbekannt
0763	Hypothetisches Protein	6,54	unbekannt
1333	Hypothetisches Protein	4,72	unbekannt
1335	Hypothetisches Protein	21,75	unbekannt
1563	Hypothetisches Protein	3,10	unbekannt
1682	Hypothetisches Protein	8,41	unbekannt
1962	Hypothetisches Protein	5,34	unbekannt
2316	Hypothetisches Protein	3,06	unbekannt
2374	Hypothetisches Protein	5,26	unbekannt
2722	Hypothetisches Protein	8,23	unbekannt
3012	Hypothetisches Protein	3,80	unbekannt
3065	Hypothetisches Protein	3,04	unbekannt
3979	Hypothetisches Protein	3,27	unbekannt
0005	Konserviertes Protein	3,50	unbekannt
0025	Konserviertes Protein	3,38	unbekannt
0196	Konserviertes Protein	4,74	unbekannt
0200	Konserviertes Protein	3,03	unbekannt
0344	Konserviertes Protein	4,39	unbekannt
0345	Konserviertes Protein	3,50	unbekannt
0348	Konserviertes Protein	3,21	unbekannt
0578	Konserviertes Protein	9,22	unbekannt
0872	Konserviertes Protein	6,36	unbekannt
1177	Konserviertes Protein	22,78	unbekannt
1178	Konserviertes Protein	13,95	unbekannt
1332	Konserviertes Protein	5,99	unbekannt
1337	Konserviertes Protein	6,73	unbekannt
1807	Konserviertes Protein	5,53	unbekannt
2215	Konserviertes Protein	5,07	unbekannt
2216	Konserviertes Protein	3,65	unbekannt
2369	Konserviertes Protein	5,66	unbekannt
2577	Konserviertes Protein	3,83	unbekannt
2724	Konserviertes Protein	4,17	unbekannt

Die vollständige Tabelle ist im Anhang D gezeigt.

Zu den Genen, die beim Wachstum auf Methanol um mindestens den Faktor drei im Vergleich zu Acetat-Zellen hochgefahren wurden (insgesamt 53 Gene), zählten solche die für die folgenden Proteine kodieren: Enzyme des oxidativen Zweigs der Methanbildung aus Methanol, die Untereinheiten der $F_{420}H_2$ -Dehydrogenase, Methyltransferasen zur Verwertung von C_1 -Substraten, Transkriptionsregulatoren, hypothetische und konservierte Proteine.

Tab. 10B: Mittelwert der Expressionsverhältnisse von Acetat- bzw. Methanol gewachsenen Zellen aller Experimente (Tabellenauszug).

ORF-Nr.	Annotierung	Expressions Quotient Ac/Me	Funktion
0739	F ₄₂₀ H ₂ -Dehydrogenase, Untereinheit B	0,12	Elektronentransport
0740	F ₄₂₀ H ₂ -Dehydrogenase, Untereinheit C	0,16	Elektronentransport
0741	F ₄₂₀ H ₂ -Dehydrogenase, Untereinheit D	0,06	Elektronentransport
2378	F ₄₂₀ H ₂ -Dehydrogenase, Untereinheit F	0,11	Elektronentransport
0747	F ₄₂₀ H ₂ -Dehydrogenase, Untereinheit L	0,14	Elektronentransport
0748	F ₄₂₀ H ₂ -Dehydrogenase, Untereinheit M	0,13	Elektronentransport
0749	F ₄₂₀ H ₂ -Dehydrogenase, Untereinheit N	0,07	Elektronentransport
0750	F ₄₂₀ H ₂ -Dehydrogenase, Untereinheit O	0,08	Elektronentransport
2377	N5-N10-Methylen-H ₄ MPT-Dehydrogenase	0,05	Methanol-Verrwertung
1781	Formyl-MFR:H ₄ MPT-Formyltransferase	0,20	Methanol-Verrwertung
0333	Dimethylamin:Corrinoid-Protein	0,14	Methyltransferase
0294	Methanol:Corrinoid Methyltransferase Mta	0,05	Methyltransferase
0295	Methanol-Corrinoid Protein MtaC	0,31	Methyltransferase
0831	Methylcobalamin: Coenzyme M Methyltransfase	0,08	Methyltransferase
0336	Trimethylamin:Corrinoid-Protein	0,29	Methyltransferase
1964	Transkriptionsregulator, MarR Familie	0,11	regulatorische Funktion
0032	Hypothetisches Protein	0,30	unbekannt
1219	Hypothetisches Protein	0,28	unbekannt
1527	Hypothetisches Protein	0,16	unbekannt
1718	Konserviertes Protein	0,30	unbekannt
1747	Konserviertes Protein	0,28	unbekannt
3015	Konserviertes Protein	0,13	unbekannt
3240	Konserviertes Protein	0,31	unbekannt

Die vollständige Tabelle ist im Anhang D gezeigt.

Tabelle 10 stellt nur einen Auszug aus den Gesamtergebnissen dar. In der ersten Spalte ist die ID-Nummer des Gens dargestellt, und in zweite Spalte enthält die Annotation aus der ERGO-Data-Suite. Die dritte Spalte gibt den Median aller Verhältnisse der einzelnen Teilexperimente wieder. Die einzelnen Verhältnisse wurden durch Division der Expression beim Wachstum auf Acetat durch die Expression beim Wachstum auf Methanol errechnet. Hierbei wurden zuerst die Verhältnisse der Mediane der acht einzelnen Experimente (vier Wiederholungen mit jeweils umgekehrten Markierungen) berechnet. Aus den erhaltenen Verhältniswerten wurde das geometrische Mittel (nicht das arithmetische Mittel) errechnet. Ein Wert von größer drei definierte eine signifikant höhere Expression beim Wachstum auf Acetat, und ein Wert kleiner als 0,33 eine signifikant höhere Expression beim Wachstum auf Methanol.

Aus den Daten wird ersichtlich, dass sehr viele Gene, entsprechend der Beteiligung der korrespondierenden Proteine an bestimmten Stoffwechselwegen, reguliert waren. Für den acetoclastischen Weg der Methanogenese waren wie erwartet die Gene hoch exprimiert, die für Enzyme der Acetat-Aktivierung wie z. B. Acetat-Kinase, Phosphotransacetylase und Acetyl-CoA-Synthetase kodieren. Dies galt auch für die CO-Dehydrogenase/Acetyl-CoA-Synthase, die beim Wachstum mit Acetat für die Einspeisung von Acetyl-CoA in die Methanogenese benötigt wird. In Abb. 21 werden diese Zusammenhänge verdeutlicht.

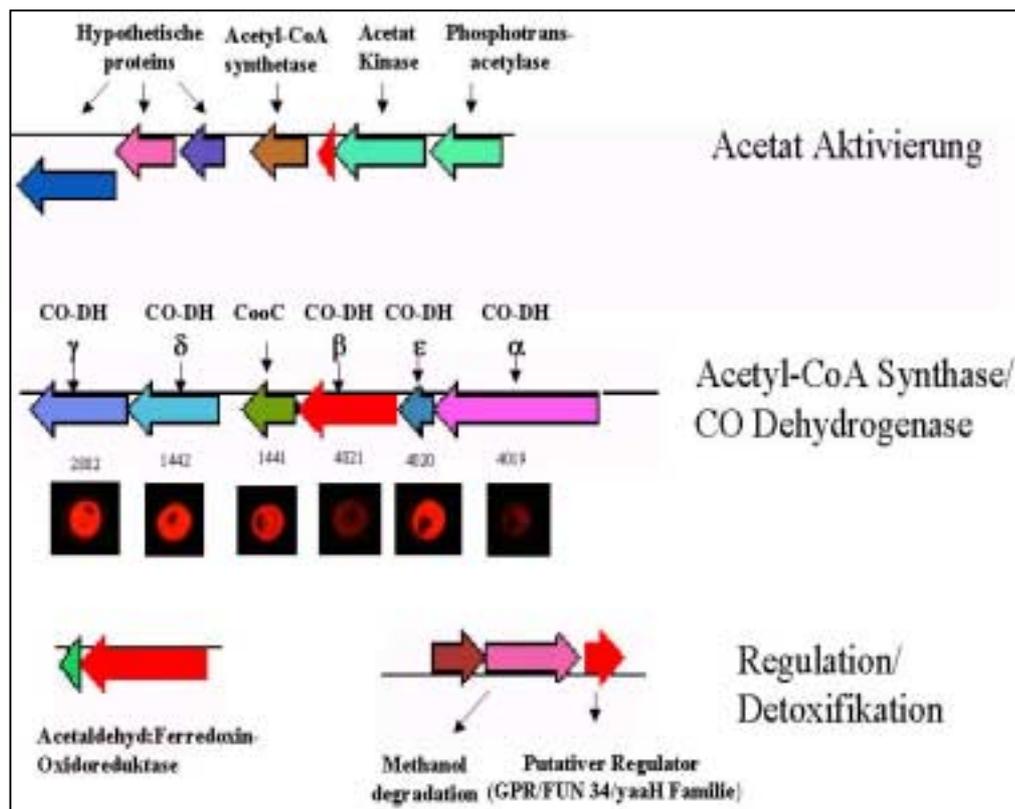


Abb. 21: Beispiele von verändert exprimierten Genen

Die dargestellten Spots stammen von einem Chip, bei dem die cDNA aus Acetat-Zellen mit Cy5 markiert wurde. Die Annotationen stammen aus der ERGO-Data-Suite.

In Abb. 21 sind einige Gene dargestellt, die beim Wachstum von *Ms. mazei* auf Acetat im Vergleich zu Methanol eine höhere Expression zeigten. Darunter befanden sich die Gene 1176-1181 und 3785, die offensichtlich ein Operon darstellen. Die Funktion der Acetat-Kinase, der Phosphotransacetylase und der Acetyl-CoA-Synthetase im Zuge des acetoclastischen Weges der Methanogenese wurde bereits erwähnt. Es finden sich jedoch noch andere ORFs in diesem Cluster, deren Proteine als hypothetisch eingestuft wurden. Durch die Expressionsanalyse kann jetzt jedoch davon ausgegangen werden, dass die Proteine eine Funktion im Acetat-Stoffwechsel erfüllen. In Abb. 21 sind außerdem die Spots gezeigt, die die Gene der Acetyl-CoA Synthase/CO-Dehydrogenase repräsentieren. Das Enzym wird von fünf Strukturgenen kodiert. Zwischengelagert ist zudem ein Gen, das für ein Protein zum Einbau von Nickel in die CO-Dehydrogenase kodiert. Interessanterweise enthält das *Ms. mazei* Genom zwei fast identische Kopien dieser Operonstruktur (neben dem in der Abb. 21 dargestellten Operon bilden die ORFs 567-572 ein zweites Operon). Alle Gene der beiden Operons wurden in gleicher Weise exprimiert, wobei die Transkriptmenge beim Wachstum auf Acetat deutlich höher war als in Methanol-gewachsenen Zellen.

Neben den zu erwartenden Expressionsmustern wurden jedoch auch eine Reihe von neuen Erkenntnissen gewonnen. So fiel z.B. ein Gencluster auf, das offensichtlich für eine Aldehyd:Ferredoxin-Oxidoreduktase kodiert. Die Konzentration an mRNA dieses Gens in Acetat-gewachsenen Zellen war im Vergleich zu Methanol-Zellen deutlich erhöht. Diese Ergebnis legt den Schluss nahe, dass dieses Enzym im Acetat-Stoffwechsel eine wichtige Funktion hat. Es kann vermutet werden, dass die Aldehyd:Ferredoxin-Oxidoreduktase als eine Art Detoxifizierungssystem arbeitet. Im Zuge der Methanbildung aus dem C₂-Körper kann es zur Freisetzung von toxischem Acetaldehyd kommen. Die Funktion der Oxidoreduktase würde damit darin bestehen, Acetaldehyd zu Acetat zu oxidieren (Andrade et al, 2003), das dann wieder in den Stoffwechsel eingespeist wird. Zudem wurde ein interessantes regulatorisches Protein entdeckt, dessen Gen im Bereich eines Gencluster zur Kodierung von löslichen Methyltransferasen lag. Im Gegensatz zu den letztgenannten Genen, war die Transkriptmenge des Gens zur Synthese des Regulatorproteins beim Wachstum auf Acetat stark erhöht. Daraus lässt sich schließen, dass der Regulator als Repressor der Methyltransferase-Gene beim Wachstum auf Acetat fungiert. Dies ist ein

beeindruckendes Beispiel für die Ergebnisse, die mir der Genom-weiten Transkriptionsanalyse erzielt werden können. Nach der Identifizierung eines solchen Regulatorgens, eröffnen sich nunmehr eine Vielzahl von Möglichkeiten, die entsprechenden regulatorischen Netzwerke weiter zu analysieren.

Die verminderte Expression der Gene für die Methylen-Tetrahydromethanopterin-Reduktase, die Methylen-Tetrahydromethanopterin-Dehydrogenase und die Formyl-MFR-H₄MPT-Formyltransferase beim Wachstum auf Acetat als Kohlenstoff- und Energiequelle wurde ebenfalls erwartet (siehe Tabelle Anhang D und Kapitel 4.4.1), da die entsprechenden Proteine nur am oxidativen Zweig der Methanogenese bei der Verwertung von Methanol beteiligt sind und beim Wachstum auf Acetat keine besondere Rolle spielen. Ähnliches gilt für die F₄₂₀H₂-Dehydrogenase als Bestandteil des F₄₂₀H₂:Heterodisulfid-Oxidoreduktase-Systems. Das Enzym katalysiert die Oxidation des reduzierten Coenzym F₄₂₀, welches als Intermediat während der Methanogenese aus Methanol auftritt (siehe 4.2.1). Im Zuge des aceticlastischen Weges der Methanbildung wird dieser Cofaktor nicht reduziert und demzufolge wird die F₄₂₀H₂-Dehydrogenase nicht benötigt.

Interessanterweise waren die Gene, die für Proteine der aromatischen Aminosäure-Biosynthese kodieren (z. B. Shikimat-5-Dehydrogenase, 3-Dehydrochinat-Synthase und 3-Dehydrochinat-Dehydratase) beim Wachstum auf Acetat höher exprimiert als bei einer Kultivierung auf Methanol. Auf die genauere biochemische Relevanz dieser und der restlichen Expressionsdaten wird in der Diskussion in dieser Arbeit näher eingegangen (siehe 4.4).

Die aus den Experimenten gewonnenen Daten wurden anschließend in das Softwareprogramm „Expressionist“ geladen und durch Cluster-Algorithmen (Analyst Expressionist, Fa. Genedata, Wien) zu co-regulierten Gruppen zusammengefasst. Das ‚Clustern‘ kann als das Zusammenfassen von Objekten in Gruppen definiert werden. Diese Gruppen enthalten Objekte mit ähnlichen Eigenschaften. Die Cluster-Algorithmen werden in hierarchische und partitionelle Algorithmen unterteilt. Hierarchische Cluster-Algorithmen basieren auf dem Prinzip, dass jede Gruppe, der mehr als zwei Objekten zugeteilt werden, wiederum in weitere Subgruppen unterteilt wird, bis allen Gruppen und Subgruppen nur noch maximal zwei Objekte

zugewiesen sind. Partitionelle Algorithmen erstellen am Anfang eine feste Menge an dynamischen Gruppen, wobei jedes Objekt einer Gruppe zugeordnet wird (Fasulo, 1999). Der am häufigsten verwendete partitionelle Algorithmus ist der sogenannte ‚k-means‘ Algorithmus. Beide Methoden waren in der verwendeten Analyst-Expressionist Software (Fa. Genedata, Wien) verfügbar.

Das Ziel des Clusters bei der Auswertung von DNA-Microarray Experimenten war es, mit Hilfe der Algorithmen die Expressionsprofile von Genen zu Gruppen zusammenzufassen, die ein ähnliches Expressionsverhalten zeigten. Das Gruppieren der Objekte durch Cluster-Algorithmen basiert auf Distanz-Matrix Berechnungen. Der mathematische Abstand zweier Expressionsprofile im n-dimensionalen Raum definiert die Ähnlichkeit der Expression (je kleiner der Abstand, desto ähnlicher ist das Expressionsverhalten). Zur Berechnung des Abstands als Maß der Ähnlichkeit werden in der Bioinformatik meist nur drei unterschiedliche Methoden verwendet.

- 1) Die Berechnung des Manhattan Abstands (auch bekannt unter den Namen L_1 und ‚City-block-metric‘). Hierbei werden die Differenzen aller Koordinaten aufaddiert.
- 2) Der euklidische Abstand (auch als L_2 bezeichnet), welcher als die Quadratwurzel aus dem Quadrat der Koordinatenunterschiede definiert ist.

$$ED = \sqrt{(X_{ik} - X_{jk})^2}$$

X_{ik} und X_{jk} = Messwert der k-ten Variablen unter den Bedingungen i und j.

- 3) Die Berechnung des Pearson-Korrelations-Koeffizienten, die wie folgt definiert ist:

$$\text{Pearson} = \sqrt{\frac{(X_{ik} - X_{jk})^2}{V_k}}$$

X_{ik} und X_{jk} = Messwert der k-ten Variablen unter den Bedingungen i und j.

V_k = Varianz der k-ten Variablen

Der größte Unterschied der Methoden liegt in der Exaktheit der Abstandsberechnung welcher von 1) bis 3) zunimmt. Jedoch nimmt mit einer exakteren Berechnung auch die Rechenzeit des Computers zu. Dies ist, wenn man berücksichtigt, dass die Menge der Berechnungen gleich der Fakultät aller Expressionsdaten über alle Experimente ist, ein signifikanter Nachteil der höheren Genauigkeit. Andere Methoden des Clusters basieren meist auf einer Modifikation dieser Berechnungsarten (z.B. Eisen et al., 1998). Es wurden alle Abstandsberechnungsmethoden im Rahmen dieser Arbeit getestet, und es gab keine signifikanten Abweichungen, weshalb die L_1 -Berechnung (Standardeinstellung der Software) für die Auswertung benutzt wurde. Ein Auszug eines Cluster-Baums ist in Abb. 22 dargestellt.

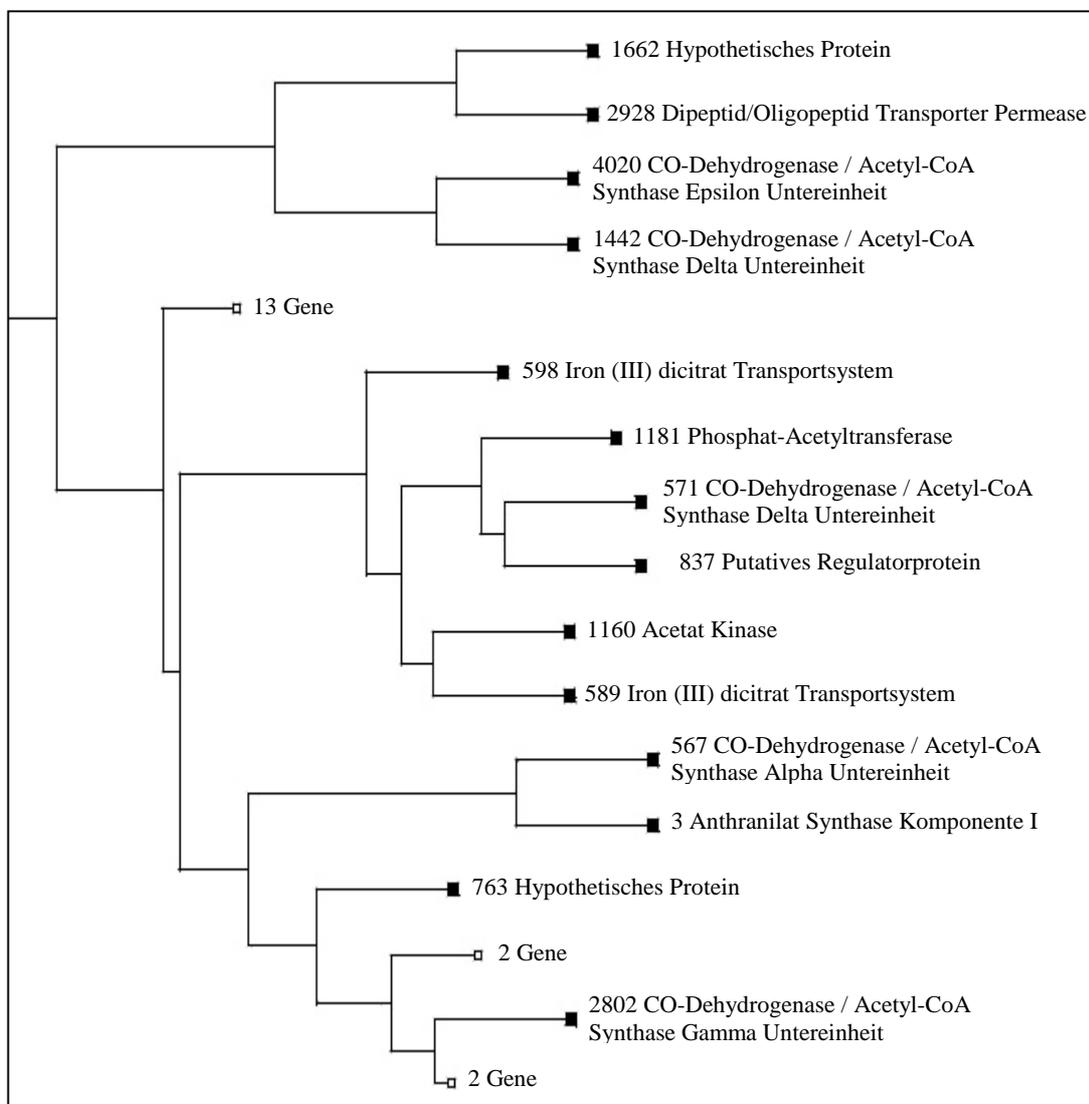


Abb. 22: Teil eines Clusterbaums aller Experimente zur Transkriptionsanalyse von *Ms. mazei*

Aus den gebildeten Gruppen dieses Ausschnitts, lässt sich erkennen, dass die Expressionsprofile für die Gene der Untereinheiten der CO-Dehydrogenase/Acetyl-CoA-Synthase als ähnlich zusammengefasst wurden. Es wurde ein hierarchischer Algorithmus gewählt, weshalb jede Gruppe maximal zwei Objekte enthält. Die Entfernung über die Verzweigungen des Baumes definieren die Höhe der berechneten Ähnlichkeit der Expressionsprofile. Je länger die Strecke zwischen zwei Gruppen, um so niedriger ist die relative Ähnlichkeit im Expressionsverhalten. Aus technischen Gründen ist es nicht möglich gewesen, einen komplett ausgefalteten Baum abzubilden. Jedoch wird aus der Abbildung 22 deutlich, dass die errechneten Ähnlichkeiten der Expressionsprofile der dargestellten Gene sehr hoch war, da in dem gezeigten Ausschnitt nur 30 Gene gruppiert waren. Die restlichen 3341 Gene lagen weiter entfernt.

3.4.2 Quantitative Expressionsanalyse ausgewählter Gene durch Real-Time-PCR

Die Auswertung der Chip-Daten ergab, dass beim Wachstum auf Methanol die Expression vieler Gene höher war als bei der Anzucht auf Acetat und umgekehrt. Allerdings kann die Methode der Transkriptionsanalyse durch DNA-Mikroarrays starken Schwankungen durch folgende Faktoren unterliegen: 1) Kreuzhybridisierungen der markierter cDNA-Sonden 2) unterschiedliche Eigenschaften der Farbstoffe in Bezug auf Fluoreszenz und cDNA-Inkorporation. 3) Cytosin-Gehalt der mRNA. 4) Schwankende Qualitäten der Aminosilan Beschichtungen der leeren Chips. 5) Instabilität der PCR-Produkt-Sonden durch die UV-Bestrahlung beim ‚crosslinking‘. 6) zu schnelles oder zu langsames Eintrocknen der PCR-Produkt-Sonden nach dem Spot-Vorgang führt zu Abweichungen bei der Spotgröße und Form. (Tseng et al., 2001; Caldwell et al., 2001; Lee et al., 2000; Liu et al., 2003 ; Fa. Amersham Biosciences, Schulung zur Transkriptionsanalyse). Diese Faktoren lassen sich durch Computer-Algorithmen, Wiederholungsexperimente oder mit einer reversen Markierung nur begrenzt relativieren. Ferner ist mit der DNA-Microarray-Technologie, vor allem bei niedrigen Expressionswerten, nur eine ungefähre relative Expressionsanalyse möglich; absolute Werte können nicht gemessen werden. Die aus der Auswertung von DNA-Chips errechneten Expressionsverhältnisse weichen oftmals um den

Faktor 2-10 von den realen Verhältnissen ab (Celis et al., 2000; Duggan et al., 1999). Diese Erkenntnisse führten dazu, dass in dieser Arbeit die Real-Time-PCR Methode angewandt wurde, um die Ergebnisse der Microarray Experimente zu überprüfen und zu quantifizieren (Stintzi, 2003; Liu et al., 2003; Caldwell et al., 2001).

Bei dieser Methode diente eine Reverse-Transkriptase zur Umwandlung von mRNA in cDNA. Anschließend wurde durch Hitzebehandlung die Hot-Start Taq-Polymerase aktiviert und die Reverse-Transkriptase deaktiviert. Danach folgten bis zu 45 PCR Zyklen, wobei jeweils zwei ORF-spezifische Primer zum Einsatz kamen. Zur Quantifizierung wurde ein Farbstoff angewandt, der 200fach stärker fluoresziert, wenn er an doppelsträngige DNA gebunden ist (SYBR Green I; Wittwer et al., 1997; Skeidsvoll und Ueland, 1995). Nach jedem Zyklus der PCR wurde die Fluoreszenz gemessen und eine Kurve aus diesen Werten erstellt. Der aus diesen Kurven errechnete c_t -Wert (der Zeitpunkt an dem zuerst eine signifikante Fluoreszenzzunahme detektiert wurde) ermöglicht es, die Ausgangsmenge an mRNA-Template zu berechnen. Hierbei wurde davon ausgegangen, dass in den frühen PCR-Zyklen eine Verdopplung der PCR-Produkte stattfand. Die Konzentrationen an mRNA folgender Gene aus *Ms. mazei*-Zellen, die auf Methanol- bzw Acetat gewachsen waren, wurden analysiert (Tab. 11):

Tabelle 11: Quantitative Expressionsanalyse ausgewählter Gene durch Real-Time-PCR

ORF-Nr.	Annotation	Funktionszuordnung nach Chip-Analyse
2378	F ₄₂₀ -Dehydrogenase, Untereinheit F	Methanolstoffwechsel
0739	F ₄₂₀ -Dehydrogenase, Untereinheit B	Methanolstoffwechsel
2377	Methylen-H ₄ MPT-Reduktase	Methanolstoffwechsel
0294	Methyltransferase MtaB	Methanolstoffwechsel
2842	Ech Hydrogenase, Untereinheit B	Acetatstoffwechsel
1180	Acetat-Kinase	Acetatstoffwechsel
2720	Aldehyd:Ferredoxin-Oxidoreduktase	Acetatstoffwechsel
4019	Acetyl-CoA-Synthase	Acetatstoffwechsel
0919	Methyl-CoM-Reduktase, Untereinheit α	Zentraler Stoffwechsel
0915	Methyl-CoM-Reduktase, Untereinheit β	Zentraler Stoffwechsel

Es wurden die Transkripte von zwei Genen der Untereinheiten der $F_{420}H_2$ -Dehydrogenase gewählt, da diese in verschiedenen Operons auf dem Chromosom lokalisiert sind. Die beiden Transkripte der Gene der Methyl-CoM-Reduktase dienten als Eichwert, da diese unter beiden Bedingungen (Wachstum auf Acetat bzw. Methanol) gleichstark exprimiert wurden. Als Farbstoff wurde SYBR-Green-I gewählt, und die PCR durchlief 40 Zyklen. Die Template RNA wurde in 1:1 und 1:2 Verdünnungen einer 600 ng/ μ l-Lösung in die Real-Time-PCR-Reaktionen eingesetzt. Ferner wurden drei Negativ-Kontrollen durchgeführt. 1) Eine PCR mit Primern für das Transkript des Gens für die Methyl-CoM-Reduktase, Untereinheit α ohne RNA-Template. 2) Eine Reaktion ohne Primer in der PCR mit Gesamt-RNA aus Methanol-gewachsenen Zellen. 3) Eine Reaktion ohne Primer in der PCR mit Gesamt-RNA aus Acetat-gewachsenen Zellen. Wie erwartet zeigten diese Kontrollreaktionen keinen Anstieg der Fluoreszenz während der PCR. Die Kurvenverläufe der Kontrollreaktionen sind in Anhang B dokumentiert. In Abbildung 23 wird am Beispiel der Transkripte für die F-Untereinheit der F_{420} -Dehydrogenase die Auswertung eines Teilerperiments verdeutlicht. Die Grafiken der restlichen PCR-Verläufe sind in Anhang B dokumentiert.

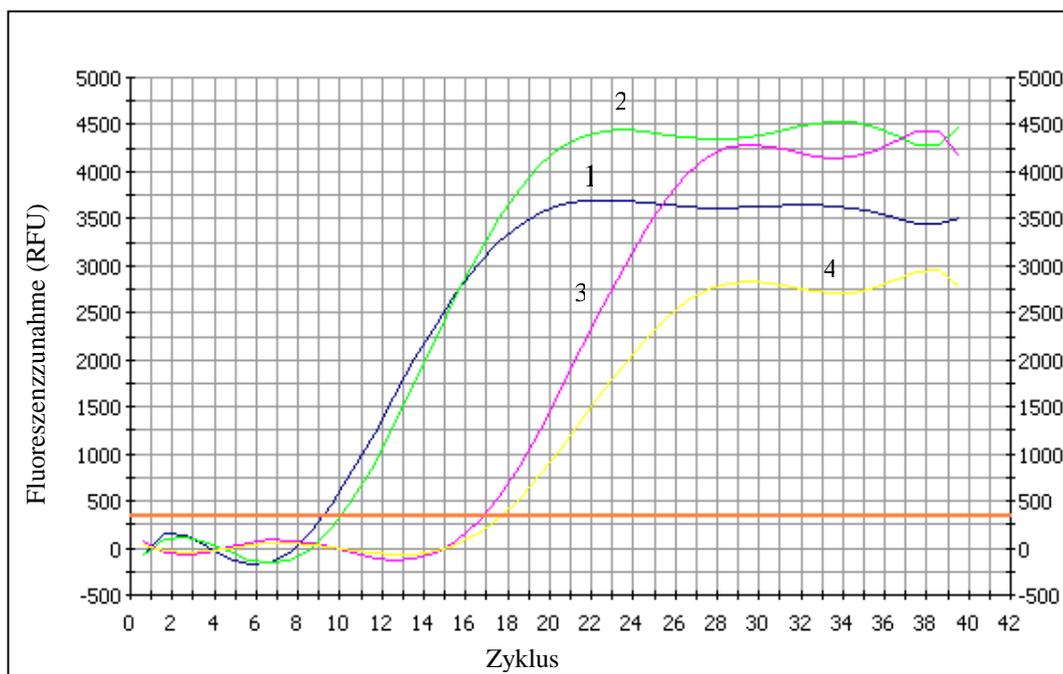


Abb. 23: Fluoreszenzzunahme während der Real-Time PCR für das Transkript des Gens der Untereinheit FpoF der $F_{420}H_2$ -Dehydrogenase.

Abb. 23 zeigt die vier Kurven der Fluoreszenzzunahme während der Real-Time-PCR mit spezifischen Primern für das Transkript des Gens *fpoF*, das für die F-Untereinheit der F₄₂₀H₂-Dehydrogenase kodiert. Als Matrize in der Real-Time-PCR diente Gesamt-RNA in einer 1:1 (Kurve 1) und einer 1:2 Verdünnung (Kurve 2) aus Methanol-gewachsenen Zellen und Gesamt-RNA in einer 1:1 (Kurve 3) und einer 1:2 Verdünnung (Kurve 4) aus Acetat-gewachsenen Zellen. Die beiden Kurven innerhalb des rechten und des linken Paares schnitten jeweils etwa um einen Zyklus versetzt die rote Linie. Die rote Linie ist der Punkt, an dem die erste signifikante Zunahme der Fluoreszenz dieser vier Experimente berechnet wurde. Die Schnittpunkte der Kurven mit der Linie markieren den errechneten c_t-Wert über diese vier Experimente. Da hierbei jeweils 1:2 Verdünnungen verwendet wurden, war zu erwarten, dass die Schnittpunkte der Verdünnungen jeweils einen Zyklus auseinander liegen.

Nach Beendigung des letzten Zyklus der Real-Time-PCR wurde eine Schmelzkurve von allen Reaktionen aufgenommen. Hierbei wird über einen steigenden Temperaturgradienten mit 0.5 °C pro 30 s die Abnahme der Fluoreszenz (RFU = Relative Fluoreszenz Units) gemessen, welche entsteht, wenn die Wasserstoffbrücken der doppelsträngigen PCR-Produkte durch die Temperatur gebrochen werden. Alle Schmelzkurven sollten nur einen Peak besitzen, und die Kurvenverläufe aller Schmelzkurven mit denselben Primern sollten annähernd gleich sein. In Abb. 24 sind als Beispiel die Schmelzkurven für die PCR-Produkte aus den Real-Time-PCR-Reaktionen angegeben, welche mit Primern für die Transkripte der Gene der F-Untereinheit der F₄₂₀H₂-Dehydrogenase generiert wurden. In Anhang B sind die Schmelzkurven aller PCR Produkte dokumentiert.

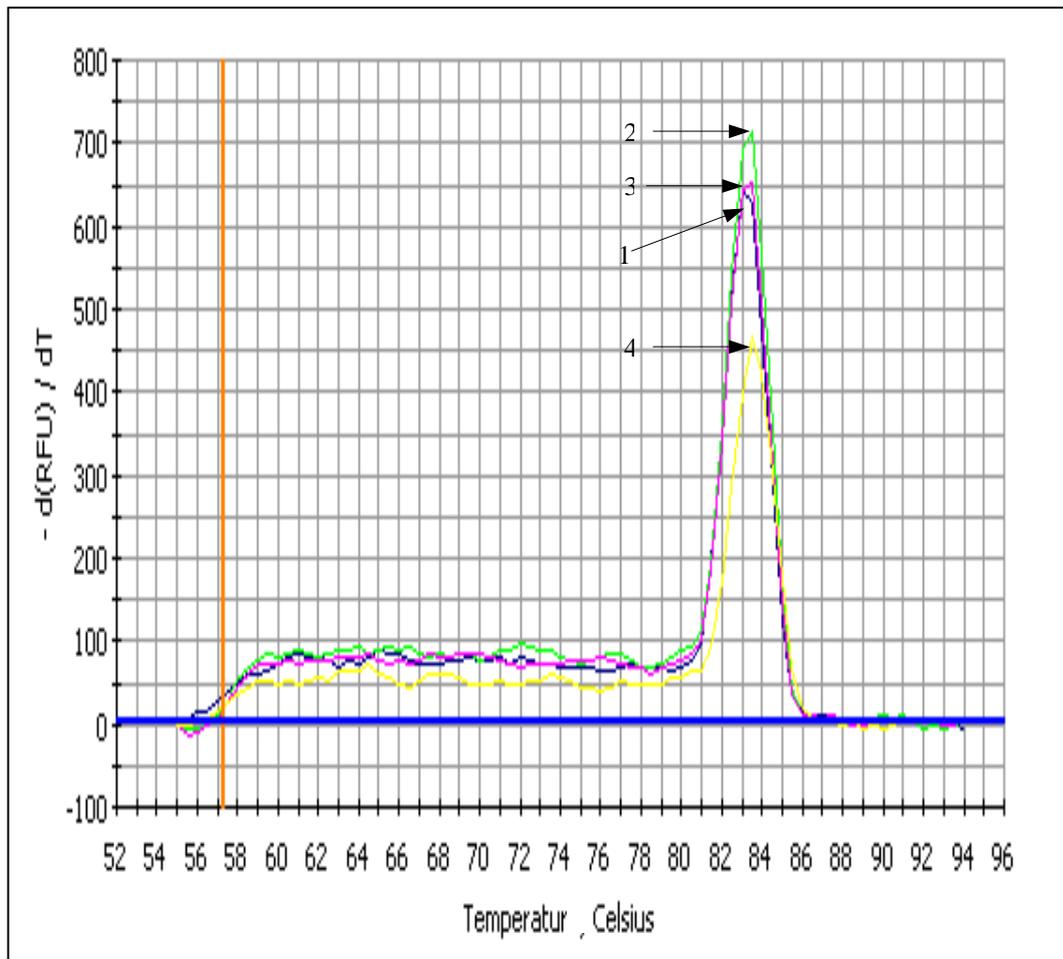


Abb. 24: Schmelzkurven der PCR-Produkte nach 40 Zyklen der Real-Time-PCR mit Primern für das Transkript des Gens *fpoF*

Diese Abbildung zeigt die vier Kurven der Fluoreszenzabnahme in Relation zur Temperatur. Jede Kurve repräsentiert den Verlauf von einer Reaktion mit spezifischen Primern für die mRNA Transkripte der F-Untereinheit der $F_{420}H_2$ -Dehydrogenase. Es lässt sich ersehen, dass alle Kurven nur einen schmalen Peak bei derselben Temperatur haben. Die Höhe der Peaks ist abhängig von dem Sättigungspunkt der Reaktion, welcher in Abb. 23 dargestellt ist.

Wie aus Abb. 24 zu ersehen ist, zeigten die Schmelzkurven einen sehr ähnlichen Verlauf mit nur einem Peak. Dies lässt darauf schließen, dass reine identische PCR-Produkte generiert wurden. Die Höhe der Peaks entsprach den Endkonzentrationen an PCR-Produkten, welche aus Abb. 23 (analoge Nummerierung) zu ersehen sind.

Aus den Kurvenverläufen wurden durch die Auswertungssoftware des Real-Time-PCR Geräts (iCycler, Fa. Biorad) die c_t -Werte errechnet, welche ein Maß für die Menge an eingesetztem Template sind. Hierbei wurden die c_t -Werte über den Gesamtverlauf aller Reaktionen berechnet, weshalb die errechneten c_t -Werte der Software leicht von den Schnittpunkten in den Verläufen der einzelnen Kurven abweichen. Diese errechneten c_t -Werte sind in Tabelle 12 aufgelistet; hierbei wurden die Zahlenwerte der Übersicht halber gerundet.

Tabelle 12: Errechnete c_t -Werte der Real-Time-PCR Reaktionen

ORF	C_t -Wert 1:1 Methanol RNA	C_t -Wert 1:2 Methanol RNA	C_t -Wert 1:1 Acetat RNA	C_t -Wert 1:2 Acetat RNA
Acetyl-CoA-Synthase	11,9	13,2	12,3	12,7
Aldehyd:Ferredoxin- Oxidoreduktase	16,8	17,5	13,4	13,7
Acetat-Kinase	13,2	13,9	13,2	13,9
Ech Hydrogenase, Untereinheit B	12,1	13	14,3	15,2
Methylen- H_4 MPT- Reduktase	9,8	10,5	17	17,5
F_{420} -Dehydrogenase, Untereinheit F	10,3	11,2	18,1	19,3
F_{420} -Dehydrogenase, Untereinheit B	12,1	13,3	19,9	26,2
Methyltransferase MtaB	12,8	8,8	20,3	27,1
Methyl-CoM ReduktaseUE- α	9,1	9,5	10,8	16,5
Methyl-CoM ReduktaseUE- β	8,6	9,3	9,7	15

Mit Hilfe der c_t -Werte ließ sich über die Relation zu den Werten der Methyl-CoM-Reduktase eine absolutes Mengenverhältnis an Template-RNA errechnen. Diese Vorgehensweise war notwendig, da die Konzentrationen der mRNA aus Acetat- bzw. Methanol- gewachsenen Zellen nicht bekannt war. Das Prinzip der Berechnung, das in diesen Auswertungen angewandt wurde, beruht auf der Tatsache, dass die relative Menge an Transkripten für die Untereinheiten der Methyl-CoM-Reduktase in beiden Gesamt- RNA Lösungen gleich war. Aus den unterschiedlichen c_t -Werten der Methyl-CoM-Reduktase-Transkripte ließ sich erkennen, dass die Ausgangsmenge an Transkripten dieser Gene in der Gesamt-RNA Lösung von Acetat-Zellen niedriger war. Dies ist der Grund, warum die Berechnung auf einer Relation der Δc_t -Werte und nicht auf den absoluten Werten beruhte. Die Berechnung erfolgte über eine exponentiale Funktion, da in der Formel davon ausgegangen wird, dass an dem gemessenen Punkt pro PCR-Zyklus eine Verdopplung der PCR-Produkten stattfand.

Zur Berechnung diente folgende Formel:

$$X = \frac{1}{\left(\frac{\text{Exp}(\text{GoI})^{\Delta c_{t_{\text{GoI}}}}(\text{GoI1}-\text{GoI2})}{\text{Exp}(\text{HK})^{\Delta c_{t_{\text{HK}}}}(\text{HK1}-\text{HK2})} \right)}$$

X: n-fache Expression

GoI: Gene of Interest (= zu untersuchendes Gen)

GoI1: c_t -Wert des GoI unter Bedingung 1

GoI2: c_t -Wert des GoI unter Bedingung 2

HK: House-Keeping Gene (= Referenz-Gen)

HK1: c_t -Wert des HK unter Bedingung 1

HK2: c_t -Wert des HK unter Bedingung 2

Die Berechnung des n-fachen Mengenunterschieds an Template erfolgte jeweils mit den beiden Referenz-Genen und den beiden Verdünnungen getrennt. Das arithmetische Mittel der Ergebnisse wurde als absoluter n-facher Mengenunterschied an Template (=Expressionerhöhung) definiert. In Tabelle 13 sind diese absoluten Mengenunterschiede an Template-RNA aus Acetat- und Methanol-Zellen dargestellt.

Tabelle 13: n-fache absolute Mengenunterschiede (=Expressionserhöhung) der Transkripte

ORF	n-fache Expressionserhöhung Methanol-/Acetat-Zellen	n-fache Expressionserhöhung Acetat-/Methanol-Zellen
Acetyl-CoA-Synthase	0,293	3,414
Aldehyd:Ferredoxin- Oxidoreduktase	0,026	38,627
Acetat-Kinase	0,293	3,414
Ech Hydrogenase, Untereinheit B	1,172	0,854
Methylen-H ₄ MPT- Reduktase	37,490	0,027
F ₄₂₀ H ₂ -Dehydro- genase, <i>fpoF</i>	74,981	0,013
F ₄₂₀ H ₂ -Dehydro- genase, <i>fpoB</i>	74,981	0,013
Methyltransferase <i>mtaB</i>	37,490	0,027

Die Abweichungen in den absoluten Werten der Expressionsänderungen zwischen der Real-Time-PCR-Experimenten und den DNA-Chip-Experimenten lagen in der Natur der Microarray-DNA-Chip-Technologie. Abweichungen um bis zu Faktor zehn sind von vielen Arbeitsgruppen beobachtet worden. (Conway und Schoolnik, 2003; Stintzi, 2003; Lee et al., 2001). Es ist daher verständlich, dass die Verwendung von DNA-Microarrays keine Aussage über den absoluten Faktor der Transkriptionsänderung zulässt. Daher muss zur Verifizierung der Daten eine zweite Methode der Quantifizierung von Transkriptmengen vorgenommen werden, wobei im Rahmen dieser Arbeit die Real-Time-PCR eingesetzt wurde. Zusammenfassend kann festgestellt werden, dass die aus der Real-Time PCR erhaltenen Daten über das

Expressionsverhalten der analysierten Gene weitgehend mit den Ergebnissen der Microarray-Chip-Experimente übereinstimmten. Die Real-Time-PCR-Analysen zeigten auch, dass die Gene, die Protein zur Nutzung von Acetat kodieren, beim Wachstum der Kulturen auf diesem Substrat hochexprimiert wurden. Hierunter fielen, wie auch schon in Kapitel 3.4.1 beschrieben, folgende Gene bzw. deren Proteine: Acetat-Kinase, Acetyl-CoA-Synthetase und Acetaldehyd:Ferredoxin Oxidoreduktase.

Für die Transkripte, die in höherer Anzahl in jenen Zellen vorkamen, die mit Methanol als Kohlenstoff- und Energiequelle wuchsen, entsprachen die Werte ebenfalls den Erwartungen. Die Expression der Gene der *fpoF* und *fpoB*, die für Untereinheiten der $F_{420}H_2$ -Dehydrogenase kodieren, war in diesen Zellen höher, da das Enzym nur an der Umsetzung von Methanol beteiligt ist. Die identischen Daten der Real-Time-PCR dieser beiden Transkripte, ließen darauf schließen, dass beide Gene einer gemeinsamen Regulation unterliegen, obwohl die Gene in unterschiedlichen Operonen an unterschiedlichen Positionen des Genoms lokalisiert sind (Bäumer et al., 2000). Die erhöhte Expression des Gens der MtaB Methyltransferase beim Wachstum auf Methanol war ebenfalls zu erwarten gewesen, da dieses Enzym nur an dem Methylgruppentransfer von Methanol zu Coenzym M beteiligt ist, und daher im Acetatstoffwechsel nicht benötigt wird.

Dass die Ech-Hydrogenase kein signifikant verändertes Expressionsverhalten zeigte, ist damit zu begründen, dass diese neben der Funktion im Ferredoxin:Heterodisulfid-Oxidoreduktase-System des Acetatstoffwechsels wahrscheinlich noch weitere Funktionen hat. Es wird inzwischen angenommen, dass das Enzym auch eine Rolle bei der Methanbildung aus Methanol einnimmt, auf die in der Diskussion näher eingegangen wird (4.4.3). Somit wird die Ech-Hydrogenase in beiden Stoffwechselwegen benötigt und die entsprechenden Gene sind sowohl beim Wachstum auf Methanol als auch auf Acetat aktiviert.

3.5 Zusammenfassende Betrachtungen zu den Experimenten und Ergebnissen dieser Arbeit

Die in dieser Arbeit beschriebenen Experimente zeigten, dass die genomweite Transkriptionsanalyse in einer aussagekräftigen Weise zur Charakterisierung des Metabolismus von *Ms. mazei* Göl angewendet werden kann. Die DNA-Array-Technologie ermöglicht es, neue Hypothesen aufzustellen, bereits existierende zu beweisen und den Weg zu weisen, um bestimmte Gene von Interesse zu identifizieren oder genauer zu charakterisieren.

Ferner haben die aus der Cluster-Analyse erhaltenen Daten, neben der Bestätigung der Microarray-Daten, eine weitere Bedeutung, die über den Rahmen dieser Arbeit hinausreicht. Der tiefgreifende Sinn dieser Daten ist, dass sie in einer Oracle Transkriptionsdatenbank des Analyst-Servers jedem Mitarbeiter des Instituts für weitere Microarray-Experimente mit anderen Fragestellungen in leicht zugänglicher Form zur Verfügung stehen. Somit bilden diese Daten den Grundstock für eine zukünftige Datensammlung, die es zu einem späteren Zeitpunkt ermöglichen wird, über viele verschiedene Experimente hinweg neue Co-Regulationen von Genen zu erfassen und damit neue Gene und Verknüpfungen zwischen Stoffwechselwegen schnell und einfach zu detektieren.

Abschließend sei noch auf die bioinformatischen Ergebnisse dieser Arbeit verwiesen, da durch die in dieser Arbeit erstellten Computerprogramme die Etablierung, Handhabung und Auswertung eines genomweiten Transkriptionsanalyseprojektes erst ermöglicht wurde. Somit wird es, basierend auf diese Programme, anderen Arbeitsgruppen erleichtert, eigene Transkriptionsanalyseprojekte durchzuführen.

4. Diskussion

4.1 Betrachtungen zum Stoffwechsel von *Methanosarcina mazei* Gö1

Am Ende der anaeroben Nahrungskette stehen die methanogenen Archaea, zu denen der in der Arbeit untersuchte Organismus *Ms. mazei* Gö1 gehört. Der Stamm setzt die durch gärende- und syntrophe Organismen produzierten Verbindungen H_2/CO_2 , Acetat, Methanol und Methylamine zu CH_4 bzw. zu CH_4 und CO_2 um und führt diese dem globalen C-Kreislauf in gasförmigen Zustand wieder zu. Die Mineralisierung der oben erwähnten Substrate ist ein bedeutender biologischer Prozess zum Abbau von organischem Material unter anaeroben Bedingungen (Zinder, 1993). Es ist daher verständlich, dass die Stoffwechselprozesse zur Methanbildung intensiv untersucht wurden (Ferry, 1992; Thauer et al., 1993; Blaut 1994; Deppenmeier et al., 1996; Thauer, 1998; Deppenmeier, 2002a). In dieser Arbeit wurden vergleichende Untersuchungen zur relativen Expression aller Gene von *Ms. mazei* Gö1 beim Wachstum auf Acetat und Methanol durchgeführt. In den folgenden Kapiteln sollen daher die Prozesse der Methanogenese aus Acetat und Methanol näher beschrieben werden.

4.1.1 Eigenschaften von ungewöhnlichen methanogenen Kofaktoren

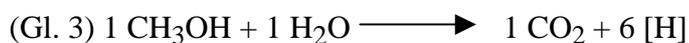
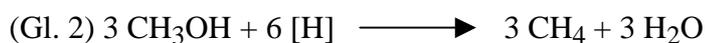
Diverse ungewöhnliche Coenzyme und prosthetische Gruppen wurden entdeckt, die an der Bildung von Methan beim Wachstum auf Acetat bzw. Methanol beteiligt sind (Wolfe, 1985). Die Umsetzung von methanogenen Substraten zu Methan verläuft über C_1 -Intermediate, welche an Methanofuran (MFR) (Escalante-Semerena et al., 1984; Leigh et al., 1985), Tetrahydromethanopterin (H_4MPT) (Leigh et al., 1985) und Coenzym M (HS-CoM) gebunden sind. Methanofuran ist ein substituiertes C4-Furfurylamin, welches Carboxyl- und Formyl-Gruppen zu binden vermag. H_4MPT ist ein C_1 -Überräger mit einem zentralen Pterin-Ringsystem, das im Stande ist, Formyl-, Methenyl-, Methylen- und Methyl-Gruppen zu tragen. Strukturell ähnelt es dem Tetrahydrofolat, wobei die beiden Träger aber nicht funktionell äquivalent sind (Maden, 2000). Der einfachste methanogene Cofaktor ist Coenzym M (HS-CoM; 2-

Mercaptoethan-Sulfonat), das als zentraler Methylgruppen-Träger in der Methanogenese dient (Taylor und Wolfe, 1974). Die methylierte Form ((2-Methylthio) Ethansulfonat) ist das Substrat der Methyl-CoM-Reduktase, die den terminalen Schritt in der Methanogenese katalysiert. Dieses Enzym enthält einen weiteren ungewöhnlichen und einzigartigen methanogenen Cofaktor, bekannt unter dem Namen Coenzym F₄₃₀ (Ellefson et al., 1982). Aus chemischer Sicht ist dies ein Tetrapyrrol-Ringsystem, welches im Gegensatz zu allen anderen Porphyrinen, ein Nickel-Ion koordiniert (Pfaltz et al., 1982; Livingston, 1984).

Neben den methanogenen Cofaktoren, die C₁-Gruppen übertragen, existieren noch drei andere Cofaktoren, die am Transfer von Elektronen bzw. von Reduktionsäquivalenten beteiligt sind: 1. F₄₂₀ ist ein Deazaflavin mit einem Mittelpunktpotential von -360 mV, das als zentraler Elektronenträger im Cytoplasma der Methanogenen fungiert (Walsh, 1986). Auf den ersten Blick scheint es ein Flavin zu sein; allerdings ist es näher mit den Nikotinamiden verwandt. Bei näherer Betrachtung des aktiven Zentrums (Kohlenstoffatom 4), wird deutlich, dass die chemische Struktur dieses Moleküls identisch mit der von Pyridin-Nukleotiden ist. Das bedeutet, dass Coenzym F₄₂₀, wie NAD⁺ und NADP⁺, im Verlauf der enzymatischen Reaktionen als Hydridionen-Donor oder -Akzeptor fungiert. 2. Neben F₄₂₀ agiert auch Coenzyme B (HS-CoB) als Elektronenüberträger während der Methanogenese. Die Struktur von HS-CoB wurde als N-7-Mercaptohetanoyl-L-Threoninphosphat charakterisiert und durch chemische Synthese bestätigt (Noll und Wolfe, 1987). Kobelt et. al. (1987) demonstrierten, dass nur die L-Form von HS-CoB aktiv ist. 3. Der Elektronentransport in der Atmungskette wird normalerweise durch Chinone vermittelt. In der Cytoplasmamembran der methanogenen Organismen wurde bisher nur Tocopherolchinone in sehr geringen Mengen identifiziert (Hughes und Tove, 1982), welche offensichtlich keine Funktion in der Elektronentransportkette haben. Stattdessen ist der neu entdeckte Cofaktor Methanophenazin im Stande, den Elektronentransport zwischen Membran gebundenen Enzymen in Vertretern der Ordnung *Methanosarcinales* zu vermitteln. Methanophenazin ist das erste Beispiel eines Phenazin-Derivats, das am Elektronentransport in biologischen Systemen beteiligt ist. Ferner wurde bewiesen, dass die Rolle des Methanophenazin im Energiemetabolismus von *Methanosarcina*-Stämmen ähnlich zu der des Ubichinons in Mitochondrien und Bakterien ist.

4.1.2 Methanbildung aus Methanol

Die wichtigsten methanogenen Substrate sind $\text{H}_2 + \text{CO}_2$ und Acetat. In marinen und Brackwasser-Umgebungen dominieren hingegen methylierte Substrate. Die Methylamine und Methylthiole entstammen dem anaeroben Zerfall zellulärer Osmolyte (Cholin-Derivate, Betain, Trimethylamin-N-oxid und Dimethylsulfonpropionat), die in Phytoplankton und bestimmten Pflanzen gefunden werden. Mitglieder der Familie *Methanosarcinaceae* sind im Stande, Methan aus methylierten Substraten wie Methanol, Methylaminen, Methylmercaptopropionaten oder Dimethylsulfiden zu produzieren (Boone et al., 1993; Finster et al., 1990; Van der Maarel et al., 1995). Der Großteil des gegenwärtigen Wissensstandes beruht auf Studien an *Methanosarcina*-Stämmen beim Wachstum auf Methanol. In Abwesenheit von H_2 sind die methylotrophen Methanarchaea im Stande, Methanol nach Gleichung 1 zu verwerten.



Dieser Prozess kann als Disproportionierung von vier Methanol-Molekülen beschrieben werden. Drei Methylgruppen werden zu Methan reduziert (Gleichung 2) und der vierte Methylrest wird zu CO_2 oxidiert (Gleichung 3). Ein Schema dieses Prozesses ist in Abbildung 25 dargestellt.

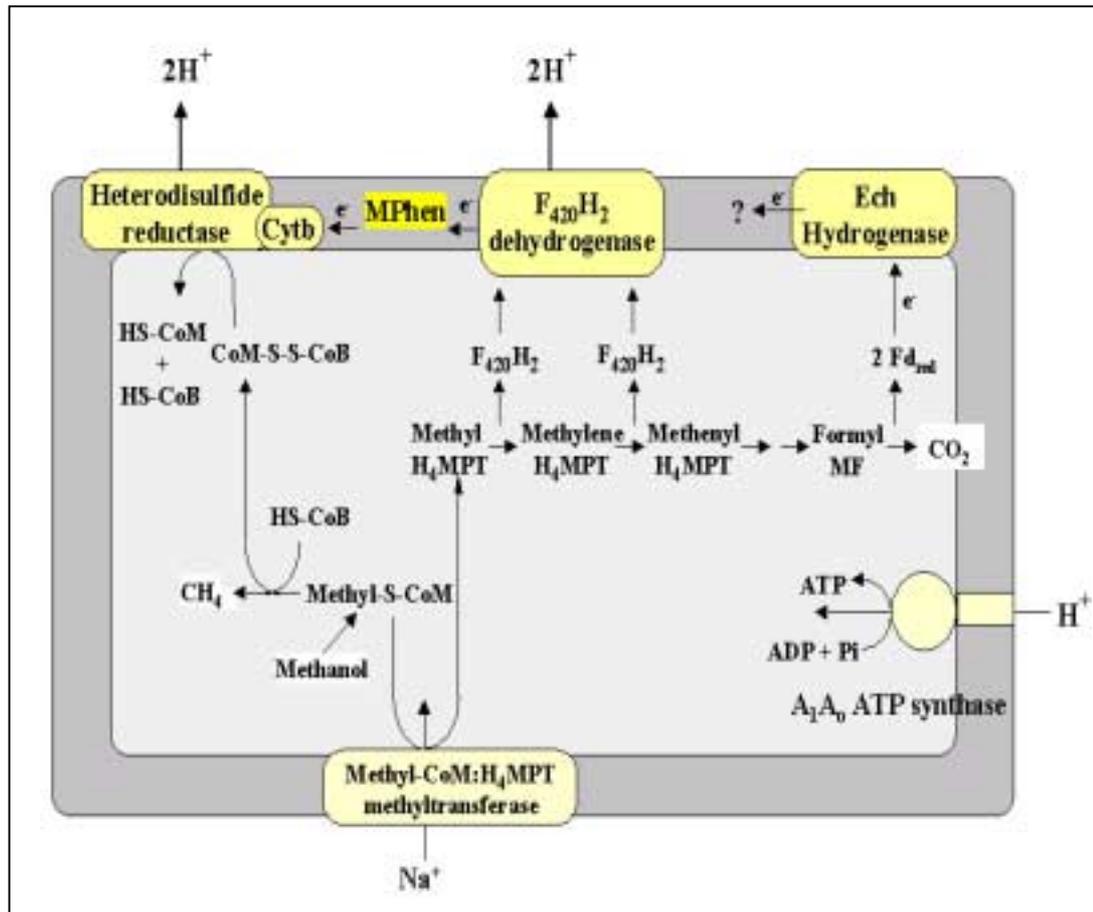


Abb. 25: Darstellung der Methanogenese aus Methanol in *Ms. mazei* Gö1

In dieser Abbildung sind einige der an der Methanbildung aus Methanol beteiligten Enzyme und die entsprechenden Intermediate abgebildet. Die graue Umrandung stellt die Cytoplasmamembran mit den gebundenen Enzymen dar.

Der Methyltransfer von Methanol zu HS-CoM wird durch ein lösliches Methyltransferasesystem bestehend aus den drei Polypeptiden MtaA, MtaB und MtaC katalysiert, wobei das letztere eine Corrinoid-Gruppe enthält (Sauer und Thauer, 1999; Sauer et al., 1997).

An dieser Stelle verzweigt sich der Stoffwechselweg in den reduktiven- und oxidativen Zweig (Abb. 25). Im reduktiven Zweig werden 3 von 4 Methyl-S-CoM Molekülen reduktiv zu Methan demethyliert. Methyl-S-CoM wird durch die Methyl-CoM-Reduktase reduktiv gespalten. Hierbei dient Coenzym B der Reduktase als

Elektronendonator (Kobelt et al., 1987; Bobik et al., 1987; Ellermann et al., 1988; Hartzell et al., 1988; Ermler et al., 1997; Becker und Ragsdale, 1998). Zusammengefasst, führt diese Reaktion zur Bildung von Methan und einem gemischten Disulfid bestehend aus Coenzym M und Coenzym B, welches als Heterodisulfid bezeichnet wird (CoM-S-S-CoB) (Bobik et al., 1987; Ellermann et al., 1988). Die Entstehung des Heterodisulfids ist der letzte Schritt in der Methanbildung (Abb. 25). Daraus ergibt sich die Frage, wie dieses Intermediat reduziert wird, um eine Regeneration von HS-CoM und HS-CoB für die nächsten Reaktionszyklen zu ermöglichen. Als synthetisches Heterodisulfid zur Verfügung stand, wurde herausgefunden, dass CoM-S-S-CoB der terminale Elektronenakzeptor einer verzweigten Atmungskette in diesen Organismen ist (Deppenmeier, 2002b). Das Enzym, das das Heterodisulfid reduziert, ist die Membran gebundene Heterodisulfid-Reduktase, welche als terminale Reduktase fungiert (Abb. 25). Daraus wird ersichtlich, dass der gesamte Prozess der Methanogenese dazu dient, eine organische Disulfid-Verbindung zu produzieren, welche als Elektronenakzeptor in der anaeroben Atmung eingesetzt wird. In dieser Hinsicht ist der methanogene Metabolismus dem Prozess der Schwefelatmung vieler Bakterien und Archaeen ähnlich.

In Abwesenheit von molekularem Wasserstoff sind Vertreter der Gattung *Methanosarcina* dazu gezwungen, alternative Elektronendonoren für die Reduktion von CoM-S-S-CoB zu benutzen. Beim Wachstum auf methylierten C₁-Verbindungen wird reduziertes Coenzym F₄₂₀ (F₄₂₀H₂) gebildet, welches als Elektronendonator dient.

Die F₄₂₀H₂:Heterodisulfid-Oxidoreduktase katalysiert die Übertragung der Elektronen vom F₄₂₀H₂ auf das Heterodisulfid. Die Übertragung stellt eine Elektronentransportkette dar und lässt sich mit folgender Formel zusammenfassen (Deppenmeier et al., 1990):



In einer ersten Reaktion überträgt die F₄₂₀H₂-Dehydrogenase (Abken und Deppenmeier, 1998) die Elektronen von F₄₂₀H₂ auf Methanophenazin, wobei pro

übertragenem Elektronenpaar zwei Protonen durch das Enzym zur Außenseite der Cytoplasmamembran transloziert werden (Bäumer et al., 2000). Das reduzierte Methanophenazin wird dann in einer zweiten Reaktion von der Heterodisulfid-Reduktase reoxidiert und die Elektronen auf Heterodisulfid übertragen, welches in CoM-SH und CoB-SH zerfällt. Coenzym B steht somit wieder als Elektronendonator für den reduktiven Zweig zur Verfügung. Die Reaktion geht ebenfalls mit einer Translokation von zwei Protonen pro übertragenen Elektronenpaar einher (Ide et al., 1999). Somit sind beide Teilreaktionen des $F_{420}H_2$:Heterodisulfid-Oxidoreduktase-Systems an dem Aufbau eines elektrochemischen Gradienten gekoppelt, welcher zur ATP-Synthese genutzt werden kann (Deppenmeier, 2002b; Müller et al., 1999).

Im oxidativen Zweig des Stoffwechselweges, wird eine von vier Methylgruppen schrittweise zu Kohlendioxid oxidiert. Eine Membran gebundene Methyltransferase katalysiert die Übertragung der Methylreste von HS-CoM auf H_4MPT . Dies ist eine endergone Reaktion ($\Delta G_0' = 30 \text{ kJ/mol}$) und wird von einem elektrochemischen Natriumionen-Gradienten angetrieben (Müller et al., 1988; Müller et al., 1992). Das entstandene N^5 -Methyl- H_4MPT wird anschließend schrittweise zu N^5 -Formyl- H_4MPT oxidiert. Intermediate dieses voll reversiblen Prozesses sind N^5 : N^{10} -Methylen- H_4MPT und N^5 : N^{10} -Methenyl- H_4MPT . Die aus den Oxidationen freigesetzten Reduktionsäquivalente werden zur Reduktion von F_{420} verwendet. Diese Reaktionen werden von der Methylen- H_4MPT -Reduktase (Te Brömmelstroet et al., 1991; Ma und Thauer, 1990; Te Brömmelstroet et al., 1990, Shima et al., 2000), der Methylen- H_4MPT -Dehydrogenase (Ensle et al., 1991; Te Brömmelstroet et al., 1991) und der Methylen- H_4MPT -Cyclohydrolase katalysiert (Vaupel et al., 1996; Grabarse et al., 1999), die alle im Zytoplasma der methanogenen Organismen lokalisiert sind. Die Formylgruppe wird danach auf Methanofuran (MFR) durch die katalytische Aktivität der Formyl-MFR: H_4MPT -Formyltransferase übertragen. Am Ende des Reaktionsweges katalysiert die Formyl-MFR-Dehydrogenase die Oxidation von CHO-MFR, was zur Freisetzung von CO_2 und zur Regeneration von MFR führt. Der Elektronenakzeptor dieses Proteins ist noch nicht eindeutig charakterisiert. Der Elektronentransfer ist ein exergoner Prozess und ist wahrscheinlich an der Bildung eines transmembranen Ionengradienten gekoppelt. An diesem Prozess ist die Ech-Hydrogenase beteiligt (siehe 4.5.3).

4.1.3 Methanbildung aus Acetat

Der Großteil der Methanproduktion in der Natur stammt aus der Mineralisation von Acetat (aceticlastischer Weg). Dieser Stoffwechselweg wird nur von Mitgliedern der Gattungen *Methanosarcina* und *Methanosaeta* (Ferry, 1997) durchgeführt. Der Stoffwechselweg in *Ms. mazei* Gö1 ist in Abbildung 26 dargestellt.

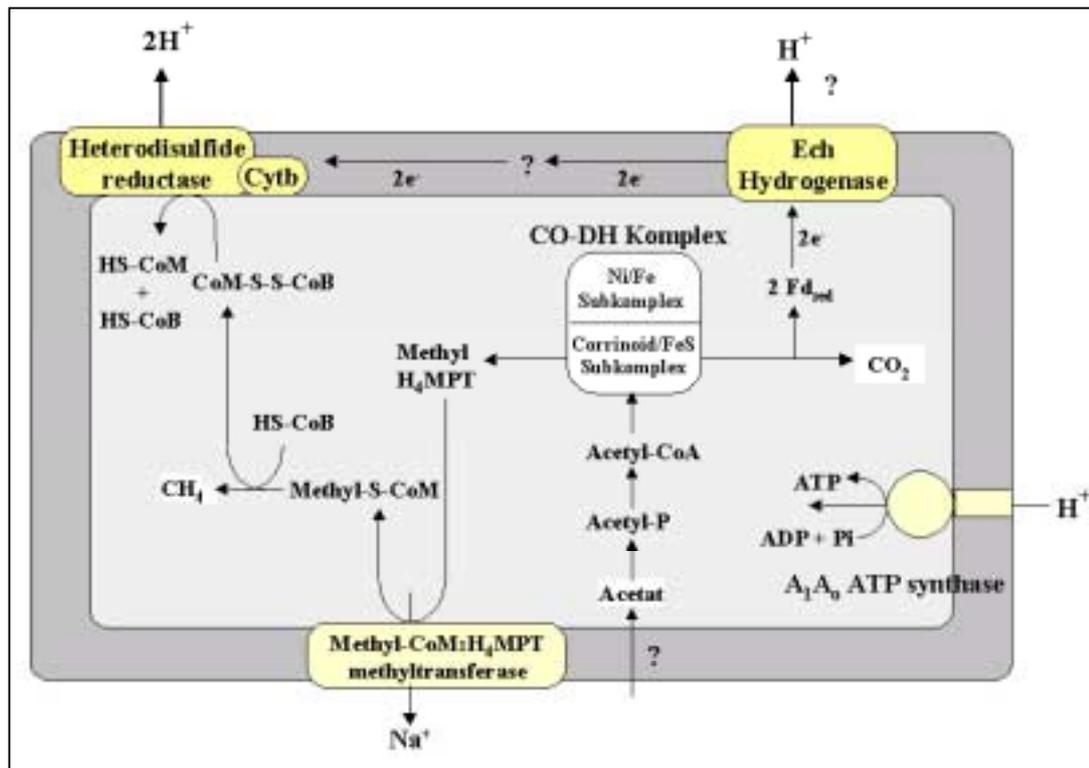


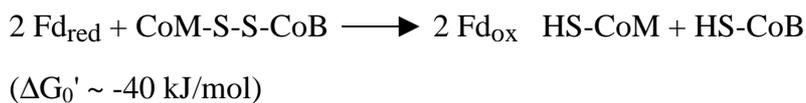
Abb. 26: Darstellung der Methanogenese aus Acetat in *Ms. mazei* Gö1

In dieser Abbildung sind die meisten der an der Methanbildung aus Methanol beteiligten Enzyme und die entsprechenden Intermediate abgebildet. Die graue Umrandung stellt die Cytoplasmamembran mit den gebundenen Enzymen dar.

Nach dem Eintritt in die Zelle, wird Acetat entweder durch eine Acetat-Kinase und eine Phosphotransacetylase (*Methanosarcina*) oder durch eine Acetyl-CoA-Synthase (*Methanosaeta*) aktiviert. Das resultierende Produkt Acetyl-CoA wird durch einen CO-Dehydrogenase/Acetyl-CoA-Synthase Komplex gespalten. Im Verlauf dieser Reaktion wird enzymgebundenes CO zu CO₂ oxidiert und die Elektronen werden zur Reduktion von Ferredoxin (FR) verwendet. Der Methylrest wird auf H₄MPT übertragen. Das entstandene Methyl-H₄MPT wird durch die katalytische Aktivität der Natrium-Ionen-translozierenden Methyl-H₄MPT:HS-CoM-Methyltransferase und

der Methyl-S-CoM-Reduktase (s.o.) zu Methan umgewandelt. Im letzten Schritt des acetoclastischen Weges reduziert die Heterodisulfid-Reduktase das CoM-S-S-CoB. Die Elektronen stammen von reduziertem Ferredoxin. Die Elektronenübertragung erfolgt durch eine dritte Membran gebundene Elektronentransportkette (Ferredoxin:Heterodisulfid-Oxidoreduktase; Deppenmeier et al, 1996; Terlesky und Ferry, 1988).

Die Methanogenese aus Acetat ist an die kleinste Änderung der freien Energie ($\Delta G^{\circ} = -36 \text{ kJ/mol CH}_4$) aller methanogenen Substrate gekoppelt. Da in *Methanosarcina*-Stämmen ein ATP zur Acetat-Aktivierung benötigt wird, ist es notwendig, dass die Methanogenese aus Acetyl-CoA eine höhere Ausbeute als 1 ATP pro gebildetes CH_4 ergibt. Deshalb benötigt der Organismus ein effizientes Energiekonservierungssystem. Der Methylgruppentransfer von $\text{CH}_3\text{-H}_4\text{MPT}$ zu HS-CoM führt zum Aufbau eines elektrochemischen Natriumionen-Gradienten über die Cytoplasmamembran. Die verbleibenden Intermediate CoM-S-S-CoB und Fd_{red} werden in dem Ferredoxin:Heterodisulfid-Oxidoreduktase System weiter verwertet (s.o.). Auch diese Reaktion ist wahrscheinlich mit dem Aufbau eines elektrochemischen Protonengradienten gekoppelt.



Die exakte Zusammensetzung und Reaktionsmechanismen des System sind noch nicht im Detail bekannt. 1) Hedderich et al. (1999) vermuten, dass die Ech-Hydrogenase und die Heterodisulfid-Oxidoreduktase an diesem Prozess beteiligt sind. In diesem Fall würde die Oxidation von reduziertem Ferredoxin durch die Ech-Hydrogenase katalysiert werden. Die Reaktion resultiert in der Freisetzung von molekularem Wasserstoff. H_2 würde dann durch die F_{420} -nichtreduzierende-Hydrogenase reoxidiert und mit Hilfe von Methanophenazin zur Heterodisulfid-Reduktase geleitet werden (Deppenmeier et al., 1991). 2) Eine andere Hypothese postuliert, dass H_2 kein eigentliches Intermediat ist. Statt dessen werden die Elektronen aus dem reduzierten Ferredoxin direkt von der Ech-Hydrogenase zum

Methanophenazin übertragen. Der reduzierte Cofaktor könnte dann durch die Heterodisulfid-Reduktase reoxidiert werden (Deppenmeier, 2002a).

4.2 Anmerkungen zur Genom-weiten Transkriptionsanalyse von *Ms. mazei* durch die DNA-Microarray-Chip-Technologie

Im Zeitalter der vollautomatisierten Sequenzierung gesamter Genome (wie z.B. des Stammes Gö1) wird die Anzahl an Sequenzen in Datenbanken wie GenBank immer größer (Benson et al., 1998). Dadurch wächst jedoch auch die Kluft zwischen der Anzahl bekannter Sequenzen und der Kenntnis über ihre Funktion. Hieraus ist ein neuer Zweig der modernen Molekularbiologie entstanden, der oft als funktionelle Genomanalyse (functional genomics) bezeichnet wird. Diese Wissenschaft hat es sich zum Ziel gesetzt, diese Kluft zu überbrücken.

Die Frage, ob es sich bei einer DNA-Sequenz um ein Gen handelt und ob dieses Gen auch wirklich exprimiert wird, ist die Grundlage für die spätere biochemische Charakterisierung des Genproduktes, die Aufklärung der Struktur und der molekularen Mechanismen. Hierfür waren bisher jedoch immer zeitaufwendige Methoden wie die Erzeugung von Knock-out-Mutanten oder rekombinante Genexpressionen notwendig.

Um mit dem Wachstum der Sequenzdaten Schritt halten zu können, bedient sich die funktionelle Genomanalyse eines anderen Ansatzes, dem der differentiellen Expression. Unterschiede in der Genexpression eines Organismus sind eine Reaktion der Zellen auf innere und äußere Stimuli. Aus der Aktivierung bzw. Inaktivierung eines Gens unter gegebenen inneren oder äußeren Einflüssen, lässt sich auf die Funktion dieses Gens rückschließen.

Im Laufe der Zeit sind für die Analyse der differentiellen Genexpression eine Reihe von molekulargenetischen Methoden beschrieben worden, die trotz unterschiedlicher Ansätze die gleiche Zielstellung verfolgen: Es sollen die Transkripte von Genen (mRNAs) erfasst und möglichst quantitativ analysiert werden, die nach Veränderung

gegebenen Bedingungen (Wachstum, Stress, zelluläre Signale, etc.) in unterschiedlicher Häufigkeit in der Zelle vorliegen.

Eine vollständige Erfassung wird allerdings durch die Verteilung limitiert mit der mRNAs in einer Zelle vorliegen. So entspricht etwa die Hälfte der gesamten mRNA-Menge einer typischen Zelle wenigen, aber extrem häufig vorhandenen Transkripten (500 bis 12000 Kopien/Zelle), die lediglich 1 % der Gesamtmenge aller exprimierten Gene ausmachen. Im Gegensatz dazu entfällt die andere Hälfte der mRNA-Masse einer Zelle auf die restlichen 99 % der verschiedenen Transkripte einer Zelle. Sie stellen damit seltene mRNAs dar (15-500 Kopien/Zelle; Axel et al., 1976).

Zu den Methoden, die eine Analyse der Genexpression von bereits bekannten Sequenzen erlauben, zählen u.a. RT-PCR, Northern-Blots sowie „Ribonuclease protection assays“. Die Charakterisierung der Genexpression einzelner Gene anhand dieser Methoden funktioniert zuverlässig. Jedoch ist eine gleichzeitige, parallele Analyse einer großen Zahl von Genexpressionen mit diesen Methoden nicht möglich (Alwine et al., 1987). Um die Expression einer Vielzahl von Genen gleichzeitig zu untersuchen, wurden zunächst Methoden entwickelt, die auf DNA/DNA-Hybridisierungen beruhen. Hierbei wird durch reverse Transkription der mRNAs eine cDNA-Bibliothek für die Hybridisierung generiert. Beispiele für solche Methoden sind die differenzielle Hybridisierung (St. John et al., 1979) und die daraus weiterentwickelte subtraktive Hybridisierung (Sargent et al., 1983). Beide Methoden ermöglichen grundsätzlich auch die Identifizierung unbekannter, differenziell exprimierter Gene. Allerdings ist die Identifizierung von seltenen Transkripten sowie eine Quantifizierung der Genexpression hiermit nicht möglich.

Eine relativ neue Methode beruht auf der Identifizierung von Genexpressionen durch Hybridisierung von cDNA mit kurzen DNA-Fragmenten, die systematisch auf einer Microarray-Slide-Oberfläche immobilisiert sind (Schena et al., 1995). Die Sequenz der immobilisierten DNA-Fragmente sowie ihre Position auf dem Slide sind dabei bekannt und dienen zur Identifizierung der Transkripte. Die zu analysierende cDNA wird dazu mit einem Fluoreszenzfarbstoff markiert und ermöglicht so die Detektion der Hybridisierung auf dem Microarray-Chip. Aus der Positionen des Signals lässt sich das Gen identifizieren und anhand der Intensität des Signals auch die

Expressionsstärke quantifizieren. Die große Zahl der DNA-Fragmente auf dem Chip lässt dabei eine parallele Analyse einer Vielzahl von Transkripten zu.

Gegenwärtig besitzt diese Methode aber noch gravierende Nachteile. So ist die Analyse auf bekannte Transkripte beschränkt und die Quantifizierung von selten transkribierten Genen wird durch das Vorhandensein der großen Menge an konstitutiven Transkripten sehr erschwert. Durch die Fortschritte in der DNA-Microarray-Technologie und durch Genom-Projekte, besitzt diese Methode, im Gegensatz zu den anderen Techniken, aber ein sehr hohes Entwicklungspotential in der effizienten Analyse der differentiellen Genexpression (Schena et al., 1998). Diese Zusammenhänge sind der Grund, weshalb die Methode in dieser Arbeit gewählt wurde, um Genom-umfassende Transkriptionsanalysen von *Ms. mazei* Göl durchzuführen.

4.3 Anmerkungen zu bioinformatischen Ansätzen zur Genom- weiten Transkriptionsanalyse von *Ms. mazei*

Die bei einer genomumfassenden Transkriptionsanalyse benötigten und produzierten Datenmengen sind in einer Größenordnung, die von einem Menschen nicht mehr überschaubar ist. Dies führt dazu, dass die moderne Wissenschaft der funktionellen Genomanalyse sehr eng mit der Bioinformatik verknüpft ist, um diese Datenmengen bearbeiten und auswerten zu können. Im Zeitalter des World-Wide-Web und schneller Internet-Verbindungen, wächst die Datenmenge in öffentlichen Datenbanken wie GenBank exponentiell an und viele freie Programme im Internet können Lösungen und Hilfen für anfallende Probleme schneller bieten, als man sie selbst lösen könnte. Durch die Zugänglichkeit dieser effizienten und kostengünstigen Informationsquellen wird die Notwendigkeit von Computer gestützten biologischen Forschungsansätzen immer bedeutender.

4.3.1 Anmerkungen zur Programmiersprache PERL

Für das Projekt, das in dieser Arbeit dargestellt wird, übernahm die Bioinformatik eine Schlüsselrolle in der Etablierung der genomweiten Transkriptionsanalyse. Computerprogramme wurden erstellt, die eine manuelle Arbeit von Wochen in wenigen Sekunden vollbrachten und Datenmengen durch Computernetzwerke transferierten, die in ausgedruckter Form eine kleine Bibliothek füllen könnten. Als Sprache zum Erstellen der Computerprogramme wurde PERL gewählt.

An dieser Stelle soll ein kleiner Überblick über diese Sprache gegeben werden, da es sich bei PERL um die wichtigste Sprache in der Bioinformatik handelt und jeder, der sich mit Bioinformatik beschäftigt, in Kontakt mit dieser Sprache kommen wird.

Ursprünglich konzipiert, um große Mengen Text schnell zu durchsuchen und zu analysieren, wird PERL mittlerweile für nahezu alle denkbaren Zwecke eingesetzt. Daher kommt auch der Name "Practical Extraction and Report Language". Die Sprache wurde von einem einzelnen Entwickler entworfen, und zwar 1987 von Larry Wall.

Ein Grund, warum diese Sprache so stark in der Bioinformatik verbreitet ist, könnte sein, dass sich PERL vieler Elemente anderer Programmiersprachen bedient, z. B. C und der Tools sed und awk. Somit kann jemand, der diese Tools anwenden kann, ohne allzu viel Aufwand PERL-Programme schreiben. Des Weiteren kommen viele alltägliche englische Wörter zum Einsatz, die dann auch noch oft genau das tun, was man erwarten würde. So steht dem C- Befehl "include" das für den "Amateur" vielleicht klarere "use" gegenüber (Hansen, 2001; Lesk, 2002).

Bei PERL handelt es sich um eine interpretierte Sprache, d. h. PERL-Programme werden als Bytecode kompiliert und zur Laufzeit interpretiert. Als Folge seiner Herkunft ist PERL besonders optimiert für das Arbeiten mit Texten (z.B. Durchsuchen, Ersetzen). Es bietet jedoch auch wichtige Features für viele andere Gebiete, darunter Systemverwaltung, Prozesssteuerung, Internetskriptprogrammierung oder Objekt-orientiertes Programmieren. PERL ist Freeware und auf vielen Rechnerplattformen verfügbar wie z. B. UNIX/Linux, DOS/Windows und Mac OS. Die Stärken von PERL sind unter anderem:

- Erschließung, Aufbereitung und Neuformatierung beliebiger Texte,
- kann C-Programme sowie sed-, awk- und shell-Skripte ersetzen und
- überwindet alle Beschränkungen dieser Sprachen.

Am bekanntesten ist die Verwendung von PERL als Programmiersprache für CGI-Skripte im WWW sowie im Bereich der UNIX-Systemadministration. Aber auch für viele andere Aufgaben hat sich PERL bewährt:

- Datenbankmanipulation
 - Text- und Dokumentenverarbeitung
 - Datenextraktion und -reduktion
 - Netzwerkadministration,
 - Filesystemmanipulation
 - sicherheitsorientiertes Programmieren
 - WWW-basiertes Informationsmanagement
 - objekt-orientierte und funktionale Programmierung
- etc.

Die von PERL benutzten Typen und Strukturen sind einfach zu verstehen und zu benutzen. Es gibt keine künstlichen Beschränkungen für die zu verarbeitenden Daten. Strings und Arrays können beliebig wachsen (solange der Speicher reicht) und sie sind darauf ausgelegt, trotzdem gut zu skalieren. Obwohl PERL in vieler Hinsicht eine einfache Sprache ist, ist es auch eine sehr reiche Sprache, bei der es sehr viel zu lernen gibt. Dies ist der Preis dafür, dass die simplen Probleme einfach und die komplexen Aufgaben überhaupt zu lösen sind. Trotzdem ist PERL keine Sprache für große Projekte (Gundavaram, 1996).

PERL ist im besten und positivsten Sinn eine "Hacker-Sprache". Larry Wall hat dazu noch bemerkt, dass man in PERL die Lösung eines Problems auf vielen Wegen erreichen kann, und dass alle Wege richtig sind. PERL ist somit auch eine Sprache, die der Bewusstseinsverengung der Programmierer entgegenwirkt (Wall et al., 1992). PERL wurde von Beginn an als Mischung zwischen Shell- und C-Programmierung entwickelt. Die von PERL benutzten Typen und Strukturen sind einfach zu verstehen

und zu benutzen. Obwohl eigentlich eine Interpretersprache, ist PERL eher eine Kreuzung aus Interpreter und Compiler und ähnelt damit Java. Zur Laufzeit wird das PERL-Skript vom Interpreter relativ schnell in ein internes, bytecode-ähnliches Format umgesetzt. Dabei werden, wie bei jedem Compiler, Syntaxchecks und Optimierungen durchgeführt (z. B. syntaktische und zum Compilierungszeitpunkt erkennbare semantische Fehler, Probleme beim Einbinden von Bibliotheken, usw.). Nach dem erfolgreichen Durchlauf des Compilers wird der Zwischencode an den PERL-Interpreter zur Ausführung weitergereicht. Compiler und Interpreter sind effizient und der typische Übersetzungszyklus dauert meist nur Sekunden. Zusammen mit den vielen Möglichkeiten PERLs, auf Fehler "weich" zu reagieren erlaubt diese kurze Zykluszeit „Rapid-Prototyping“ in PERL. (Wall et al., 1992)

4.3.2 Anmerkungen zur Computer-gestützten Analyse der genomweiten Transkriptionsdaten (=Data Mining).

In der modernen Wissenschaft der ‚functional genomics‘ werden - wie im Rahmen dieser Arbeit (siehe 3.4.1, Anhang A) - große Mengen an Daten erzeugt, die nur durch computergestützte Analysen gehandhabt werden können. Aus dieser Aufgabenstellung entstand die Methode des Data Mining, auf das an dieser Stelle näher eingegangen werden soll, um die Komplexität der Auswertung der Ergebnisse dieser Arbeit zu verdeutlichen.

Data Mining ist ein relativ junges in äußerst dynamischer Entwicklung stehendes Forschungsgebiet, welches sich im Schnittpunkt von verschiedenen Wissenschaftsdisziplinen befindet. Data Mining umfasst ebenso Inhalte der Statistik wie Aspekte der Datenbanktheorie, des „Pattern Recognition“ und der „Artificial Intelligence“ - Forschung.

In Bezug auf die Ziele und Aufgaben des Data Mining existiert eine große Übereinstimmung mit denen der klassischen und angewandten Statistik, insbesondere in dem Bereich der explorativen Datenanalyse. Für beide Disziplinen ist es das primäre Ziel, relevante Strukturen und Muster in multivariaten Datenkörpern zu finden. Formulierungen wie „learning from data“ oder „turning data

into information“ lassen sich gleichermaßen für beide Disziplinen anwenden. Die klassischen Methoden der angewandten Statistik gehen von kleinen, überschaubaren Datensätzen aus, die häufig als unabhängige Realisierungen identisch verteilter Zufallsvariablen interpretiert werden können und oft spezifisch für die Beantwortung einer bestimmten Fragestellung gesammelt wurden. Datensätze mit einigen tausend Beobachtungen gelten hier schon als groß.

Im Bereich des Data Mining werden große Datenkörpern analysiert, die Millionen von Datensätzen, welche keineswegs primär zum Zwecke der statistischen Analyse gesammelt wurden, umfassen. Die Entstehung des Data Mining ist eng mit der Computerisierung unserer Gesellschaft verknüpft. Die moderne Computertechnologie und Datenbanktechnik ermöglicht es, kostengünstig riesige Datenmengen elektronisch zu speichern. Hieraus entsteht der Wunsch aus dieser reichen Vielfalt verfügbarerer Datensammlungen in Forschungsinstituten (z.B. GenBank) nützliche Erkenntnisse zu gewinnen. Der bekannte Trendforscher John Naisbett stellte einmal provokativ die These auf *“We are drowning in information, but starving for knowledge“*. Die Vertreter des Data Mining erkannten sehr bald, dass der klassische Weg der Statistik nicht ausreichend ist, für diese neue Herausforderung (Fayyad et al., 1996).

Vereinfacht lässt sich das Verfahren des Data Mining als eine Methodensammlung ansehen, die es ermöglicht, die komplexen Prozesses des Findens valider, neuer, interessanter und auch nützlicher Muster und Regeln aus großen Datenbanken durchzuführen. Daher wird Data Mining oft auch als *“Knowledge discovery in databases“* bezeichnet.

Abgesehen von der häufigen Größe der zu analysierenden Datenbanken, ist ein wesentlicher Unterschied im Entstehungsprozess der Daten auszumachen. Die zu untersuchenden Daten, wurden keineswegs primär zum Zweck statistischer Auswertungen erzeugt. Vielfach sind diese Daten als Ergebnisse von experimentellen Analysen zu sehen. Die pragmatische Nutzung vorhandener Daten steht hierbei im Vordergrund, während theoretische Überlegungen über ideale Stichproben oder optimale Versuchspläne meist in den Hintergrund gedrängt werden. Es ist offensichtlich, dass hieraus Probleme in Bezug auf die Qualität der Daten und

die Interpretation der Ergebnisse entstehen. Die Größe der untersuchten Datenkörper sowohl in Bezug auf die Anzahl der Experimente als auch in Bezug auf die Anzahl der Variablen stellen die klassischen Vorgehensmodelle der induktiven Statistik infrage. In Anbetracht der großen Zahl an Beobachtungen erscheint z.B. das klassische Vorgehensmodell im Sinne von Neyman und Pearson, statistische Signifikanztests mit vorgegebenem Fehler 1. Art (= Alpha-Fehler) zu rechnen, fragwürdig. Aufgrund der großen Fallzahl verfügen die Tests häufig über eine solche Macht, dass selbst minimale, inhaltlich nicht relevante Effekte als statistisch signifikant eingestuft werden.

Die Methoden des Data Mining basieren oft auf problembezogenen „Ad-hoc“-Überlegungen. Im Gegensatz dazu basieren klassische statistische Verfahren auf stochastischen Modellen. Dieser Aspekt hat dazu geführt, dass bei Statistikern heuristische Methoden, für welche keine allgemeinen theoretische Ergebnisse vorliegen, kaum verbreitet sind.

Bei den Methoden der klassischen Statistik steht nach wie vor das Modell im Vordergrund. In der Vergangenheit bediente man sich dabei relativ starrer nur gering parametrisierter Modelle, die eine einfache Berechenbarkeit und theoretische Behandlung auch komplexer multivariater Fragestellungen gewährleisten, sich jedoch oft als nicht robust gegenüber Verletzungen der Modellannahmen erwiesen und daher für die Praxis nicht immer angemessen waren. Die Entwicklung der multivariaten Statistik ist von einem Trend zu immer flexibleren komplexeren bzw. auch robusteren Modellen charakterisiert, welche jedoch selbst mit modernen Hochleistungscomputern insbesondere bei größeren Datensätzen beträchtliche computationale Probleme aufwerfen.

Als Beispiel sei hier die Entwicklung unterschiedlicher Verfahren des „model-based clustering“ (Banfield und Raftery, 1993) erwähnt, bei der Clusterstrukturen unter verschiedenen Annahmen über die Mischung von multivariat normalverteilten Gruppen mittels Maximierung unterschiedlicher Kriterien gesucht werden. Allerdings erweist sich hierbei die algorithmische Suche nach einer optimalen Partition bei großen Datenmengen nicht als trivial (Coleman et al., 1999; Hudec und Steiner, 2002). Ein anderes Beispiel bei dem der Algorithmus die Eigenschaften des

Schätzers quasi dominiert, ist die robuste Schätzung von Kovarianzmatrizen (Woodruff und Rocke, 1994).

Data Mining ist im Wesentlichen ein suchendes Vorgehensmodell, bei dem Muster, Strukturen und Hypothesen auf semi-automatische Weise direkt aus eventuell nicht repräsentativen Daten abgeleitet werden. Deshalb handelt es sich um eine Hypothesen generierende Vorgehensweise, bei der die Ergebnisse in Bezug auf die Anwendbarkeit auf andere Populationen bzw. Generalisierung nur mit größter Vorsicht interpretiert werden dürfen. Bei Data Mining geht es in der Regel nicht darum, „wahre Gesetzmäßigkeiten“ über den Daten generierenden Prozess aufzuzeigen. Im Vordergrund steht, ob die Ergebnisse für den intendierten Zweck brauchbar bzw. praxistauglich sind.

Ein weiteres Problem kann durch das Überschätzen der Allmacht des Algorithmus liegen. Der Anwender begibt sich in die Gefahr, dass er jegliche Beziehung zu den Daten und ihrer Semantik verliert. Der komplexe Algorithmus wird dann zu einer undurchschaubaren Black-Box, die ihn von der Realität, welche von den Daten repräsentiert werden, trennt. Das Überprüfen von experimentellen Fragestellungen entfällt, und die vom Algorithmus generierten Ergebnisse, welche oft nur eine von vielen möglichen Interpretationen eines komplexen Datensatzes darstellen, werden fälschlicherweise als erwiesenes Ergebnis gedeutet.

4.3.3 Anmerkungen zur Gruppierung von Daten durch Cluster-Algorithmen im Prozess des Data Minings

Im Rahmen der Auswertung der DNA-Micorarray Daten wurden die einzelnen Expressionsdaten durch die spezielle Auswertungssoftware Expressionist Analyst einer computergestützten Analyse unterzogen, um die einzelnen Genexpressionen zu clustern (siehe 3.4.1). An dieser Stelle soll ein kurzer Überblick über die Arten der Algorithmen und deren Funktionsweise gegeben werden.

Das Ziel des ‚Clusterings‘ war es, aus einer Menge von Daten zu extrahieren, um Kerninformationen über deren interne Struktur zu erhalten. Dabei werden

Kerninformationen definiert als die reduzierte Information großer Datenmengen auf einen wesentlich kleineren repräsentativen Teil der Daten. Ein Teil der Daten ist repräsentativ, wenn jedes darin enthaltene Objekt, stellvertretend eine Teilmenge (=Cluster) der ursprünglichen Daten ist, und das Verhalten der durch dieses Objekt repräsentierten Daten nach bestimmten ausgewählten Eigenschaften widerspiegelt.

Bei einer Aufteilung der Daten in Cluster, ist es das Ziel Cluster zu bilden, bei denen die Objekte in den Clustern maximal homogen und die Objekte zwischen den Clustern maximal heterogen sind.

Alle folgenden Clusteralgorithmen versuchen dies durch verschiedenste Techniken zu erreichen. Dies ist jedoch in den seltensten Fällen möglich. Meistens kann nur ein lokales Optimum erreicht werden, was aber in den meisten Fällen auch hilfreich ist. Hierbei werden die Daten als Punkte eines n -dimensionalen Raumes definiert. In den folgenden Clusteralgorithmen wird zum Vergleich dieser vektoriellen Punkte ihre Distanz zueinander verwendet. Die folgenden Algorithmen leisten alle das Gleiche, d. h. dienen zum Vergleich von vektoriellen Punkten, aber je nach Algorithmus können diese verschieden eingesetzt werden. Z. B. gibt es Algorithmen die Distanzschwellen als Eingabe erhalten. Bei solchen Algorithmen ist es wichtig, welches Distanzmaß man wählt, da verschiedene Maße verschieden hohe Werte für ein und dasselbe Paar von Vergleichsobjekten liefern. Für die einzelnen gängigen Methoden der Distanzberechnung wird auf Kapitel 3.4.1 dieser Arbeit verwiesen.

4.3.3.1 Hierarchische Algorithmen

Für das Clustern der Expressionsdaten, welche im Rahmen dieser Arbeit gewonnen wurden, diente ein hierarchischer Algorithmus. Auf die genaueren Hintergründe hierfür wurde in Kapitel 3.4.1 eingegangen.

Hierarchische Algorithmen zeichnen sich durch ihren Aufbau einer Baumstruktur aus. In den Blättern des Baumes stehen die zu clusternden Objekte und bilden den jeweiligen Cluster. Auf der Tiefe i des Baumes steht eine Zusammenfassung der

Cluster aus der Tiefe $i+1$ zu neuen Clustern in den Knoten. Die Anzahl der Cluster sinkt mit der Tiefe bis in der Wurzel alle Objekte zu einem Cluster vereinigt sind.

Bei diesen Algorithmen ergibt sich ein typisches Dendrogramm (siehe Kapitel 3, Abb. 22). Die hierarchischen Clusteralgorithmen lassen sich in die zwei Gruppen der agglomerativen und der divisiven Algorithmen einteilen (Everitt, 1998).

Agglomerative Algorithmen:

Die agglomerativen Algorithmen sind im Bereich des hierarchischen Clustering die „*bottom-up*“-Algorithmen. Der Ausgangspunkt eines agglomerativen Algorithmus ist eine Menge, in der alle Objekte in einem eigenen Cluster sind, und eine Matrix, welche die Distanz zwischen den Objekten angibt. Innerhalb des Algorithmus wird eine Distanzschwelle benutzt, um aus der Matrix eine zweite Matrix zu erzeugen, die dazu verwendet wird, um mit verschiedenen graphentheoretischen Ansätzen Komponenten zu erzeugen, die dann Cluster bilden.

Divisive Algorithmen:

Die divisiven Algorithmen sind im Gegensatz zu den agglomerativen Algorithmen „*top-down*“-Algorithmen. Wie bei agglomerativen Algorithmen werden die zu clusternden Objekte mit einer Distanzmatrix eingelesen und ein Dendrogramm wird ausgegeben. Der Unterschied liegt in der Arbeitsrichtung der Algorithmen. Ein divisiver Algorithmus beginnt mit allen Objekten in einem Cluster und versucht die Cluster immer am günstigsten zu teilen, so dass immer möglichst optimale Cluster entstehen. Am Ende liegt dann jedes Objekt in einem eigenen Cluster.

4.3.3.2 Partitionelle Algorithmen

Partitionelle Algorithmen geben nur eine Partition einer Menge von Objekten wieder, d. h. nur eine Aufteilung der zu clusternden Objekte, und sind in Aufbau der Cluster nicht hierarchisch, d. h. im Algorithmus selbst wird keine hierarchische Baumstruktur aufgebaut. Dabei kann die Clusterzahl bei partitionellen Algorithmen Teil der Eingabe oder Teil der Ausgabe sein. Die weiteren Einzelheiten eines partitionellen Algorithmus hängen dann von dem einzelnen Verfahren ab, auf dessen Idee der Algorithmus dann basiert. Folgende partizionelle Algorithmen waren in der

Analyst Expressionist Software (Fa. GeneData, Wien), welche im Rahmen dieser Arbeit verwendet wurde, integriert.

K-mean Algorithmus:

Der K-mean Algorithmus ist ein iterativer Algorithmus, der die zu clusternden Objekte beim ersten Schritt zufällig auf n Cluster verteilt und das Mittel eines jeden Clusters berechnet. Dann wird jedes Objekt wieder aus seinem Cluster entnommen und dem Cluster zugeteilt, zu dem es den geringsten Abstand hat. Die Mittel der Cluster werden wieder neu berechnet und die Objekte dann wieder entnommen. Dies wird solange iteriert, bis sich die Clusterzentren (Mittel der Cluster) stabilisieren, d.h. wenn sich die Differenz zwischen den alten Clusterzentren und den neuen Clusterzentren unterhalb einer definierten Schwelle befinden (Hartigan, 1975). Der K-mean-Algorithmus hat die Eigenschaft in den meisten Fällen sphärische Cluster zu finden. Cluster die beliebige Formen haben, werden dabei oft, je nachdem was man als gewünschte Clusterzahl eingegeben hat, in mehrere sphärische Cluster getrennt. Außenseiterobjekte, die weder zu einem Cluster gehören, noch einen bilden, werden beim K-mean nicht als solche erkannt. Sie mischen sich immer unter die gefundenen Cluster und können so die Varianz in einem bis auf Außenseitern eher homogenen Cluster sehr hoch werden lassen.

SOM Algorithmus:

Der SOM Algorithmus (SOM = Self-Organized-Maps) ist dem K-mean Algorithmus in der Idee und den Ergebnissen sehr ähnlich. Hierbei werden im ersten Schritt jedoch nicht n zufällige Cluster gebildet, sondern es wird am Anfang ein Raster aus vordefinierten Clustern zufällig in die Datenwolke projiziert. Anschließend werden iterativ die Abstände der Objekte zu den Clusterzentren berechnet, und die Zentren werden verschoben, so dass die Größe aller errechneten Abstände minimiert wird. Diese Iteration erfolgt bis keine signifikante Änderung mehr stattfindet (Tamayo et al., 1999; Törönen et al., 1999).

4.4 Untersuchungen der unterschiedlichen Expressionsmuster beim Wachstum von Zellen auf Methanol oder Acetat

Die Gene, deren Transkriptmengen beim Wachstum auf Methanol bzw. Acetat signifikant unterschiedlich waren, lassen sich in verschiedene Gruppen einteilen:

1) Eine erhöhte Expression um mindestens den Faktor drei im Vergleich zu Methanol-gewachsenen Zellen zeigten 118 Gene, wenn die Zellen auf Acetat angezogen wurden. Hierunter fielen Gene, die für folgende Proteine kodieren:

- a) spezifische Enzyme des aceticlastischen Stoffwechselweges der Methanogenese,
- b) Proteine für die aromatische Aminosäure-Biosynthese,
- c) Methylthiol:Coenzyme M-Methyltransferasen,
- d) Aldehyd:Ferredoxin-Oxidoreduktase,
- e) Aufnahmesysteme (Kobalt- und Dipeptid/Oligopeptid-Transporter),
- f) Transkriptionsregulatoren,
- g) hypothetische und konservierte Proteine.

2) Zu den Genen, die beim Wachstum auf Methanol um mindestens den Faktor drei im Vergleich zu Acetat-Zellen hochgefahren wurden (insgesamt 53 Gene), zählten solche, die für die folgenden Proteine kodieren:

- a) Enzyme des oxidativen Zweigs der Methanbildung aus Methanol,
- b) $F_{420}H_2$ Dehydrogenase,
- c) Methyltransferasen zur Verwertung von C_1 -Substraten,
- d) Transkriptionsregulatoren,
- e) hypothetische und konservierte Proteine.

Auf die wichtigsten Klassen und die enthaltenen Proteine soll nachfolgend näher eingegangen werden, wobei die Funktion der Enzyme, der Regulatoren und der Transporter im Vordergrund stehen soll. Abb. 27 gibt einen zusammenfassenden Überblick über die Expressionsänderungen von Genen, die für Enzyme kodieren, die direkt an der Methanogenese aus Methanol oder Acetat beteiligt sind.

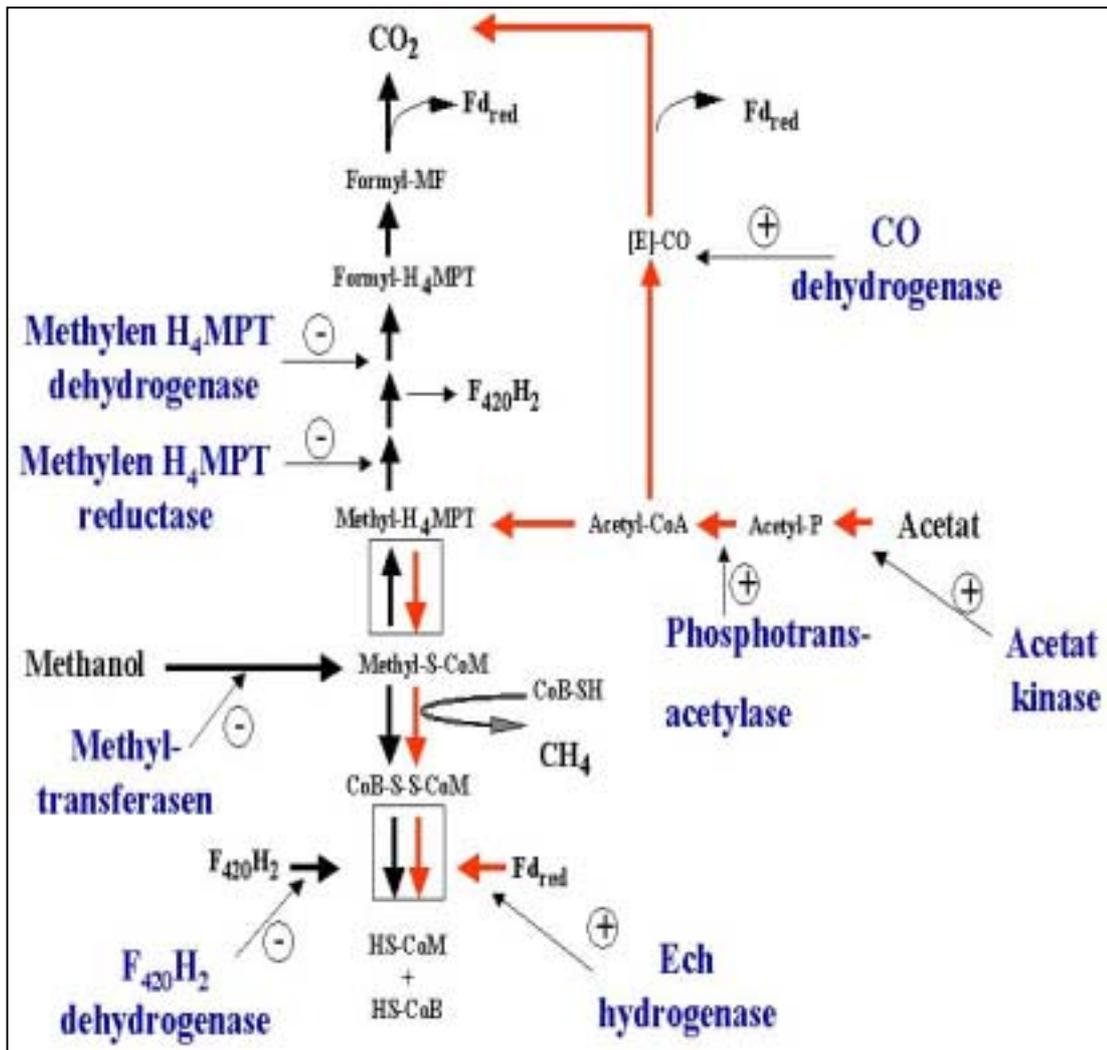


Abb. 27: Zusammenfassende Darstellung der Ergebnisse der genomweiten Transkriptionsanalyse beim Wachstum auf Methanol bzw. Acetat

Diese Abbildung zeigt die Stoffwechselwege der Methanogenese aus Methanol (schwarze Pfeile) und Acetat (rote Pfeile). Aufgeführt sind die wichtigsten Enzyme, die in diesen Stoffwechselwegen involviert sind. Das Expressionsniveau der korrespondierenden Gene dieser Enzyme beim Wachstum auf Acetat ist durch + bzw. - gekennzeichnet. (+) = hohe Expression beim Wachstum auf Acetat; (-) geringe Expression beim Wachstum auf Acetat (jeweils im Vergleich zum Transkriptionslevel beim Wachstum auf Methanol)

4.4.1 Enzyme des oxidativen Zweigs der Methanbildung aus Methanol

Wie bereits beschreiben, sind die Methylen-H₄MPT-Dehydrogenase, die Methylen-H₄MPT-Reduktase, die Methylen-H₄MPT-Cyclohydrolase und die Formyl-H₄MPT:MFR-Formyl-transferase an der Oxidation der Methylgruppe aus Methanol beteiligt. Ein Teil der Methylgruppen muss von den Zellen in Abwesenheit von Wasserstoff oxidiert werden, um Reduktionsäquivalente für die Reduktion des Cofactors F₄₂₀ zu gewinnen.

Aus Tabelle 10 (Kapitel 3) wird ersichtlich, dass der Expressionslevel aller Gene, die für die oben erwähnten Proteine kodieren, beim Wachstum der Zellen auf Methanol signifikant höher war als beim Wachstum auf Acetat. Dieses Ergebnis stimmt mit der Vermutung überein, dass die Methylen-H₄MPT-Dehydrogenase, die Methylen-H₄MPT-Reduktase, und die Formyl-H₄MPT:MFR-Formyltransferase für den acetoclastischen Weg der Methanogenese nicht benötigt werden (Abb. 27) und daher eine verminderte Transkription der entsprechenden Gene beobachtet wird. Zellen, die auf dem C₂-Körper Acetat wuchsen, benötigten zwar Methyl- und Formyl-Reste für die Biosynthese von Aminosäuren und Purinen/Pyrimidinen, diese können in *Ms. mazei* aber wahrscheinlich im Zuge der Umsetzung von Serin zu Glycerin gewonnen werden, wobei Methylen-THF entsteht. Diese Verbindung kann vermutlich mit Hilfe eines alternativen Sets von Enzymen, bestehend aus einer Methylen-THF-Dehydrogenase, einer Methylen-THF-Reduktase und einer Methenyl-THF-Cyclohydrolase, in Methyl- bzw. Formyl-THF umgewandelt werden. Interessanterweise, zeigen diese Proteine hohe Homologien zu bakteriellen Enzymen und wurden bislang in der Domäne der Archaea nur in Vertretern der Gattung *Methanosarcina* gefunden (Deppenmeier et al., 2002, Galagan et al., 2002). Für die Gene, die für THF-abhängige Enzyme kodieren, wurden keine unterschiedlichen Expressionsmuster beim Wachstum auf Methanol bzw. Acetat beobachtet, was die Hypothese bestätigt, dass diese Proteine eine Rolle im Anabolismus beim Wachstum auf Acetat und Methanol spielen.

4.4.2 F₄₂₀H₂ Dehydrogenase

Ebenso wie die oben erwähnten Proteine des oxidativen Zweigs der Methanogenese aus Methanol, ist die F₄₂₀H₂-Dehydrogenase nur an der Umsetzung von Methanol beteiligt. Das Enzym ist Bestandteil des F₄₂₀H₂:Heterodisulfid Oxidoreduktase-Systems, welches für die Oxidation von F₄₂₀H₂ und die Reduktion des Heterodisulfids beim Wachstum auf Methanol verantwortlich ist (Bäumer et al., 2000). Demgegenüber wird F₄₂₀H₂ im Zuge der Methanogenese aus Acetat nicht gebildet. Aus diesen Überlegungen wird deutlich, dass es keine offensichtliche Funktion für eine F₄₂₀H₂-Dehydrogenase im aceticlastischen Weg der Methanbildung gibt. In Übereinstimmung mit dieser Hypothese stehen die Ergebnisse der Genomweiten Transkriptionsanalyse und der „Real-Time PCR“. Beide Untersuchungsmethoden zeigen einen hohen relativen Expressionslevel der entsprechenden Gene in Methanol-gewachsenen Zellen und nur geringe Mengen an Transkripten in Acetat-gewachsenen Kulturen.

Die F₄₂₀H₂-Dehydrogenase wurde aus den Cytoplasmamembranen von *Ms. mazei*, *Methanobus tindarius* und dem Sulfat reduzierenden Archaeon *Archaeoglobus fulgidus* aufgereinigt (Abken und Deppenmeier, 1997; Haase et al., 1992; Westenberg et al., 1999; Kunow et al., 1994). Das Protein aus *Ms. mazei* besteht aus 13 unterschiedlichen Polypeptiden (Bäumer et al., 2000). Die Untereinheiten sind durch das *fpo*-Cluster kodiert, das die Gene für die zwölf Untereinheiten *fpo* A, B, C, D, H, I, J, K, L, M, N und O enthält (Bäumer et al., 2000). Das Gen, welches für die 37 kDa Untereinheit (Fpo F) des Enzyms kodiert, ist nicht Teil des Operons und ist an einer anderen Stelle des Chromosoms lokalisiert (Westenberg et al., 1999; Bäumer et al., 2000). Kürzlich, wurden die Gene der F₄₂₀H₂-Dehydrogenase von *Archaeoglobus fulgidus* im Rahmen des Genomprojektes dieses hyperthermophilen Archaeons sequenziert (Klenk et al., 1997). Das Operon, das für dieses Enzym kodiert, ist dem aus *Ms. mazei* sehr ähnlich mit der Ausnahme, dass *fpoF* – das Äquivalent zu *fpo F* – am Ende des Operons von *A. fulgidus* lokalisiert ist (Brüggemann et al., 2000). Die abgeleiteten Aminosäuresequenzen von *fpo F* und *fpo F* zeigten Ähnlichkeiten zu bestimmten Untereinheiten der F₄₂₀-reduzierenden Hydrogenase (Halboth und Klein, 1992) und zur F₄₂₀-abhängigen Format-Dehydrogenase (Bult et al., 1996; Smith et al., 1997). Deshalb wird angenommen,

dass FqoF und FpoF die Untereinheiten der F₄₂₀H₂-Dehydrogenase darstellen, welche im Stande sind, reduziertes F₄₂₀ zu oxidieren (Brüggemann et al., 2000). Diese Theorie ist inzwischen durch die heterologe Produktion der beiden Untereinheiten in *E. coli* und deren Charakterisierung bestätigt (Hofmann, 2003). Die abgeleiteten Aminosäuresequenzen der anderen Untereinheiten der F₄₂₀H₂-Dehydrogenase wurden mit den Sequenzen anderer Organismen und anderer phylogenetisch verwandten Enzyme verglichen. Für elf Untereinheiten, kodiert durch *fqo* A bis *fpo* N wurden verwandte Gegenstücke in bakteriellen und eukaryotischen NADH-Dehydrogenasen gefunden (Bäumer et al., 2000).

Sowohl das Gen *fpoF* als auch die Gene *fpoA-N* zeigten einen um den Faktor 74 höheren Expressionslevel beim Wachstum auf Methanol im Vergleich zu Acetat. Die Ergebnisse der Real-Time-PCR (Tabelle 12), die auf eine gleiche Menge an Transkripten für die F- und B-Untereinheiten hindeuteten, weisen auf eine simultane Regulation der beiden Gencluster hin. Diese simultane Regulation (An- und Abschaltung) der Gene lässt auf ein gemeinsames Regulatorprotein schließen.

4.4.3 Ech-Hydrogenase

Vor kurzem wurde eine neuartige Membran gebundene [NiFe] Hydrogenase (Ech) aus Acetat gewachsenen *Ms. barkeri* Zellen isoliert. Dieses Enzym ist sehr wahrscheinlich Teil eines Ferredoxin:Heterodisulfid-Oxidoreduktase Systems. Das aufgereinigte Enzym besteht aus sechs unterschiedlichen Untereinheiten (EchABCDEF) und katalysiert die reversible H₂-abhängige Reduktion eines 2[4Fe-4S] Ferredoxins aus *Ms. barkeri* (Meuer et al., 1999; Kunkel et al.; 1998). Weiterhin wurde gezeigt, dass das reduzierte Ferredoxin (Fd_{red}), welches im Zuge des methanogenen Acetyl-CoA Abbaus entsteht, durch die Ech-Hydrogenase reoxidiert wird (Meuer et al., 1999). Der Weg des Elektronentransfers von der Ech-Hydrogenase zu CoM-S-S-CoB ist jedoch bis heute nicht eindeutig geklärt. Derzeit werden zwei Hypothesen favorisiert: 1) Die Elektronen aus der Oxidation des reduzierten Ferredoxins werden zur Reduktion von Protonen verwendet. Dadurch würde molekularer Wasserstoff entstehen, welcher von der F₄₂₀-nichtreduzierenden Hydrogenase reoxidiert werden könnte. Die Elektronen gelangen dann über

Methanophenazin auf die Heterodisulfid- Reduktase. 2) Die Elektronen könnten direkt von der Ech-Hydrogenase auf Methanophenazin übertragen werden, welches dann von der Heterodisulfid-Reduktase reoxidiert wird. In jedem Fall ist kein Zweifel daran, dass die Ech Hydrogenase Bestandteil des Fd_{red} :Heterodisulfid Oxidoreduktase Systems ist. Die gesamte Reaktion, die von der Ech-Hydrogenase katalysiert wird, ist wahrscheinlich an den Aufbau eines Protonengradienten gekoppelt (Hedderich et al., 1998).

Kürzlich wurde postuliert, dass die Ech-Hydrogenase eventuell eine zweite Funktion in der Bildung und dem Abbau von Formyl-MFR haben könnte (Meuer et al., 2002). Im methylotrophen Stoffwechselweg der Methanogenese wird Formyl-MRF durch die Formyl-MFR-Dehydrogenase oxidiert, wobei CO_2 und reduziertes Ferredoxin entstehen. Es wird ferner angenommen, dass das reduzierte Ferredoxin durch die Ech-Hydrogenase oxidiert und Wasserstoff freigesetzt wird (Meuer et al., 2002). Unter Berücksichtigung des geringen H_2 -Drucks, welcher in der natürlichen Umgebung von methanogenen Organismen herrscht, ist die Wasserstoff-abhängige Reduktion des Oxidation von reduziertem Ferredoxin zu H_2 ein exergoner Prozess. Es wird daher angenommen, dass die Ech-Hydrogenase im Zuge der Reaktion Protonen über die Membran pumpt (Hedderich et al., 1998)

Die Hypothese wird durch die Ergebnisse der relativen Transkriptionsanalyse zwischen dem aceticlastischen und dem methylotrophen Weg der Methanogenese zumindest teilweise unterstützt. In den Real-Time-PCR-Experimenten war keine signifikante Änderung (1,2-fach) im Expressionsverhalten der Ech-Hydrogenase nachweisbar (siehe Tabelle 12). Dieser Befund deutet an, dass das Enzym in beiden Stoffwechselwegen von Bedeutung ist. In den DNA-Array-Experimenten jedoch wurde eine moderate Steigerung der Expression der Gene der sechs Untereinheiten der Ech-Hydrogenase beim Wachstum auf Acetat beobachtet (Tab. 13 E+E, Anhang D). Eine Erklärungsmöglichkeit ist, dass das Enzym beim Wachstum auf Acetat essentiell ist, da Protonen im Zuge der Reaktion nach außen gepumpt werden. Diese Protonentranslokation könnte von entscheidender Bedeutung für das Wachstum der Zellen sein, da die Spanne der Änderung der freien Energie beim Wachstum auf Acetat sehr klein ist ($-36 \text{ kJ/mol } CH_4$) und jede Form der Energiekonservierung genutzt werden muss. Anders liegen die Dinge beim Wachstum auf Methanol, da

hier die Änderung der freien Energie wesentlich höher ist (-106 kJ/mol CH₄). Es ist demnach vorstellbar, dass die Funktion der Ech-Hydrogenase im Methanolstoffwechsel von einer F₄₂₀-reduzierenden Hydrogenase übernommen werden kann, von der gezeigt wurde, dass auch sie reduziertes Ferredoxin unter Freisetzung von H₂ oxidieren kann (Fox et al., 1987). Dieses Enzym trägt jedoch nicht zur Energiekonservierung bei. Unter diesen Gesichtspunkten würde eine erhöhte Expression der Ech-Hydrogenase-Gene beim Wachstum auf Acetat von Vorteil sein.

4.4.4 Methyltransferasen

Die Methanbildung aus Methanol beginnt durch ein System löslicher Methyltransferasen (MtaA, MtaB, MtaC), die den Methylgruppentransfer von Methanol auf Coenzym M katalysieren (Sauer und Thauer, 1998/1999/2000; Sauer et al., 1997). Dabei fungiert MtaB als initiales Enzym, das die Methylgruppe von Methanol auf das Corrinoid-Protein MtaC transferiert. MtaB ist somit als Methanol:Corrinoid-Methyltransferase zu bezeichnen. Die methylierte Form des Corrinoid-Proteins dient dann dem Protein MtaA als Substrat zur Übertragung der Methylgruppe auf HS-CoM. Das MtaA Protein fungiert somit als Methyl-Corrinoid:HS-CoM-Methyltransferase. Die Genom-sequenzierung von *Ms. mazei* offenbarte allerdings, dass insgesamt drei Gencluster und ein Einzelgen existieren, die für Komponenten des Mta-Systems kodieren (Deppenmeier et al., 2002). Dies bedeutet, dass *Ms. mazei* theoretisch über drei Isoenzyme verfügt, die in der Lage sind, die Methylgruppe aus Methanol in den Prozess der Methanogenese einzuspeisen. In Tab. 14 sind die entsprechenden ORFs und deren Annotation aufgeführt.

Tabelle 14: Darstellung der MtaABC-Systeme zur Methanolverwertung

ORF-Nummer	Gen	Annotation	Verhältnis
0294	<i>mtaB1</i>	Methanol:Corrinoid-Methyltransferase	0,05
0295	<i>mtaC1</i>	Methanol-Corrinoid Protein	0,31
0831	<i>mtaA1</i>	Methylcobalamin:HS-CoM-Methyltransferase	0,08
0835	<i>mtaC3</i>	Methanol-Corrinoid Protein	.*
0836	<i>mtaB3</i>	Methanol:Corrinoid-Methyltransferase	1,17
2611	<i>mtaC2</i>	Methanol-Corrinoid Protein	1,81
2612	<i>mtaB2</i>	Methanol:Corrinoid-Methyltransferase	.*
2613	<i>mtaA2</i>	Methylcobalamin:HS-CoM-Methyltransferase	1,56

*Die Gene sind nicht auf dem Chip repräsentiert; doch sind sie in vermeidliche Operonstrukturen eingebunden, so dass davon ausgegangen werden muss, dass sich die Expressionslevel nicht sehr von den benachbarten Genen unterscheiden sollten.

Die Expressionsanalysen zeigten jedoch, dass nur eines der *mtaABC* Gen-Cluster eine veränderte Expression beim Wachstum auf Methanol im Vergleich zur Anzucht auf Acetat aufwies. Hierbei handelte es sich um die Gene 0294 (*mtaB1*), 0294 (*mtaC1*) und um ein Gen (0831), das sich außerhalb dieses Clusters befindet und für die Methy-Corrinoid:HS-CoM-Methyltransferase (*mtaA1*) kodiert. Diese Ergebnisse wurden durch Real-Time-PCR Untersuchungen bestätigt, die zeigten, dass die Expression des Gens für die Methyltransferase MtaB1 beim Wachstum auf Methanol um den Faktor 37 höher war, im Vergleich zum Wachstum mit Acetat als Kohlenstoff- und Energiequelle. Zusammengefasst deuten diese Ergebnisse an, dass die Methyltransferasen (MtaA1, MtaB1, MtaC1), ebenso wie die F₄₂₀H₂-Dehydrogenase, nur an der Methanbildung aus Methanol als Kohlenstoffquelle beteiligt sind. Alle anderen *mta*-Gene (Tab. 13 E+E, Anhang D) wiesen keine signifikante Änderung im Expressionsverhalten auf. Zwar lässt sich mittels der DNA-Chip Technologie keine direkte Aussage über die Menge der Transkripte machen, doch lassen die absoluten Werte der Fluoreszenzmessung der markierten DNA-Arrays den Schluss zu, dass beim Wachstum auf Methanol nur das System bestehend aus den Genen 0294, 0295 und 0831 aktiv ist.

Neben den Methyltransferasen für die Verwertung von Methanol existieren derartige Enzym-Systeme ebenfalls für die Nutzung von Methylaminen (Ferguson et al., 2000; Paul et al., 2000; Burke und Krzycki 1997/1998). Auch hier finden sich 2-3 Gencluster im Genom, die für Methyltransferasen für die Verwertung von Tri-, Di- und Monomethylamin kodieren. Interessanterweise zeigte keines dieser Gene eine signifikante Änderung des Expressionslevels beim Vergleich der Transkriptome aus Methanol- und Acetat-gewachsenen Zellen. Einige als hypothetische Methyltransferasen annotierte Gene (z.B. ORF 2929, ORF 2277) wiesen allerdings ein verändertes Expressionsverhalten auf. Da diese Methyltransferasen nicht eindeutig charakterisiert sind, stellen sie einen interessanten Ausgangspunkt für weitergehende Analysen der Methanogenese dar.

4.4.5 Enzyme des aceticlastischen Stoffwechselwegs der Methanogenese

Wie in 4.1.3 erwähnt, sind im aceticlastischen Stoffwechselweg der Methanbildung aus Acetat die Acetat-Kinase und die Phosphotransacetylase an der Aktivierung von Acetat beteiligt, wobei Acetyl-CoA gebildet wird. Die Acetyl-CoA-Synthetase/CO-Dehydrogenase spaltet anschließend Acetyl-CoA, wobei CO zu CO₂ oxidiert wird, und die freigesetzten Elektronen werden verwendet, um Ferredoxin zu reduzieren. Das Ferredoxin_{red} wird durch die Ech-Hydrogenase wieder reoxidiert (Meuer et al., 1997; Kunkel et al., 1998). Die Elektronen dienen letztendlich zur Reduktion des Heterodisulfids (siehe 4.4.3)

Die Gene für die Acetat-Kinase, die Phosphotransacetylase und die Acetyl-CoA-Synthetase/CO-Dehydrogenase zeigten in der Real-Time-PCR und den Microarray-Experimenten ein höheres Expressionsniveau in Zellen, die mit Acetat als Kohlenstoff- und Energiequelle wuchsen (siehe Tabelle 10/12 und Tabelle 13 in E+E und Anhang D). An dieser Stelle wird deutlich, dass ein umfangreiches Regulationsnetzwerk in *Ms. mazei* vorhanden sein muss, das in Abhängigkeit des Wachstumssubstrats arbeitet. Steht nur Acetat zur Verfügung, so werden alle Gene, die speziell an der Umsetzung des C₂-Körper im aceticlastischen Stoffwechselweg der Methanogenese beteiligt sind, stark exprimiert.

4.4.6 Biosynthese von Aminosäuren und Dipeptid/Oligopeptid

Aufnahmesysteme

Die Auswertung der Microarray-Experimente zeigte eine erhöhte Menge an Transkripten für die Proteinsynthese von Enzymen, die an der Synthese von aromatischen Aminosäuren beteiligt sind, wenn *Ms. mazei* auf Acetat angezogen wurde. Zudem wurde mehr mRNA für die Produktion von Di- und Oligopeptide Transportern gebildet. Eine mögliche Erklärung hängt mit den thermodynamischen Rahmenbedingungen des aceticlastischen Stoffwechselweges zusammen. Fakt ist, dass beim Wachstum auf Acetat die geringste Änderung der freien Energie mit -36 kJ/mol auftritt. Trotzdem muss der ATP-Spiegel und das Membranpotential hoch gehalten werden, um lebenswichtige Zellfunktionen (Proteinsynthese, Nucleotidsynthese, Ionentransport) aufrechtzuerhalten. D.h. die Zellen müssen alle verfügbaren Ressourcen nutzen, um optimal wachsen zu können. Dies impliziert natürlich, das Intermediate des Anabolismus, die in Form von Hefeextrakt und Casition im Medium zur Verfügung stehen, genutzt werden. In diesem Zusammenhang ist die vermehrte Produktion von mRNA zur Synthese von Transportern zur Aufnahme von Di- und Oligopeptiden zur Bereitstellung von Aminosäuren sinnvoll. Die Energie aufwendige *de novo* Synthese von Aminosäuren kann auf diese Weise vermieden werden. Wie bereits in Kapitel 3.4.1 erwähnt, scheinen *Ms. mazei*-Zellen keine aromatischen Aminosäuren aus dem Medium aufnehmen zu können. Dies trifft möglicherweise auch auf Dipeptide und Oligopeptide zu, die aromatische Aminosäuren enthalten. Diese Zusammenhänge könnten dazu führen, dass Phe, Tyr und Trp in geringeren Mengen in der Zelle vorkommen, im Vergleich zu den anderen Aminosäuren, die aus dem Medium aufgenommen werden können. Dies würde zu einem relativen Mangel an aromatischen Aminosäuren führen, da der Konzentration aller Aminosäuren als Synthesebausteine in etwa gleich hoch sein sollte. Nun ist es vorstellbar, das *Ms. mazei* über sensorische Systeme verfügt, die dieses Missverhältnis erfassen. Entsprechende regulatorische Proteine könnten nachfolgend zur vermehrten Produktion von mRNA zur Synthese von Proteinen zur Herstellung von aromatischen Aminosäuren beitragen. Dies würde erklären, warum die Gene zur Kodierung von Proteinen der aromatischen Aminosäureproduktion beim Wachstum auf Acetat hochreguliert sind.

4.4.7 Detoxifizierung im aceticlastischen Stoffwechselweg

Die Expression des Gens für die Acetaldehyde:Ferredoxin-Oxidoreduktase war beim Wachstum mit Acetat 38-fach höher als beim Wachstum mit Methanol, was darauf schließen lässt, dass das entsprechende Enzym im aceticlastischen Weg der Methanogenese eine bedeutende Rolle spielt. Bei der Methanbildung aus Acetat wird der C₂-Körper zuerst durch die Acetat-Kinase aktiviert. Aus dem entstandenen Acetylphosphat wird anschließend durch die Phosphotransacetylase Acetyl-CoA gebildet. Die CO-Dehydrogenase wiederum spaltet die CoA-Verbindung zu Methyl-H₄MPT und enzymgebundenes CO. In den letzten beiden Schritten der Reaktionsfolge kann Acetaldehyd als Nebenprodukt entstehen, welches für die Zelle toxisch ist. Um diesem Effekt entgegen zu wirken, könnte die Acetaldehyde:Fd-Oxidoreduktase die Oxidation der Aldehydgruppe des Acetaldehyds zu einer Carboxylgruppe katalysieren, wobei Acetat entsteht. Sollte diese Hypothese zutreffen, dann stellt die Acetaldehyde:Ferredoxin-Oxidoreduktase ein Acetat-abhängiges Detoxifizierungssystem dar.

4.4.8 Verwertung von Methylthiolen

Dimethylsulfide (DMS) und verwandte methylierte Thiole spielen eine wichtige Rolle in marinen und Brackwasser- Ökosystemen. Den Dimethylsulfiden aus den Ozeanen wird die wichtigste Rolle in der Kondensation von Wolken in der marinen Atmosphäre zugesprochen, was starke Auswirkungen auf das Klima hat (Charlson et al., 1987). Die Hauptquelle für DMS und andere methylierte Thiole ist Dimethyl-Sulfoniopropionat (DMSP), eine Verbindung, die in diversen Organismen wie Algen, Riffkorallen und Pflanzen vorkommt (Kirst et al., 1989; Paquet et al., 1995). DMSP kann durch zwei metabolische Wege in der Natur direkt umgesetzt werden (Visscher et al., 1994). 1) Es kann von Bakterien und Algen durch DMSP-Lyasen zu DMS und Acrylsäure gespalten werden. (De Souza et al., 1996; De Souza et al., 1995; Ledyard et al., 1993). 2) Alternativ, kann DMSP unter anaeroben und aeroben Bedingungen auch sequentiell demethyliert werden. Hierbei entstehen 3-Methyl-Mercaptopropionat (MMPA) und 3-Mercapto-Propionsäure (MPA) (Taylor und Gilchrist, 1991; Visscher et al., 1995). Methanogene Organismen, die im Stande

sind, Methan aus methylierten Thiolen zu produzieren, verwenden normalerweise auch andere methylotrophe methanogene Substrate wie Methylamine, Acetat oder Methanol. Coenzym M wird hierbei, in einem Zwischenschritt bei der Bildung von Methan und Kohlendioxid, methyliert (Ferry, 1992). Dieser Schritt wird von der Methylthiol:CoM-Methyltransferase katalysiert. In *Ms. barkeri* wurde nachgewiesen, dass dieses Enzym beim Wachstum auf Acetat als Kohlenstoff- und Energiequelle (ein 480kDa Corrinoïd-Protein) aktiv ist, aber nicht bei Wachstum auf Methanol (Tallan und Krzycki, 1997, Paul und Krzycki, 1996). Diese Beobachtungen stimmen mit den Ergebnissen der DNA-Microarray Experimente überein, in denen die Gene für die Methylthiol:CoM-Methyltransferase in Acetat-gewachsenen Zellen um den Faktor sechs stärker exprimiert waren. Da in den Medien keine methylierten Thiole als Substrat enthalten waren, wurde eine Coregulation der Gene der Methylthiol:CoM-Methyltransferase mit anderen Genen des acetoclastischen Stoffwechselweges postuliert.

4.4.9 Kobalt-Aufnahme

Die Kobalt-Komplexierung ist für die Umwandlung von Porphyrinoiden zu Corrinoïden und damit zur Bildung von Cobamiden notwendig. Cobamide sind für viele zentrale Prozesse der prokaryotischen und eukaryotischen Zellen notwendig. Hierunter fallen die Desoxyribonucleotid-Synthese, die Synthese modifizierter tRNAs und die Aminosäure-Biosynthese (Halpern, 1985). Während umfangreiche Studien mit Bakterien durchgeführt wurden, ist über die Synthese von Cobamiden in Archaeen nur wenig bekannt (Blanche et al., 1991; Thomas und Escalante-Semerena, 2000). Es ist jedoch eindeutig, dass viele Archaeen Cobamide synthetisieren. Zum Beispiel benötigen methanogene Archaea Cobamide zur Methanogenese aus H₂ und CO₂, Acetat oder Methanol (DiMarco et al., 1990). Für die CO-Dehydrogenase/Acetyl-CoA-Synthase aus *Methanosarcina thermophila* wurde nachgewiesen, dass das Enzym mindestens aus zwei Protein-Komplexen besteht. Es handelt sich hier um eine CO-oxidierende Nickel/Eisen-Schwefel (Ni/Fe-S) Komponente und eine Kobalt-enthaltene Corrinoïd/Eisen-Schwefel Komponente (Co/Fe-S). Dieser Enzymkomplex ist, wie oben beschrieben, essenziell im acetoclastischen Stoffwechselweg der Methanogenese. Unter Berücksichtigung der

Notwendigkeit für den Einbau von Kobalt in die CO-Dehydrogenase/Acetyl-CoA-Synthase und der höheren Expression der Gene dieses Enzyms beim Wachstum mit Acetat als Kohlenstoff- und Energiequelle, war davon auszugehen, dass die Gene für die Aufnahme von Kobalt ebenfalls höher exprimiert waren. Die Auswertung der vergleichenden DNA-Microarray-Expressionsanalyse ergab, dass Aufnahmesysteme für Kobalt um bis zum Faktor sieben höher exprimiert waren, wenn *Ms. mazei* Gö1 mit Acetat als Kohlenstoffquelle wuchs.

4.4.10 Eisen(III)-Dicitrat-Aufnahme

Das Aufnahmesystem von dreiwertigem Eisen über das Eisen(III)-Dicitrat Siderophor-System wird durch das Fec-Operon kodiert und ist in *E. coli* nur unter aeroben Bedingungen aktiv (Kammler et al., 1993). In den Microarray-Experimenten war die Expression der Gene für dieses Aufnahmesystem in *Ms. mazei* Gö1 beim Wachstum auf Acetat bis zum Faktor zehn höher als beim Wachstum auf Methanol. Dieses Ergebnis lässt darauf schließen, dass in gewissen natürlichen Umgebungen von *Ms. mazei* Gö1, z. B. an der oberen Grenze von Teichsedimenten, eventuell dreiwertiges Eisen vorliegen kann. Dies bedeutet allerdings eine höhere Menge an Sauerstoff in der Umgebung. *Ms. mazei* Gö1 besitzt jedoch Gene für Proteine, die eine Toleranz gegenüber Sauerstoff vermitteln. Hierunter fallen z. B. die Katalase und die Superoxid-Dismutase (Deppenmeier et al., 2002). Die Kopplung dieses Aufnahmesystems an den acetoclastischen Weg der Methanogenese wurde durch Untersuchungen von ‚Funktionellen Clustern‘ des Eisen(III)-Dicitrat-Aufnahmesystems in der ERGO-Data-Suite weiter verifiziert (nicht gezeigt). Bei einer solchen Analyse wird überprüft, an welche anderen Expressionen diese Gene in anderen Organismen gekoppelt sind. Das Clustern ergab, dass die Funktion des Fec-Systems an Cobalamin-Bindungssysteme (in *Pyrococcus horikoshii* und *Pyrococcus abyssi*), 2-Ketosäure:Ferredoxin-Oxidoreduktasen (in *Pyrococcus furiosus*) und an ein Nickelinsertionsystem (CooC) der CO-Dehydrogenase in *Methanococcus janaschii* gekoppelt ist. Da die Expression der Gene für die CO-Dehydrogenase, die Cobalaminsynthese und die Acetaldehyd:Ferredoxin-Oxidoreduktase in den Experimenten dieser Arbeit beim Wachstum mit Acetat ebenfalls höher war, ließ dies auf eine Coexpression des Aufnahmesystems für dreiwertiges Eisen schließen.

4.4.11 Vermutliche Regulatoren

Regulatorproteine sind die Schlüssel in der Regulation der Expression von Genen. An dieser Stelle muss erwähnt werden, dass mehr als 80 ORFs im Genom von *Ms. mazei* als wahrscheinliche Regulatorproteine annotiert wurden (Deppenmeier et al., 2002). Bis auf wenige Ausnahmen ist die Funktion dieser putativen Regulatoren unbekannt. Die genomweite Transkriptionsanalyse unter den beiden untersuchten Wachstumsbedingungen Minimal/Komplex-Medium und Metanol/Acetat-Verwertung zeigte zudem nur wenige regulatorische Gene mit signifikanten Unterschieden im Expressionsverhalten. Hierfür könnten zwei Gründe verantwortlich sein: 1) Die regulatorischen Proteine haben keine Funktion in den oben erwähnten Vergleichsexperimenten und die zugehörigen Gene werden daher nicht unterschiedlich exprimiert. 2) Die Regulatorproteine werden konstitutiv gebildet und werden auf Proteinebene ein- bzw. ausgeschaltet. Die korrespondierenden Gene unterliegen keiner direkten Expressionsregulation.

Jedoch fielen auch einige Gene mit verändertem Expressionsverhalten auf, deren zugehörige Proteine als Regulatoren annotiert wurden. Ein Beispiel hierfür ist das Gen 0660 als Bestandteil des *feo*-Systems, das für ein putatives Fe^{2+} -abhängiges Regulatorprotein kodiert (Kapitel 3.3.2). Die Expression dieses Gens war äquivalent zu der Funktion des korrespondierenden Proteins. Es wurde festgestellt, dass die Transkriptmengen der *feo*-Gene beim Wachstum auf Minimalmedium stark vermindert waren. Demgegenüber war das relative Transkriptionsniveau des Gens 0660 unter diesen Bedingungen stark erhöht. Es ist daher wahrscheinlich, dass der Regulator als Repressor der Transkription des *feo*-Operons dient. Ein ähnliches Beispiel ist das Gen 0837 als Bestandteil des *mtaB3/mtaC3*-Genclusters, das für ein Methanol-Corrinoid-Protein und eine Methanol:Corrinoid-Methyltransferase kodiert. Das Expressionsverhältnis beim Wachstum auf Acetat im Vergleich zu Methanol betrug hier 12:1. Somit ist auch hier davon auszugehen, dass das entsprechende Protein als Repressor der *mta*-Cluster dient (siehe Kapitel 4.4.4).

Ferner wurde ein verändertes Expressionsverhalten von Genen in *Ms. mazei* beobachtet, die aufgrund ihrer Proteinsequenzen als putative Regulatoren annotiert wurden. Diesen Proteinen ist keine definierte Funktion zugewiesen. Als Beispiele seien hier das Gen 1308 (Expressionsverhältnis Acetat zu Methanol = 1:0,44) erwähnt. Durch die nicht geklärte Funktion dieser Regulatoren stellen diese ein weiteres interessantes Gebiet für zukünftige Analysen der genetischen Regulation in Archaeen dar.

Als weiteres Beispiel für ein verändert exprimiertes Gen soll an dieser Stelle auf den ORF 1987 eingegangen werden, dessen Protein zur Familie der MerR-Regulatoren gehört. Das Gen zeigte beim Wachstum auf Acetat die stärkste Änderung im Expressionsniveau (siehe Anhang D). Ursprünglich wurde den Proteinen der MerR-Familie eine Funktion als Transkriptionsregulatoren in der Schwermetallantwort in Bakterien zugesprochen (Summers, 1992). Archaeen besitzen diese Repressorfamilie normalerweise nicht, was auf einen lateralen Gentransfer dieses Gens aus einem Bakterium in *Ms. mazei* Gö1 schließen lässt (Brown et al., 2003; Deppenmeier et al., 2002). Nach neusten Erkenntnissen haben einige Regulatoren der MerR-Familie Funktionen in der Antwort auf Umgebungsstimuli wie z. B. oxidativem Stress (Amabile-Cuevas und Demple, 1991; Nunoshiba et al., 1992), Antibiotika (Holmes et al., 1993) oder Schwermetallstress (Lund et al., 1986; Lund und Brown, 1989). Wie in 4.4.11 erwähnt, kann der aceticlastische Stoffwechselweg der Methanogenese in *Ms. mazei* an den natürlichen Standorten des Organismus eventuell mit oxidativem Stress verbunden sein, weshalb die höhere Expression des ORFs 1987 sinnvoll ist.

Aus diesen Darstellungen wird auch ein großer Vorteil der DNA-Chip-Technologie deutlich, der das Auffinden von regulatorischen Genen/Proteinen betrifft. Es kann sicher behauptet werden, dass ohne diese Technik die Identifizierung der Gene 0660, 0837, 1308 und 1987 schwierig gewesen wäre. Die unterschiedlichen Expressionsniveaus dieser Gene unter verschiedenen Wachstumsbedingungen bieten interessante Ansätze für weiterführende Arbeiten, die die Funktion der korrespondierenden Regulatoren im Stoffwechsel näher beleuchten können.

4.5 Gene mit gleichem Expressionsmuster beim Wachstum von Zellen auf Methanol oder Acetat

4.5.1 Methyl-CoM Reduktase und Methyl-H₄MPT:HS-CoM Methyltransferase

Wie bereits erwähnt, wird im aceticlastischen Stoffwechselweg der Methanogenese Acetat von der Acetat-Kinase zuerst aktiviert und von der Phosphotransacetylase die Acetylgruppe auf Coenzym A transferiert. Das entstandene Acetyl-CoA wird von der CO-Dehydrogenase gespalten und die Methylgruppe auf H₄MPT übertragen (Ferry, 1995; Ferry, 1997). Die Methylgruppe des Methyl-H₄MPT wird, wie bei der Methanogenese aus Methanol von dem gleichen Enzym, der Methyl-H₄MPT:HS-CoM Methyltransferase, auf Coenzym M übertragen. Diese exergone Reaktion ($\Delta G^{\circ} = -29$ kJ/mol) ist an der Bildung eines elektrochemischen Natriumionen-Gradienten gekoppelt. Der nächste Schritt ist ebenfalls mit dem methylotropen Stoffwechselweg identisch. Im Zuge einer reduktiven Demethylierung durch die Methyl-CoM-Reduktase (Ellermann et al., 1988; Hartzell et al., 1988; Ermler et al., 1997) wird mit Hilfe des Elektronendonators HS-CoB (Noll et al., 1986; Kobelt et al., 1987) Methyl-S-CoM zu Methan und einem aus HS-CoM und HS-CoB bestehenden Heterodisulfid (CoM-S-S-CoB) umgesetzt (Jablonski und Ferry, 1991).

Aufgrund der Tatsache, dass die Methyl-CoM Reduktase und die Methyl-H₄MPT:HS-CoM Methyltransferase für den methylotropen und den aceticlastischen Stoffwechselweg der Methanogenese benötigt werden, wurde keine Veränderung in der Expression der Gene für diese beiden Enzyme erwartet. Die relativen Transkriptionswerte dieser Gene in der Auswertung der DNA-Microarray-Experimente ergab tatsächlich keine signifikante Veränderung in der Expression der für diese Enzyme kodierenden Gene (siehe Tabelle 13 E+E und Anhang D).

4.5.2 F₄₂₀H₂-nicht-reduzierende Hydrogenase

Beim Wachstum von *Ms. mazei* Gö1 auf H₂ + CO₂ spielt das H₂:CoB-S-S-CoM-Oxidoreduktase-System eine entscheidende Rolle. Es setzt sich aus der membrangebundenen, F₄₂₀-nichtreduzierenden Hydrogenase und der Heterodisulfid-Reduktase zusammen, die über den membranintegralen Elektronenüberträger Methanophenazin gekoppelt sind (Deppenmeier et al., 1996; Brodersen et al., 1998; Beifuß et al., 2000; Beifuss und Tietze, 2000). Die F₄₂₀-nichtreduzierende Hydrogenase zeigt im gereinigten Zustand im nativen Gradienten-Gel eine Bande bei 77 kDa mit Hydrogenase-Aktivität und im SDS-Gel zwei Banden mit den molekularen Massen von 40 kDa und 60 kDa (Deppenmeier et al., 1992). Ein Mol Protein enthält 0,8 mol Ni, 15 mol Nicht-Häm-Eisen und 13 mol säurelabilen Schwefel (Deppenmeier et al., 1992). Das gereinigte Enzym zeigte zwar mit Methylviologen eine Aktivität von 107,2 U/mg Protein, konnte aber im Gegensatz zur membrangebundenen Form 2-Hydroxyphenazin nicht umsetzen (Brodersen, 1998). Es wurde daher postuliert worden, dass *in vivo* mindestens eine weitere Komponente am Elektronentransport beteiligt ist. Dafür spricht auch, dass das Hydrogenase-Operon aus drei Genen besteht (Deppenmeier et al., 1995). Das Gencluster *whoGAC* kodiert für die 40 kDa-Untereinheit (*whoG*), die 60 kDa-Untereinheit (*whoA*) und für ein b-Typ-Cytochrom (*whoC*). Dabei fällt die Struktur des Operons auf, die, ebenso wie die abgeleiteten Aminosäuresequenzen der Gene, Homologien zu membrangebundenen Hydrogenasen von verschiedenen Bakterien aufweist. So konnte für *Ralstonia eutropha* gezeigt werden, dass das b-Typ-Cytochrom als physiologischer Elektronenakzeptor des Hydrogenase-Kernenzym fungiert und verantwortlich für die Einschleusung von Elektronen in die Atmungskette ist (Kortlücke et al., 1992).

Die F₄₂₀-nichtreduzierenden Hydrogenase könnte jedoch auch eine Funktion bei der Methanbildung aus Methanol und Acetat haben. Wenn die in Kapitel 4.4.3 vorgestellte Hypothese von Hedderich et al. korrekt ist, dann verwertet dieses Enzym den von der Ech-Hydrogenase produzierten molekularen Wasserstoff aus der Oxidation von reduziertem Ferredoxin. Dieser reduzierte Elektronenüberträger wird im Fall der Methanbildung aus Methanol im Zuge der Oxidation von Formyl-MFR und im Fall der Methanogenese aus Acetat von der CO-Dehydrogenase gebildet.

Weiterhin wurde schon in früheren Arbeiten gezeigt, dass die F₄₂₀-nicht-reduzierende Hydrogenase konstitutiv gebildet wird (Deppenmeier, 1995). In Übereinstimmung mit diesen Befunden stehen die Microarray-Experimente in dieser Arbeit, die keinen signifikanten Expressionsunterschied der Gene *whoGAC* in *Ms. mazei* Gö1 zeigten (siehe Tabelle 13 und Anhang D).

5. Zusammenfassung

(1) Im Rahmen dieser Arbeit wurden verschiedene Computerprogramme entwickelt, die es ermöglichten, DNA-Microarray-Chips zu produzieren, die das gesamte Genom von *Methanosarcina mazei* Gö1 repräsentierten. Ein Teil der neuentwickelten Software diente dazu, die ‚Shotgun‘-Genbank des Göttinger Genom Labors zur Generierung der PCR-Produkte zur Chip-Produktion zu nutzen. Ein anderer Teil der neuentwickelten Programme wurde genutzt, um die immensen Datenmengen, die im Rahmen dieser Arbeit generiert wurden, zu verarbeiten und zu verwalten. Die Programmiersprache PERL hat sich hierbei als sehr effektiv erwiesen.

(2) Zur Erzeugung von PCR-Produkten jedes einzelnen Gens von *Ms. mazei* wurde auf eine Plasmid-Genbank zurückgegriffen, die zur Genomsequenzierung des Organismus eingesetzt worden war. Die Inserts dieser Plasmide dienten als Template in der PCR-Reaktion, um die DNA-Microarray-Chips herzustellen. Der Vorteil dieses Ansatzes war, dass nur ein genspezifischer Primer benötigt wurde, um die PCR-Produkte zu generieren. Der zweite Primer wurde aus der Randsequenz des Vektors abgeleitet und war somit universell einsetzbar. Ferner lag ein großer Vorteil dieser Genbank darin, dass die Plasmide bekannte Inserts enthielten und damit das Risiko der Erzeugung von falschen PCR-Produkten minimiert wurde.

(3) Da für die Erstellung von Microarrays große Mengen an PCR-Produkten benötigt wurden, und weil Kooperationen mit anderen Arbeitsgruppen bestanden, wurden die unter Punkt (2) generierten PCR-Produkte in einen TOPO-Vektor kloniert. Diese TOPO-Genbank ermöglichte es, mit zwei universellen Primern alle PCR-Produkte in großen Mengen zu gewinnen. Durch diese Vorgehensweise ist auch in Zukunft die Nachproduktion von DNA-Microarray-Chips gewährleistet.

(4) Die im Rahmen dieser Arbeit erstellten DNA-Microarray-Chips von *Ms. mazei* Gö1, zeigten einen sehr hohen Grad an Genauigkeit in der Analyse des relativen Expressionsverhaltens des Genoms unter verschiedenen Wachstumbedingungen. Dies wurde bereits in Vorversuchen deutlich, in denen die Verhältnisse der Transkriptmengen beim Wachstum auf Komplexmedium im Vergleich zum Wachstum auf Minimalmedium bestimmt wurden. Es zeigte sich, dass die

Konzentration von ca. 10 % der Gesamt-mRNA-Spezies beim Wachstum der Zellen auf Minimalmedium um mehr als den Faktor drei erhöht war.

(5) Viele Gene mit einer erhöhten Expression beim Wachstum auf Minimalmedium kodierten für Enzyme, die für Biosynthesen von Aminosäuren und Nukleotiden verantwortlich sind. Zudem wurde deutlich, dass zahlreiche andere Gene beim Wachstum auf Minimalmedium vermehrt exprimiert wurden, die für Proteine kodieren, die wahrscheinlich an der Speicherstoffnutzung, dem Transport oder an der Regulation des Stoffwechsels beteiligt sind. Diese Chips ermöglichten somit, einen tiefen Einblick in das Expressionsverhalten der Gene des zentralen Metabolismus dieses methanogenen Organismus zu erlangen.

(6) Hingegen wurden nur wenige Gene gefunden, die beim Wachstum auf Vollmedium eine höhere Expression als bei der Anzucht der Zellen auf Minimalmedium aufwiesen. Ein wichtiges Beispiel für diese Klasse von Genen war das *feo*-Operon, dessen korrespondierende Proteine wahrscheinlich für die Aufnahme von zweiwertigen Eisenionen verantwortlich sind.

(7) In einem weiteren Teilprojekt wurde die Expression aller Gene aus *Ms. mazei* Gö1 beim Wachstum auf unterschiedlichen Kohlenstoff- und Energiequellen (Methanol und Acetat) untersucht. Zu den Genen, die beim Wachstum auf Methanol um mindestens den Faktor drei im Vergleich zu Acetat-Zellen hochgefahren wurden (insgesamt 53 Gene), zählten solche, die für die folgenden Proteine kodieren: Enzyme des oxidativen Zweigs der Methanbildung aus Methanol, die Untereinheiten der $F_{420}H_2$ -Dehydrogenase, Methyltransferasen zur Verwertung von C_1 -Substraten, Transkriptionsregulatoren, hypothetische und konservierte Proteine.

(8) Bei der Methanogenese aus Acetat waren insgesamt 118 Gene um mehr als den Faktor drei höher exprimiert im Vergleich zum Wachstum auf Methanol. Die korrespondierenden Proteine erfüllten folgende Funktionen: Enzyme des aceticlastischen Stoffwechselweges der Methanogenese, Proteine für die aromatische Aminosäure-Biosynthese, Methylthiol:Coenzyme M-Methyltransferasen, Aldehyd: Ferredoxin-Oxidoreduktase und Aufnahmesysteme für Kobalt und Dipeptide/

Oligopeptide sowie Transkriptionsregulatoren und viele hypothetische und konservierte Proteine.

(9) Für die abschließenden Auswertungen und Dokumentationen der erhaltenen Expressionsdaten erwies sich die Analyst Expressionist Software als sehr effektiv. Die erhaltenen Daten wurden in der integrierten Oracle Datenbank abgelegt und durch Cluster-Algorithmen zu funktionellen Gruppen zusammengefügt, die auch aus biologischer Sicht sinnvoll waren.

(10) Zur Überprüfung der Daten der Microarray-Analyse und zur Bestimmung der absoluten Transkriptmenge von Genen erwies sich die Real-Time-PCR als zuverlässige Methode. Basierend auf den Ergebnissen der DNA-Microarrays wurden durch diese Methode acht Gene untersucht, die ein unterschiedliches Expressionsverhalten beim Wachstum auf Methanol bzw. Acetat aufwiesen. Alle mittels Real-Time-PCR analysierten Transkripte zeigten das gleiche relative Expressionsverhalten wie in den DNA-Microarray-Experimenten.

6. Literaturverzeichnis

Abken, H.-J., Deppenmeier, U., 1997.

Purification and properties of an F₄₂₀H₂-Dehydrogenase from *Methanosarcina mazei* Gö1.

FEMS Lett. 154: 231-237.

Abken, H.J., 1997.

Reinigung und Charakterisierung der F₄₂₀H₂-Dehydrogenase und Identifizierung von Methanophenazin als neuartigen methanogenen Cofaktor.

Dissertation, Universität Göttingen.

Abken, H. J., Tietze, M., Brodersen, J., Bäumer, S., Beifuß, U., Deppenmeier, U., 1998.

Isolation and Characterization of Methanophenazine and Function of Phenazine in Membrane-Bound Electron Transport of *Methanosarcina mazei* Gö1.

Journal of Bacteriology 180: 2027-2032.

Alwine, J., D. Kemp, and G. Stark. 1977.

Method for the detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl paper and hybridization with DNA probes.

Proc. Natl. Acad. Sci. USA 74:5350-5354

Amabile-Cuevas, C.F. and Demple, B., 1991.

Molecular characterization of the soxRS genes of *Escherichia coli*: two genes control a superoxide stress regulon.

Nucleic Acids Research, Vol 19, Issue 16 4479-4484.

Andrade, S.L., Brondino, C.D., Kamenskaya, E.O., Levashov, A.V. and Moura, J.J., 2003.

Kinetic behavior of *Desulfovibrio gigas* aldehyde oxidoreductase encapsulated in reverse micelles.

Biochem Biophys Res Commun. 15;308(1):73-8.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidmann, J. G., Smith, J. A. and Struhl, K., 1987.

Current protocols in molecular biology.

Green Publishing Associates and Wiley-Interscience. J. Wiley & Sons, New York.

Axel, R. , Feigelson, P., Schutz, G., 1976.

Analysis of the complexity and diversity of mRNA from chicken liver and oviduct.

Cell 7, 247-254

Bachmann, B. J., 1987.

Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12.

In: Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M. and Umberger, H. E. (eds.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology.

American Society for Microbiology, Washington, D.C., 1191–1219.

Banfield, J.D. and Raftery, A.E., 1993.

Model based Gaussian and non-Gaussian clustering.

Biometrics, 49:803-821.

Barns, S. M., Delwiche, C. F., Palmers, J. D., Pace, N. R., 1996.

Perspectives on archaeal diversity, thermophily and monophily from environmental rRNA sequences.

Proc. Natl. Acad. Sci. USA 93: 9188-9193.

Bäumer, S., Murakami, E., Brodersen, J., Gottschalk, G., Ragsdale, S. W., Deppenmeier, U., 1998.

The $F_{420}H_2$:heterodisulfide oxidoreductase system from *Methanosarcina* species. FEBS Lett. 428: 295-298.

Bäumer, S., Ide, T., Jacobi, C., Johann, A., Gottschalk, G. and Deppenmeier, U., 2000.

The $F_{420}H_2$ Dehydrogenase from *Methanosarcina mazei* is a redox-driven proton pump closely related to NADH Dehydrogenases.

J. Biol. Chem. 275: 17968-17973.

Becker, D. F., Ragsdale, S. W., 1998.

Activation of methyl-S CoM reductase to high specific activity after treatment of whole cells with sodium sulfide.

Biochemistry 37: 2639-2647.

Beifuß, U., Tietze, M., 2000.

On the total synthesis of (S)-methanophenazine and the formal synthesis of (R)-methanophenazine from a common precursor.

Tetrahedron Letters. 41: 9759-9763.

Beifuß, U., Tietze, M., Bäumer, S., Deppenmeier, U., 2000.

Methanophenazine: Structure, Total Synthesis, and Function of a New Cofactor from Methanogenic Archaea.

Angewandte Chemie. 39: 2470-2472.

Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J. and Ouellette, B.F., 1998.

Genbank.

Nucleic Acids Research, 26(1): 1-7.

Bilban, M., Buehler, L.K., Head, S., Desoye, G., Quaranta, V., 2002.

Normalizing DNA microarray data.

Curr Issues Mol Biol. 2002 Apr;4(2):57-64.

Blanche, F., Robin, C., Couder, M., Faucher, D., Cauchois, L., Cameron, B. and Crouzet, J., 1991.

Purification, characterization, and molecular cloning of S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase from *Methanobacterium ivanovii*.

J. Bacteriol. 173:4637-4645 .

Bobik, T. A., Olson, K. D., Noll, K. M., Wolfe, R. S., 1987.

Evidence that the heterodisulfide of coenzyme M and 7-mercaptoheptanoylthreonine phosphate is a product of the methylreductase reaction in *Methanobacterium*.

Biochem. Biophys. Res. Commun. 149: 455-460.

Boone, D. R., Whitman, W. B., Rouviere, P. E., 1993.

Diversity and taxonomy of methanogens.

In: Ferry, J.G. (ed.), Methanogenesis, Chapman & Hall, New York/London: 35-80.

Brodersen, J., Gottschalk, G., Deppenmeier, U., 1998.

Membrane-bound $F_{420}H_2$ -dependent heterodisulfide reduction in *Mc. voltae*.

Arch. Microbiol. 171: 115-121.

Brown, J. R., Doolittle, W. F., 1997.

Archaea and the Prokaryote-to-Eukaryote transition.

Mikrobiol. Mol. Biol. Rev. 61: 457-502.

Brown, N.L., Stoyanov, J.V., Kidd, S.P., Hobman, J.L., 2003.

The MerR family of transcriptional regulators.

FEMS Microbiol Rev. 2003 Jun;27(2-3):145-63.

Brüggemann, H., Falinski, F., Deppenmeier, U., 2000.

Structure of the H_2 :quinone oxidoreductase of *Archaeoglobus fulgidus*. Identification and overproduction of the $F_{420}H_2$ -oxidizing subunit.

Eur. J. Biochem. 267: 5810-5814.

Bryant, M. P., 1972.

Commentary on the hungate technique for cultur of anaerobic bacteria.

Am. J. Clin. Nutr. 25: 1324-1328.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougerty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, B. P., Borodovsky, M., Klenk, H. P., Fraser, C. M., Smith, H. O., Woese, C. R., Venter, J. C., 1996.

Complete Genome Sequence of the Methanogenic Archaeon *Methanococcus jannaschii*.

Science 273: 1058-1073.

Burke, S.A., Krzycki, J.A., 1997.

Reconstitution of Monomethylamine:Coenzyme M methyl transfer with a corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*.

J. Biol. Chem. 1997 Jun 27;272(26):16570-7.

Burke, S.A., Krzycki, J.A., 1998.

Clustered genes encoding the methyltransferases of methanogenesis from monomethylamine.

J. Bacteriol. 1998 Jul;180(13):3432-40.

Cammack, R., 1997.

The enzyme at the end of the food chain.

Nature 390: 443-444.

Chester, N. and Marshak, D. R., 1993.

Dimethyl sulfoxid-mediated primer T_m reduction: a method for analyzing the role of renaturation temperature in the polymerase chain reaction.

Anal. Biochem. 209: 284-290.

Coleman, D., Dong, X., Hardin, J., Rocke, D.M. and Woodruff, D.L., 1999.

Some computational issues in cluster analysis with no a priori metric.

Comp. Stat. and Data Analysis, 31:1-11.

Deppenmeier, U., Blaut, M., Mahlmann, A., Gottschalk, G., 1990a.

Membrane-bound $F_{420}H_2$ -dependent heterodisulfide reductase in methanogenic bacterium strain Gö1 and *Methanobolus tindarius*.

FEBS Lett. 261: 199-203.

Deppenmeier, U., Blaut, M., Mahlmann, M., Gottschalk, G., 1990b.

Reduced coenzyme $F_{420}H_2$ -dependent heterodisulfide oxidoreductase: a proton translocating system in methanogenic bacteria.

Proc. Natl. Acad. Sci. USA 87: 9449-9453.

Deppenmeier, U., Blaut, M., Schmidt, B., Gottschalk, G., 1992.

Purification and properties of a F_{420} -nonreactive membrane-bound hydrogenase from *Methanosarcina* strain Gö1. Arch. Microbiol. 157: 505-511.

Deppenmeier, U., 1995.

Different structure and expression of the operons encoding the membrane-bound hydrogenases from *Methanosarcina mazei* Gö1.

Arch. Microbiol. 164: 370-376.

Deppenmeier, U., Müller, V., Gottschalk, G., 1996.

Pathways of energyconservation in methanogenic Archaea.

Arch. Microbiol. 165: 149-163.

Deppenmeier, U., 2002a.

The unique biochemistry of methanogenesis.

Prog. Nucleic. Acid Res. Mol. Biol. 71: 223-283.

Deppenmeier, U., 2002b.

Redox-driven proton translocation in methanogenic Archaea.

Cell. Mol. Life Sci. 59: 1-21.

Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A., Martinez-Arias, R., Henne, A., Wiezer, A., Bäumer, S., Jacobi, C., Brüggemann, H., Lienard, T., Christmann, A., Bomeke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H. P., Gunsalus, R. P., Fritz, H. J., Gottschalk, G., 2002.

The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea.

J. Mol. Microbiol. Biotechnol. 4: 453-61.

De Rosa, M., Gambacorta, A., 1988.

The Lipids of archaebacteria.

Prog. Lipid. Res. 27: 153-175.

DiMarco, A. A., Bobik, T. A. and Wolfe, R. S., 1990.

Unusual coenzymes of methanogenesis.

Annu. Rev. Biochem. 59:355-394.

DiRuggiero, J., Brown, J. R., Bogert, A. P., Robb, F. T., 1999.

DNA Repair Systems in Archaea: Mementos from the Last Universal Common Ancestor?

J. Mol. Evol. 49: 474-484.

Ellefson, W. L. , Whitman, W. B. and Wolfe, R. S., 1982.

Nickel-containing factor F430: chromophore of the methylreductase of *Methanobacterium*.

Proc. Natl. Acad. Sci. USA 79, 3707.

Ellermann, J., Hedderich, R., Böcher, R., Thauer, R. K., 1988.

The final step in methane formation.

Eur. J. Biochem. 172: 669-677.

Enssle, M., Zirngibl, C., Linder, D. and Thauer, R. K., 1991.

Coenzyme F₄₂₀ dependent N⁵,N¹⁰-methylenetetrahydromethanopterin dehydrogenase in methanol-grown *Methanosarcina barkeri*.

Arch. Microbiol. 155:483-490.

Ermler, U., Merckel, M., Thauer, R. K., Shima, S., 1997.

Formylmethanofuran: tetrahydromethanopterin formyltransferase from *Methanopyrus kandleri* - new insights into salt-dependence and thermostability.

Structure 5: 635-646.

Ermler, U., Grabarse, P., Shima, S., Goubeaud, M. and Thauer, R. K., 1997.

Formylmethanofuran: tetrahydromethanopterin formyltransferase from *Methanopyrus kandleri* - new insights into salt-dependence and thermostability.

Science 278, 1457.

Escalante-Semerena, J. C. , Leigh, J. C., Rinehart Jr., K. L., and Wolfe, R. S., 1984.

Tetrahydromethanopterin, a carbon carrier in methanogenesis.

J. Biol. Chem. 259, 9447-9455.

Everitt B.S., 1998.

Cluster Analysis, Arnold Verlag, London

Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R., 1996.

Advances in Knowledge Discovery and Data Mining.

AAAI Press, Menlo Park.

Fayyad, U.M., Piatetsky-Shapiro, G. and Smyth P., 1996.

From data mining to knowledge discovery in databases.

AI Magazine, 37-54.

Ferguson, DJ. Jr., Gorlatova, N., Grahame, D.A., Krzycki, J.A., 2000.

Reconstitution of dimethylamine:coenzyme M methyl transfer with a discrete corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*.
J. Biol. Chem. 2000 Sep. 15;275(37):29053-60.

Ferry, J. G., 1992.

Biochemistry of methanogenesis. Critic. Rev. Biochem.
Mol. Biol. 27: 473-503.

Ferry, J. G., 1995.

CO dehydrogenase.
Annu. Rev. Microbiol. 49: 305-333.

Ferry, J. G., 1997.

Enzymology of the fermentation of acetate to methane by *Methanosarcina thermophila*.
Biofactors 6: 25-35.

Ferry, J. G., 1999.

Enzymology of one-carbon metabolism in methanogenic pathways.
FEMS Microbiology Reviews, Volume 23, Issue 1, January 1999, Pages 13-38.

Finster, K., G.M. King and F. Bak. 1990.

Formation of methylmercaptan and dimethylsulfide from methoxylated aromatic compounds in anoxic marine and fresh water sediments.
FEMS Microbiol.: Ecol. 74: 295-302.

Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Manilofe, J., Dyer, T. A., Wolfe, R. S., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. N., Chen, K. N., Woese, D. R., 1980.

The phylogeny of prokaryotes.
Science 209: 457-463.

Fox, J.A., Livingston, D.J., Orme-Johnson, W.H., Walsh, C.T., 1987.

8-Hydroxy-5-deazaflavin-reducing hydrogenase from *Methanobacterium thermoautotrophicum*: 1. purification and characterization.

Biochemistry 26, 4219-4227 (1987).

Fox, J. D., He, Y. P., Shelver, D., Roberts, G. P. and Ludden P. W., 1996a.
Characterization of the region encoding the CO-induced hydrogenase of *Rhodospirillum rubrum*.

J. Bacteriol. 178: 6200-6208.

Fox, J. D., Kerby, R. L., Roberts, G. P. and Ludden, P.W., 1996b.

Characterization of the CO-induced, CO-tolerant hydrogenase from *Rhodospirillum rubrum* and the gene encoding the large subunit of the enzyme.

J. Bacteriol. 178: 1515-1521.

Friedmann, H. C., Klein, A., Thauer, R. K., 1990.

Structure and function of the Nickel porphinoid, coenzyme F₄₃₀, and of its enzyme, methyl-coenzyme M reductase.

FEMS Microbiol. Rev. 87: 339-348.

Garcia J. L., Bharat K. C. Ptel, Ollivier, B., 2000.

Taxonomic, phylogenetic, and ecological diversity of methanogenic Archaea.

Anaerobe 6: 205.

Grabarse, W., Vaupel, M., Vorholt, J. A., Shima, S., Thauer, R. K., Wittershagen, A., Bourenkov, G., Bartunik, H. D., Ermler, U., 1999.

The crystal structure of methenyltetrahydromethanopterin cyclohydrolase from the hyperthermophilic archaeon *Methanopyrus kandleri*.

Structure Fold. Des. 7: 1257-1268.110.

Gundavaram, S., 1996.

CGI programming on the world wide web.

von Shishir Gundavaram, 1996. O'Reilly Verlag.

Haase, P., Deppenmeier, U., Blaut, M., Gottschalk, G., 1992.

Purification and characterization of F₄₂₀H₂-dehydrogenase from *Methanolobus tindarius*.

Eur. J. Biochem. 203: 527-531.

Halpern, J. 1985.

Mechanisms of coenzyme B₁₂-dependent rearrangements.

Science 227:869-875.

Hanahan, D., 1983.

Studies on transformation of *Escherichia coli* with plasmids.

J. Mol. Biol. 166: 557-580.

Halboth, S., Klein, A., 1992.

Methanococcus voltae harbors four gene clusters potentially encoding two [NiFe] and two [NiFeSe] hydrogenases, each of the cofactor F₄₂₀-reducing or F₄₂₀-non-reducing types.

Mol. Gen. Genet. 233: 217-224.

Hansen, A., 2001

Bioinformatik. Ein Leitfaden für Naturwissenschaftler.

von Andrea Hansen, Birkhäuser Verlag, 2001.

Hartigan J., 1975.

Clustering Algorithms, John Wiley and Sons Verlag, New York,

Hartzell, P. L., Escalante-Semerena, J. C., Bobik, T. A., Wolfe, R. S., 1988.

A simplified methylcoenzyme M methylreductase assay with artificial electron donors and different preparations of component C from *Methanobacterium thermoautotrophicum* ΔH.

J. Bacteriol. 170: 2711-2715.

Hedderich, R., Klimmek, O., Kröger, A., Dirmeier, R., Keller, M. and Stetter, K. O., 1998.

Anaerobic respiration with elemental sulfur and with disulfides.

FEMS Microbiol. Rev. 22: 353-381.

Hedderich, R., Klimmek, O., Kröger, A., Dirmeier, R., Keller, M. & Stetter, K. O., 1999.

Anaerobic respiration with elemental sulfur and with disulfides.

FEMS Microbiol. Rev. 22: 353-381.

Hofmann, K., 2003

Charakterisierung und Kristallisierung der Elektronen-einspeisenden Module der F420H₂-Dehydrogenase.

Dissertation, Universität Göttingen.

Holmes, D.J., Caso, J.L. , and Thompson, C.J., 1993.

Autogenous transcriptional regulation of a thiostrepton-induced gene in *Streptomyces lividans*.

EMBO J. 12, 3183-3191.

Horowitz, J. P., Chua, J., Curby, R. J., Thomson, M. A., Da Rooge, M. A., Fischer, B. E., Mauricio, J., Klundt, I., 1964.

Substrates for cytochemical demonstration of enzyme activity. I. Some substituted 3-indoyl- β ,D-glycopyranosides.

J. Med. Chem. 7: 547-548.

Huber, H., Michael, J. H., Rachel, R., Fuchs, T., Wimmer, V. C., Stetter, K. O., 2002.

A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature 417: 63-67.

Hudec, M. and Steiner, P.M., 2002.

Skript zum Vortrag: Model-based classification of large data sets.

COMPSTAT conference on computational statistics, Berlin.

Hughes, P. E. and Tove, S. B., 1982.

Occurrence of α -Tocopherolquinone and α -Tocopherolquinol in microorganisms.

J. Bacteriol. 151: 1397-1402.

Hungate, R. E., 1969.

A roll tube method for cultivation of strict anaerobes.

In: Norris J.R. und Ribbons, D.W. (Ed.) Methods in Microbiology, Academic Press, New York, London. 3b: 117-132.

Ide, T., Bäumer, S., Deppenmeier, U., 1999.

Energy conservation by the H_2 :heterodisulfide oxidoreductase from *Methanosarcina mazei* Gö1: identification of two proton-translocation segments.

J. Bacteriol. 181: 4076-4080.

Inoue, H., Nojima, H. & Okayama, H., 1990.

High efficiency transformation of *Escherichia coli* with plasmids.

Gene 96: 23-28.

Jablonski, P. E., Ferry, J. G., 1991.

Purification and properties of methyl coenzyme M methylreductase from acetate-grown *Methanosarcina thermophila*.

J. Bacteriol. 173: 2481-2487.

Jarrel, K. F., Kalmokoff, M. L., 1988.

Nutritional requirement of the methanogenic archaeobacteria.

Can. J. Microbiol. 34: 557.

Jones, W. J., Nagle, D. P., Whitman, W., 1987.

Methanogens and the diversity of archaeobacteria.

Microbiol. Rev. 51: 135-177.

Kammler, M., Schon, C., Hantke, K., 1993.

Characterization of the ferrous iron uptake system of *Escherichia coli*.

J. Bacteriol. 1993 Oct;175(19):6212-9.

Kandler, O. and Hippe, H., 1977.

Lack of peptidoglycan in the cell wall of *Methanosarcina barkeri*.

Arch. Microbiol. 113: 57-60.

Kandler, O. and König, H., 1978.

Chemical composition of the peptidoglycan-free cell walls of methanogenic bacteria.

Arch. Microbiol. 118: 141-152.

Kates, M., Kushner, D. J., Matheson, A. T., 1993.

The biochemistry of Archaea (archaeobacteria).

Elsevier Biomedical press. Amsterdam. The Netherlands.

Keeling, P.J., Charlebois, R. L., Doolittle, W. F., 1994.

Archaeobacterial genomes: eubacterial form and eukaryotic content.

Curr. Opin. Genet. Dev. 4: 816-822.

Klaas, D. L., 1984.

Methane from anaerobic fermentation.

Science 223: 1021

Klenk, H.-P. and Doolittle, W. F., 1994.

Archaea and eukaryotes versus bacteria?

Curr. Biol. 4: 920-922.

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbusch, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenny, K., Adams, M. D., Loftus, B., Venter, J. C., 1997.

The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*.

Nature 390: 364-370.

Kobelt, A., Pfaltz, A., Ankel-Fuchs, D. and Thauer, R. K., 1987.

The L-form of *N*-7-mercaptoheptanoyl-*O*-phosphothreonine is the enantiomer active as component B in methyl-CoM reduction to methane.

FEBS Lett. 214, 265.

Kortlüke, C., Horstmann, K., Schwartz, E., Rohde, M., Binsack, R., Friedrich, B., 1992.

A gene complex coding for the membrane-bound hydrogenase of *Alcaligenes eutrophus* H16.

J. Bacteriol. 174: 6277-6289.

Künel, A., Vorholt, J. A., Thauer, R. K. and Hedderich, R., 1998.

An *Escherichia coli* hydrogenase-3-type hydrogenase in methanogenic archaea.

Eur. J. Biochem. 252: 467-476.

Kunow, J., Linder, D., Stetter, K. O., Thauer, R. K., 1994.

F₄₂₀H₂: quinone oxidoreduktase from *Archaeoglobus fulgidus*. Characterization of a membrane-bound multisubunit complex containing FAD and iron-sulfur clusters.

Eur. J. Biochem. 223: 503-511.

Lee, M.L., Kuo, F.C., Whitmore, G.A., Sklar, J., 2000.

Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations.

Proc. Natl. Acad. Sci. U S A. 2000 Aug 29;97(18):9834-9.

Lee, J-M., Zhang, S., Saha, S., Santa Anna, S., Jiang, C. and Perkins, J., 2001.

RNA Expression Analysis Using an Antisense *Bacillus subtilis* Genome Array.

Journal of Bacteriology 183(24): 7371-7380.

Leigh, J. A., Rinehart, K. L., Wolfe, R. S., 1985.

Methanofuran (carbon dioxide reduction factor), a formyl carrier in methane production from carbon dioxide in *Methanobacterium*.

Biochemistry 24: 995-999.

Lesk, A.L., 2002

Introduction to Bioinformatics von Arthur M. Lesk.

Oxford University Press, 2002.

Liang, M., Briggs, A.G., Rute, E., Greene, A.S., Cowley Jr., A.W., 2003

Quantitative assessment of the importance of dye switching and biological replication in cDNA microarray studies.

Physiol. Genomics. (10) [Epub ahead of print].

Livingston, D. A., Pfaltz, A., Schreiber, J., Eschenmoser, A., Ankel-Fuchs, D., Moll, J., Jaenchen, R., and Thauer, R. K., 1984.

Zur Kenntnis des Faktors F₄₃₀ aus methanogenen Bakterien: Über die Natur der Isolierungsartefakte von F₄₃₀, ein Beitrag zur Chemie von F₄₃₀ und zur konformationellen Stereochemie der Ligandperipherie von hydroporphinoiden Nickel(II)-Komplexen.

Helv. Chim. Acta, 67, 334.

Lund, P. A., S. J. Ford, and N. L. Brown, 1986.

Transcriptional regulation of the mercury-resistance genes of transposon Tn501.

J. Gen. Microbiol. **132**:465-480.

Lund, P.A. and Brown, N.L., 1989.

Up-promoter mutations in the positively-regulated mer promoter of Tn501.

Nucleic Acids Research, Vol 17, Issue 14 5517-5527.

Ma, K., and Thauer, R.K., 1990.

Single step purification of methylenetetrahydromethanopterin reductase from *Methanobacterium thermoautotrophicum* by specific binding to blue sepharose CL-6B.

FEBS Lett. 1990 Jul 30;268(1):59-62.

Maden, B. E., 2000.

Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C 1 metabolism.

Biochem. J. 350, 609.

Marck, C., 1988.

„DNA Strider“: a „C“ programm for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers.

Nucl. Acids Res. 16: 1829.

Marlovits, T.C., Haase, W., Herrmann, C., Aller, S.G., Unger, V.M., 2002.

The membrane protein FeoB contains an intramolecular G protein essential for Fe(II) uptake in bacteria.

Proc. Nat. Acad. Sci. USA 99: 16243-16248

Marmur, J., 1961.

A procedure for the isolation of deoxyribonucleic acid from microorganisms.

J. Mol. Biol. 3: 208-218.

McHugh, J.P., Rodriguez-Quinones, F., Abdul-Tehrani, H., Svistunenko, D.A., Poole, R.K., Cooper, C.E., Andrews S.C., 2003.

Global iron-dependent gene regulation in escherichia coli: A New Mechanism For Iron Homeostasis.

J. Biol. Chem. 2003 Aug 8;278(32):29478-86.

Meuer, J., Bartoschek, S., Koch, J., Künkel, A., and Hedderich, R., 1999.

Purification and catalytic properties of Ech hydrogenase from *Methanosarcina barkeri*.

Eur. J. Biochem. 265: 325-335.

Möller-Zinkhahn, D., Börner, G., Thauer, R. K., 1989.

Function of methanofuran, tetrahydromethanopterin and coenzyme F₄₂₀ in *Archaeoglobus fulgidus*.

Arch. Microbiol. 152: 362-368

Müller, V., Blaut, M., Gottschalk, G., 1988.

The transmembrane electrochemical gradient of Na⁺ as driving force for the methanol oxidation in *Methanosarcina barkeri*.

Eur. J. Biochem. 172: 601-606.

Müller, V. and Gottschalk G., 1992.

„Acetogenesis“ (H. L. Drake, ed.), p. 127, Chapman & Hall Verlag, New York, USA (1992).

Noll, K. M. , Rinehart, R. K. , Tanner, R. S. and Wolfe, R. S. 1986.

Structure of component B (7-mercaptoheptanoylthreonine phosphate) of the methylcoenzyme M methylreductase system of *Methanobacterium thermoautotrophicum*.

Proc. Natl. Acad. Sci. USA 83(12):4238-42.

Noll K. M., Donnelly M. I., Wolfe R. S., 1987.

Synthesis of 7-mercaptoheptanoylthreonine phosphate and its activity in the methylcoenzyme M methylreductase system.

J. Biol. Chem. 262: 513-515.

Nunoshiba, T., E. Hidalgo, C. Amabile Cuevas, B. Demple 1992.

Two-stage control of an oxidative stress regulon: the Escherichia coli SoxR protein triggers redox-inducible expression of the soxS regulatory gene.

J. Bacteriol. 174:6054-6060.

Park, S.J and Gunsalus, R.P., 1995.

Oxygen, iron, carbon, and superoxide control of the fumarase fumA and fumC genes of Escherichia coli: role of the arcA, fnr, and soxR gene products.

J Bacteriol. 1995 Nov;177(21):6255-62.

Paul, L., Ferguson, D.J. Jr., Krzycki, J.A., 2000.

The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri* contain in-frame and read-through amber codons.

J. Bacteriol. 182(9):2520-9.

Pfaltz, A., Jaun, B., Fassler, A., Eschenmoser, A., Jaenchen, R., Gilles, R., Diekert, G., 1982.

Zur Kenntnis des Faktors F₄₃₀ aus methanogenen Bakterien: Struktur des porphinoïden Ligandsystems Thauer.

Helv. Chim. Acta, 65, 828.

Pfennig, G. N., Lippert, K. D., 1966.

Über das Vitamin B₁₂-Bedürfnis phototropher Schwefelbakterien.

Arch. Microbiol. 55: 245-256.

Quackenbush, J., 2002, Microarray data normalization and transformation.

Nat. Genet. 2002 Dec;32 Suppl:496-501.

Sargent, T.D., Dawid, I.B., 1983.

Differential gene expression in the gastrula of *Xenopus laevis*.

Science 222, 135-139

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995.

Quantitative monitoring of gene expression patterns with a complementary DNA microarray.

Science 270, 467-470

Schena, M., Heller, R.A., Theriault, T.P., Konrad, K., Lachenmeier, E., Davis, R.W., 1998.

Microarrays: biotechnology's discovery platform for functional genomics.

Trends Biotechnol. 16, 301- 306

Sambrook, J., Fritsch, E. F. & Maniatis, T., 1989.

Molecular cloning.

A laboratory manual. 2. Auflage, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA.

Sandman, K., Krzycki, J. A., Dobrinski, B., Lurz, R., Reeve, J. N., 1990.

Hmf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermobacter fervidus*, is most closely related to histones.

Proc. Natl. Acad. Sci. USA 87: 5788-5791

Sapolsky, S., Weyler, W., Maile, R.R., Causey, S.C., and Ferrari, E., 2001.

Correlation between *Bacillus subtilis scoC* Phenotype and Gene Expression Determined Using Microarrays for Transcriptome Analysis.

Journal of Bacteriology. 183(24): 7329-7340,.

Sauer, K., Harms, U., Thauer, R.K., 1997.

Methanol:coenzyme M methyltransferase from *Methanosarcina barkeri* - Purification, properties and encoding genes of the corrinoid protein MT1
Eur. J. Biochem 243: 670-677.

Sauer, K., Thauer, R. K., 1997.

Methanol:coenzyme M methyltransferase from *Methanosarcina barkeri*. Zinc dependence and thermodynamics of the methanol:cob(I)alamin methyltransferase reaction.

Eur. J. Biochem. 249: 280-285.

Sauer, K., Thauer R.K., 1998.

Methanol : coenzyme M methyltransferase from *Methanosarcina barkeri* - Identification of the active-site histidine in the corrinoid-harboring subunit MtaC by site-directed mutagenesis.

Eur. J. Biochem 253: 698-705.

Sauer, K., Thauer R.K., 1999.

Methanol : coenzyme M methyltransferase from *Methanosarcina barkeri* - substitution of the corrinoid harbouring subunit MtaC by free cob(I)alamin.
Eur. J. Biochem 261: 674-681.

Sauer, K., Thauer, R.K., 2000.

Methyl-coenzyme M formation in methanogenic archaea - Involvement of zinc in coenzyme M activation.
Eur. J. Biochem. 267: 2498-2504.

Schmid, G., Böck, A., 1984.

Immunoblotting analysis of ribosomal proteins from archaebacteria. System.
Appl. Microbiol. 5: 1-10.

Setzke, E., Hedderich, R., Heiden, S., Thauer, R. K., 1994.

H₂:heterodisulfide oxidoreductase complex from *Methanobacterium thermoautotrophicum*.
Eur. J. Biochem. 220: 139-148.

Shima, S., Warkentin, E., Grabarse, W., Sordel, M., Wicke, M., Thauer, R. K., Ermler, U., 2000.

Structure of coenzyme F₄₂₀ dependent methylenetetrahydromethanopterin reductase from two methanogenic archaea.
J. Mol. Biol. 300: 935-950.

Skeidsvoll, J., Ueland, P.M.,1995.

Analysis of double-stranded DNA by capillary electrophoresis with laser-induced fluorescence detection using the monomeric dye SYBR green I.
Anal. Biochem; 231(2):359-65.

Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J. N., 1997.

Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics.

J. Bacteriol. 179: 7135-7155.

Steigerwald, V. J., Pihl, T. D., Reeve, J. N., 1992.

Identification and isolation of the polyferredoxin from *Methanobacterium thermoautotrophicum* strain ΔH.

Proc. Natl. Acad. Sci. USA. 89: 6929-6933.

St. John, T.P., Davis, R.W., 1979.

Isolation of galactose-inducible DNA sequences from *Saccharomyces cerevisiae* by differential plaque filter hybridization.

Cell 16, 443-452.

Stojiljkovic, I., Baumler, A.J., Hantke, K., 1994.

Fur regulon in gram-negative bacteria. Identification and characterization of new iron-regulated *Escherichia coli* genes by a fur titration assay.

J. Mol. Biol. 236(2):531-45.

Summers AO., 1992.

Untwist and shout: a heavy metal-responsive transcriptional regulator.

J. Bacteriol. 174(10):3097-101.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S., and Golub, T. R., 1999.

Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation.

Proc. Natl. Acad. Sci. USA, 96: 2907-2912.

Taylor, C. D. and Wolfe, R. S., 1974.

Structure and Methylation of Coenzyme M (HSCH₂CH₂SO₃).

J. Biol. Chem. 249, 4886.

te Brommelstroet, B. W., Hensgens, C. M., Keltjens, J. T., van der Drift, C., Vogels, G. D., 1990.

Purification and properties of 5,10-methylenetetrahydromethanopterin reductase, a coenzyme F₄₂₀-dependent enzyme, from *Methanobacterium thermoautotrophicum* strain delta H.

J. Biol. Chem. 265: 1852-1857.

te Brommelstroet, B. W., Geerts, W. J., Keltjens, J. T., van der Drift, C., Vogels, G. D., 1991.

Purification and properties of 5,10-methylenetetrahydromethanopterin dehydrogenase and 5,10-methylenetetrahydromethanopterin reductase, two coenzyme F₄₂₀-dependent enzymes, from *Methanosarcina barkeri*.

Biochim. Biophys. Acta 1079: 293-302.

Terlesky K. C. and Ferry J. G., 1988.

Ferredoxin requirement for electron transport from the carbon monoxide dehydrogenase complex to a membrane-bound hydrogenase in acetate-grown *Methanosarcina thermophila*.

J. Biol. Chem. 263: 4075-4079.

Thomas, M. G., and Escalante-Semerena, J. C., 2000.

Identification of an alternative nucleoside triphosphate: 5'-deoxyadenosylcobinamide phosphate nucleotidyltransferase in *Methanobacterium thermoautotrophicum* ΔH.

J. Bacteriol. 182:4227-4233.

Touati D. 1988.

Transcriptional and posttranscriptional regulation of manganese superoxide dismutase biosynthesis in *Escherichia coli*, studied with operon and protein fusions.

J. Bacteriol. 170:2511-20.

Tornabene, T. G. and Langworthy, T. A., 1978.

Diphytanyl and bidiphytanyl ether lipids of methanogenic archaeobacteria.

Science 203: 51-53.

Törönen, P., Kolehmainen, M., Wong, G. and Castrén, E., 1999.

Analysis of gene expression data using self-organizing maps.

FEBS Lett. 451(2): 142-146.

Tran, P.H., Peiffer, D.A., Shin, Y., Meek, L.M., Brody, J.P., Cho, K.W., 2002.

Microarray optimizations: increasing spot accuracy and automated identification of true microarray signals.

Nucleic Acids Res. 30(12):e54.

Van Beelen, P., Stassen, A. P. M., Bosch, J. W. G., Vogels, G. D., Guijt, W., Haasnoot, C. A. G., 1984.

Elucidation of the structure of methanopterin, a coenzyme from *Methanobacterium thermoautotrophicum*, using two-dimensional nuclear-magnetic-resonance techniques.

Eur. J. Biochem. 134: 563-571.

Van der Maarel, M. J. E. C., Jansen, M. and Hansen, T. A., 1995.

Methanogenic conversion of 3-S-methylmercaptopropionate to 3-mercapto-propionate.

Appl. Environ. Microbiol. 61, 48.

Vassinova, N., Kozyrev, D., 2000.

A method for direct cloning of fur-regulated genes: identification of seven new fur-regulated loci in *Escherichia coli*.

Microbiology. 146(12):3171-82.

Vaupel, M., Dietz, H., Linder, D., Thauer, R. K., 1996.

Primary structure of cyclohydrolase (Mch) from *Methanobacterium thermoautotrophicum* (strain Marburg) and functional expression of the mch gene in *Escherichia coli*.

Eur. J. Biochem. 236: 294-300.

Vorholt, J. A. and Thauer, R. K., 1997.

The active species of 'CO₂' utilized by formylmethanofuran dehydrogenase from methanogenic Archaea.

Eur. J. Biochem. 248: 919-924.

Wall, L., Christiansen, T., Orwant, J., Schwartz, R., 1992.

Programming PERL: There's more than one way to do it.

Larry Wall, Tom Christiansen, Jon Orwant, Randal Schwartz

O'Reilly Verlag, 1992,

Walsh, C., 1986.

Naturally occurring 5-deazaflavin coenzymes: Biological redox roles.

Acc. Chem. Res. 19: 216-221.

Weiss, D. und Thauer, R. K., 1993.

Methanogenesis and the unity of biochemistry.

Cell 72: 819-822.

Westenberg D. J., Braune A., Ruppert C., Muller V., Herzberg C., Gottschalk G., Blaut M., 1999.

The F₄₂₀H₂-dehydrogenase from *Methanobolus tindarius*: cloning of the *ffd* operon and expression of the genes in *Escherichia coli*.

FEMS Microbiol. Lett. 170: 389-398.

Wettach, J., Gohl, H. P., Tschochner, H., Thomm, M., 1995.

Functional interaction of yeast and human TATA binding proteins with an archaeal RNA-polymerase and promotor.

Proc. Natl. Acad. Sci. USA 92: 472-476.

Whitman, W. B., Ankwanda, E., Boone, D. R., 1991.

The methanogenic bacteria.

In: The Prokaryotes, Balows, A., Truper, H. G., Dworkin, M., Harder, W., Schleifer, K. J. (eds.), pp. 719-767, Springer Verlag, New York.

Wittwer, C.T., Herrmann, M.G., Moss, A.A., Rasmussen, R.P., 1997.

Continuous fluorescence monitoring of rapid cycle DNA amplification.

Biotechniques. Jan;22(1):130-1, 134-8.

Woodruff, D.L. and Rocke, D.M., 1994.

Computable robust estimation of multivariate location and shape in high dimension using compound estimators.

JASA, 89:888-896.

Wolin, E. A., Wolfe, R. R., Wolin, M. J ., 1964.

Viologen dye inhibition of methane formation by *Methanobacterium omelanskii*.

J. Bacteriol. 87: 993-998.

Woese, C. R., Kandler, O. W., Wheelis, M. L., 1990.

Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.

Proc. Natl. Acad. Sci. USA 87: 4576-4579.

Wolfe, R. S., 1985.

Unusual coenzymes of methanogenesis.

TIBS. 10: 396-399.

Zillig, W., Stetter, K. O., Schnabel, R., Madou, J., Gierl, A., 1982.

Transkription in Archaeobacteria.

Zbl. Bakt. Hyg. 1 Abt. Orig. C3: 218-227.

Anhang A:

Quellcode und Dokumentation der selbst erstellten Computerprogramme

Im Rahmen dieser Arbeit wurden diverse Computerprogramme erstellt. Als Sprache für diese Skripte wurde PERL gewählt. Für die Implementation einer PERL 5.0 Umgebung wurde eine freien Distribution der Firma ActiveState verwendet (2.9.2.1). Die Reihenfolge der Programme in diesem Anhang entspricht der Reihenfolge in welcher auf sie in dieser Arbeit verwiesen wurde. Die Kommentare innerhalb der Programme wurden in englischer Sprache verfaßt, da sie nach Veröffentlichung dieser Arbeit einer PERL ‚open-source‘ Distribution unterliegen werden. Der Quellcode dieser Programme ist dadurch für wissenschaftliche und nicht kommerzielle Zwecke frei zugänglich und verwendbar, wenn dabei auf diese Arbeit verwiesen wurde.

Da optimierte PERL Programme schwer für einen Menschen zu lesen sind, unterscheidet sich der Quellcode dieser Programme zum Teil von dem Quellcode welcher während des erstellen dieser Arbeit verwendet wurde. Oft wurden komplexe Abläufe auf Kosten von Geschwindigkeit umgeschrieben, um den Aufbau der Programme leserlicher zu machen. Diese Stellen sind dokumentiert. Die Funktion der Programme ist davon nicht beeinträchtigt. Obwohl aus Sicht eines Informatikers ein Unterschied zwischen einem Programm und einem Skript besteht, wurden diese Begriffe in dieser Arbeit als Synonyme verwendet.

Alle Programme starten mit `#!/usr/bin/perl`.

Dies ist ein sogenannter ‚Shebang‘ und wurde vor dem eigentlichen Start des Skriptes interpretiert. Unter Windows Betriebssystemen kommt dem keine Bedeutung zu. Unter Linux, weist es dem Betriebssystem den Pfad zu PERL welches benutzt werden soll, um das Skript auszuführen. Sollte dieser Pfad nicht `/usr/bin/perl` sein, läßt er sich über den ‚which perl‘ Befehl lokalisieren. Alle Programme benutzten ‚use strict‘ und ‚use warnings‘, um mögliche Fehler leichter zu finden bzw. zu vermeiden. Die absoluten Filenamen und Pfade wurden durch Buchstabenreihen (z.B. bbbbb) ersetzt.

Skript-1 (verwendet in 3.1.1):

```
#!/usr/bin/perl

#####
#
# Script for logging into the local ERGO Software to get all ORF sequences
# of an organism.
# The Script simulates a Web-Browser to log in.
#
# the targetfile contains html code and has to be parsed for the pure
# sequences with another Script to produce a multiFASTA file
#
# The Script asks for username and password at start.
#
# ERGO uses GET Form parsing in its CGI, this security hole enables this script to
# work.
# If at some time the Forms get converted to POST, it will no longer function.
#####

use strict;
use warnings;

use LWP::UserAgent;
use HTTP::Request;
use HTTP::Response;
use URI::Heuristic;

#define tragetfile
my $targetfile="C:/xxx/yyy/zzz.txt";

#declare variables
my $USAGE = "$0 One or more needed parameters are not given!"
my $username="";
my $password="";
my $suffix="";
my $highest_orf=0;

##### Get Data from the User #####
#
# Get the required Data from the user, remove the linefeeds and assign to variables
#
print "Please enter the 3 letter ERGO suffix of the organism.\n Example: GOX for
G.oxydans: ";
$suffix = <STDIN>;
chomp $suffix;
```

```

print "Please enter number of highest ORF in organism to look up: ";
$highest_orf = <STDIN>;
chomp $highest_orf;

print "Please enter ERGO username: ";
$username = <STDIN>;
chomp $username;

print "Please enter ERGO password: ";
$password = <STDIN>;
chomp $password;

if ( (!$username) || (!$password) || (!$highest_orf) || (!$suffix) )
{
print $USAGE;
exit;
}
#
#
##### end of user input #####

#####
# use a for loop from 1 to number of highest orfs.
# the loop requests the sequence for an orf with the number.
# if no orf with the number exists an the html document contains no sequence
# but ERGO returns no error

for (my $k=1;$k<$highest_orf;$k++)
{
my $i = sprintf "%5.5d", $k; # leading 0 to 5 decimals
my $roher_url = "http://gl-ergo.migen.bio.uni-
goettingen.de:8080/ERGO/CGI/prot.cgi?user=$username:$password;org=$su
ffix;prot=R$suffix$i;request=dna_sequence"; # For Protein Sequence
#extraktion change request=dna_sequence
#to request=aa_sequence
my $url = URI::Heuristic::uf_urlstr($roher_url); #get the URL
$| = 1; #autoflush output buffer now and after each print statement
printf "%s =>\n\t", $url; #just for checking on screen
my $ua = LWP::UserAgent->new(); # generate blank useragent
$ua->agent("Mozilla 9.1 Platinum"); # set browser of useragent object
my $anf = HTTP::Request->new(GET => $url); # assign http GET request to
#anf variable
my $antwort = $ua->request($anf); #get document at URL, content access via
#content()

if ($antwort ->is_error())
{
printf " %s\n", $antwort->status_line
}
else

```

```

    {
        my $inhalt = $antwort->content(); # get content of antwort object
        open TARGET, ">>$targetfile" || die "Cannot open footemp file for read
        :$!";
        printf TARGET "%s\n",$inhalt || die "Cannot open footemp file for write
        :$!";
        close (TARGET);
    }
}
##### End of file #####

```

Verwendete Variablen:

Alle lokal verwendeten Variablen wurden als Typ 'my' deklariert.

Filehandles:

\$targetfile - Enthielt den Pfad zum File in welches die Ergebnisse geschrieben wurden.

Andere Skalare Variablen

\$USAGE - Variable mit Fehlermeldung bei falscher Eingabe. Gibt \$0 mit zurück.

\$username - Enthielt den vom Benutzer eingegebenen ERGO Benutzernamen.

\$password - Enthielt das vom Benutzer eingegebenen ERGO Paßwort.

\$suffix - Enthielt den vom Benutzer eingegebenen ERGO-Suffix des Organismus.

\$highest_orf - Enthielt die vom Benutzer eingegebenen Nummer des höchsten ORFs.

\$k - Counter in der for()-Schleife.

\$i - Auf 5 Dezimalstellen umformatiertes \$k, für die ERGO Abfrage.

\$roher_url - Enthielt die rohe URL zum CGI des ERGO-Servers mit den User Parametern.

\$url- Enthielt \$roher_url nach parsing durch URI::Heuristic::uf_urlstr().

(= hexadezimale und ASCII Transformation)

\$ua - UserAgent Objekt, generiert durch LWP::UserAgent->new();.

\$anf - Enthielt den Code zu Anfrage über den http Befehl GET.

HTTP::Request->new(GET => \$url);

\$antwort - Enthielt das html-Dokument, auf den Inhalt wurde über content() zugegriffen.

\$inhalt – Inhalt des html-Dokumentes von \$antwort.

Funktionsweise des Programms:

Zuerst wurden die benötigten PERL-Module über ‚use‘ geladen. Die verwendeten Module sind in der normalen PERL Distribution enthalten.. Als nächstes wurden die Filehandle Variable und die anderen globalen Variablen deklariert. Dann forderte das Skript die benötigten Parameter wie Benutzername und Kennwort vom User an. Es wurde eine for()-Schleife geöffnet die von 1 an alle ORFs in ERGO abfragte, bis zu dem vom User angegebenen höchsten ORF. Hierbei wurden die vom User eingegebenen Parameter in eine rohe URL eingesetzt und in eine heuristische URL überführt. Dann wurde über LWP ein neuer UserAgent generiert und ihm ein String zur Identifikation zugewiesen. Der Variable \$antwort wurde über den request() Befehl des LWP Moduls eine html-Dokument zugewiesen. War der Rückgabewert von request() des LWP Moduls keine Fehlermeldung, wurde über den content() Befehl der Inhalt des html-Dokumentes von \$antwort der Variablen \$inhalt zugewiesen. Dieser Inhalt wurde dann in das Zielfile geschrieben.

Dieses Skript benutzte ferner eine PERL ‚autoflush‘-Routine über \$|=1;. Dadurch wurde der Schreibpuffer nach jedem print() Aufruf geleert um die Geschwindigkeit des Skriptes zu erhöhen. Der Lesebuffer wurde hiervon nicht beeinträchtigt. Zur Erstellung dieses Skriptes wurde das PERL-Kochbuch des O'Reilly Verlages verwendet, um die zu benutzenden Module zu finden. Dieses Skript extrahierte in der dokumentierten Version nur DNA Sequenzen aus ERGO. Zur Extraktion von Proteinsequenzen aus ERGO mußte im Skript lediglich request=dna_sequence abgeändert werden zu request=aa_sequence .

Skript-2 (verwendet in 3.1.1):

```
#!/usr/bin/perl

use strict;
use warnings;

# Script for Extracting DNA Sequence vom html files generated by Skript-1.
# The generated file is in multi-FASTA format.

# Define Filehandles
my $sourcefile="C:/xxxxxx/zzzzz.txt";
my $targetfile="C:/yyyyyy/vvvvv.txt";

open INPUT,"$sourcefile" || die "Error opening file for append. $!\n";

while (<INPUT>)
{
  if(<>RGOX/){$flag = 1};
  if(<>RGOX/ && /_/){$flag = 0};
  if(<>PRE/){$flag = 0};      #bei aa convert PRE => pre
  if(<>Up/){$flag = 0};
  if(<>Dn/){$flag = 0};
  if ($flag)
  {
    open TARGET, ">>$targetfile" || die "Cannot open footemp file for read :$!";
    print TARGET "$_" || die "Cannot open footemp file for write :$!";
    close (TARGET);
  }
  close (INPUT);
}
##### End of file #####
```

Verwendete Variablen:

Alle lokal verwendeten Variablen wurden als Typ 'my' deklariert.

Filehandles:

\$sourcefile - Enthielt den Pfad zum html-File aus Skript-1.

\$targetfile - Enthielt den Pfad zum Zielfile in welches die Ergebnisse geschrieben werden sollten.

Andere Skalare Variablen:

\$flag – Wurde auf 1 gesetzt wenn eine Sequenz oder Annotation innerhalb eines html files begann, und auf 0 gesetzt wenn die Sequenz oder Annotation endete.

Funktionsweise des Programms:

Dieses PERL Programm las ein html-File, welches vorher von Skript-1 generiert wurde, über eine while()-Schleife zeilenweise ein. Hierbei wurde über reguläre Ausdrücke überprüft, ob sich in der jeweiligen Zeile eine gewünschte Information (= Annotation oder Sequenz) befand. War dies der Fall wurde die Flag Variable auf 1 gesetzt und die Zeile in das Zielfile geschrieben. War dies nicht der Fall wurde die Flag Variable auf 0 gesetzt. Dieses Skript konnte nur html-files konvertieren, die über Skript-1 und einer DNA-Sequenz-Abfrage erstellt wurden. Um Files zu konvertieren, die durch ein Protein-Sequenz Abfrage generiert wurden, mußte innerhalb des Skriptes

```
if(!/PRE/){$flag = 0};  
zu  
if(!/pre/){$flag = 0};
```

geändert werden.

Skript-3 (verwendet in 3.1.2):

```

#!/usr/bin/perl

use strict;
use warnings;

# Selective BLAST script with parsing for annotated genes
# Generates a tab-delimited .txt file
# Requires a multiFASTA file with the query Protein sequences (ORFs)
# This Script can be used with DNA sequences with only a few changes.

#Set Filehandles
#multi-FASTA file with query sequences
$proteinfile="C:/aaaaaa/bbbbbbb/xxxxxxxxxxxx.txt";

#temp file for blast query sequences
$temp="c:/cccc/yyyyyyy.txt";

# define file to write results to
$targetfile="C:/eeeeee/ffffff/gggggg/zzzzzz.txt";

#only needed if raw blast output dumping is needed
# $tempblastresult="c:/bbbbbb/gggggg.txt";

#$i=0;          #counter for keeping track of progress, comment out if not needed

$tempkey=""; #temporary variable for %myhash key generation

#####
# Open the file containing the query sequences.
# The sequences gets loaded into an associative array (=hash) in line mode.
# The terminal \n gets removed and becomes merged with the sequence in the next
# line.
# The resulting hash contains key,value pairs of each ORF with Key=name and
# value=sequence.

open INPUT, $proteinfile || die "Cannot open input file for read :$!";
while (<INPUT>)
{
  if (/>/)
  {
    chomp($tempkey=$_); #savely remove \n from EOL
  }
  else { $myhash{$tempkey}="$myhash{$tempkey}$_"; #concatenate string in
hash value
}
}

```

```

close (INPUT);
#####

#remove the temporary variable
undef $tempkey;

#####
# Debug routine for dumping the hash data on the screen
#
# while ( ($key,$value)=each %myhash)
#   {
#     print "$key\n\n$value\n\n\n"
#   }
#####

#####
# Iterate over the hash (%myhash) and BLAST each sequence.
# A tempfile tempfile ($temp) gets generated with raw sequence data as BLAST
# query.
# BLAST for some reason refuses to use variables as query input, wants a file path.

#####      Start outer while Loop      #####

while ( ($key,$value)=each %myhash)
{
  my $flag=0;      #end loop flag
  my $legalhit=0;  #filter flag for annotated hits
  my $blaster;     #variable for holding blast output
  my @splitter;   #array variable for holding blast output lines. 1 line per field
  my $k=0;        #array pointer for the @splitter array of blast results

#For keeping track of process. commented out normally
#  $i++;print "\$i ist $i\n";

  open TEMP, ">$temp" || die "Cannot open temp file for read :$!";
  print TEMP "$value"|| die "Cannot open temp file for write :$!";
  close (TEMP);

#####
#
# Two Versions of the BLAST call, uncomment the needed version
#
# BLAST with tempfile as query and only 1 result.
# This routine is faster if only one hit is needed:
# my $blaster = `blastall -p blastp -d ncbi_nr -i $temp -K 1 `
# ||
# die "Error in blast call :$!";
#

```

```

# Standard BLAST with tempfile as query:
$blaster = `blastall -p blastp -d ncbi_nr -i $temp `
||
die "Error in blast call :$!";
#
#####

# print raw blast result to a file for bughunts. uncomment if needed
# open TEMP2, ">>$tempblastresult"
# || die "Cannot open footemp file for read :$!";
# print TEMP2 "$blaster" || die "Cannot open temp2 file for write :$!";
# close (TEMP2);

# split blast result at \n+ for later iteration and load the lines
# into fields of the @splitter array
@splitter = split /\n+/, $blaster;

#free chunky bit of memory
undef $blaster;

# walk through (=Iterate) @splitter with $k as array pointer to
# parse the BLAST-result and write annotated matches to the target file
# the $legalhit variable is a flag getting set when an annotated match was found
# the $flag variable is a flag getting set when the annotated hit was parsed
# the loop ends if no hit is found or if $flag=1
#
# Running scripts mostly used splice(@splitter, 1, 1); instead of the pointer
# increment
# for speed purposes.
# This was changed here for easier comprehension

##### start of inner while Loop #####

while ($splitter[$k]) # beware possible breakpoints. it is a nested loop.
{
  if (!
    (
      $splitter[$k]=~/putative/i
      || $splitter[$k]=~/hypothet/i
      || $splitter[$k]=~/unknown/i
    )
  )
  {
    if ($splitter[$k]=~/^\>/)
    {
      open TARGET, ">>$targetfile" || die "Cannot open target file for read :$!";

```

```

    print TARGET "$key\t" || die "Cannot open target file for write :$!";
    print TARGET "$splitter[$k]$splitter[$k+1]\t"
    || die "Cannot open target file for write :$!";
    close (TARGET);
    $legalhit=1; #found an annotated match
}

if ( ($splitter[$k]=~/xpect/) && $legalhit==1 ) #get score and e-value
{
    open TARGET, ">>$targetfile" || die "Cannot open target file for read :$!";
    print TARGET "$splitter[$k]\n" || die "Cannot open target file for write :$!";
    close (TARGET);
    $flag=1; #set flag to end loop. use break or last instead, if speed matters
}
}
#increment $k to move to next line (=array field) in the blast result
$k++;

last if ($flag==1); # leave inner nested loop only .
                    # otherwise use 'last TOHERE if($flag==1)' instead and put
                    # 'TOHERE: ' in front of the loop to break out of.
}
#####      End of inner while Loop      #####

#reset variables for next iteration step
$flag=0;
$k=0;
$legalhit=0;
undef @splitter; #free another chunk of memory
}
#####      End of outer while Loop      #####

#
#
#####      End of file      #####

```

Verwendete Variablen:

Alle lokal verwendeten Variablen wurden als Typ 'my' deklariert.

Filehandles:

\$proteinfile - Enthielt den Pfad zum multiFASTA-file der queries.

\$stemp - Enthielt den Pfad zu einer temporären Auslagerungsdatei.

\$targetfile - Enthielt den Pfad zum File in welches die Ergebnisse geschrieben wurden.

Andere Skalare Variablen

\$tempkey - Temporäre Variable zur Erzeugung von %myhash.

\$k - Array-Pointer zum Parsen der BLAST Ergebnisse.

\$flag - Flag Variable. Wurde gesetzt um die innere while()-Schleife zu verlassen.

\$legalhit - Flag Variable. Wurde gesetzt wenn ein annotiertes Ergebnis gefunden wurde.

\$key – Zuordnungsvariable (=Schlüssel) für Werte in %myhash.

\$value – Werte in %myhash welche einem Schlüssel zugeordnet wurden (\$key).

\$blaster – Enthielt die Rohdaten des BLAST Ergebnisses.

Arrays:

@splitter - Enthielt in jedem Feld eine Zeile der BLAST-Rohdaten.

Hashes:

%myhash - Enthielt alle Namen (= \$key) und Sequenzen (= \$value) des multiFASTA-Files als Paare .

In diesem Programm normalerweise auskommentierte Variablen:

\$tempblastresult - Enthielt den Pfad zum File in welches die Rohdaten der BLAST Ergebnisse geschrieben wurden.

\$i - Zähler für die Anzahl der prozessierten Gene.

In der Version dieses Skriptes welche für die Aufgaben in dieser Arbeit benutzt wurde, waren folgende Abläufe anders strukturiert:

\$i und die dazugehörige Zählerfunktion waren aktiviert, um den Fortschritt des Programmes zu betrachten.

\$k wurde nicht als array-pointer verwendet, sondern durch 0 ersetzt.

Anstelle der Inkremetation (\$k++) wurde das erste Feld des @splitter arrays gelöscht durch: splice(@splitter, 1, 1);

Funktionsweise des Programms:

Zuerst wurden die Pfade zu den Files, welche von dem Script benutzt wurden sollen, Variablen zu gewiesen (Filehandles).

Das multiFASTA-file mit den Query-Sequenzen wurde geöffnet und zeilenweise eingelesen durch:

```
open INPUT, $proteinfile || die "Cannot open input file for read :$!";  
while (<INPUT>) {.....}
```

Jede einzelne Zeile wurde innerhalb der Schleife überprüft, ob sie mit einem ,>' beginnt. War dies der Fall, wurde dieser Zeile durch `chomp()` der terminale Zeilenumbruch (`newline, \n`) entfernt und der Wert der Zeile wurde in die Variable `$tempkey` geladen. Ansonsten wurde der Zeilenwert als ,value' innerhalb des `%myhash` hashes gesetzt. Hierbei diente `$tempkey` als ,key' für den ,value'. Der Wert wurde an den bereits bestehenden ,value' angehängt. Hierbei ist zu beachten, daß `$tempkey` außerhalb der Schleife deklariert sein mußte. Nachdem die Variable `$tempkey` ihren Zweck erfüllt hatte, wurde ihre Speicherzuweisung durch `undef()` gelöscht.

Nun wurde eine `while()`-Schleife geöffnet, welche alle `key/value` Paare im `%myhash` hash durchlief. Hierbei wurde jeder ,key' der Variablen `$key` und jeder ,value' der Variablen `$value` zugewiesen. Innerhalb dieser Schleife wurden zuerst all lokalen Variablen deklariert und dann der Wert von `$value` in ein temporäres File geschrieben (filehandle ist `$temp`). Dieses File war notwendig, da BLAST als Eingabe keine Variablen als Query akzeptierte.

Anschließend wurde über einen sogenannten ,backticks'-call ein externer Funktionsaufruf des `blastall` Programms ausgeführt. Als Input diente hierbei das temporäre File. Der Rückgabewert von ,blastall' wurde im Speicher in der Variablen `$blaster` gespeichert. Der Wert von `$_` wurde zeilenweise in den `@splitter` array geladen durch :

```
@splitter = split /\n+/, $blaster;
```

Aufeinander folgende Leerzeilen ($\backslash n+$) wurden hierbei gelöscht, da sonst PERL nicht im Stande war, diesen @splitter zu iterieren und vorher abbrach. Durch undef() wurde die Speicherzuweisung der \$blaster Variablen gelöscht. Nun wurde eine zweite while() Schleife geöffnet, die alle Elemente des @splitter Arrays als Skalar (=\$splitter) durchlief. \$k diente hierbei als Array-Pointer und wurde am Ende der Schleife inkrementiert. Der Wert jedes Elements des @splitter arrays wurde hierbei überprüft, ob es die substrings ‚putative‘, ‚hypothet‘ oder ‚unknown‘ nicht enthielt. Dieser Abgleich wurde ‚case-insensitiv‘ durchgeführt. War dies der Fall, wurde in der ersten if()-Abfrage überprüft, ob dieses Feld mit einem ‚>‘ begann (Annotation) und gegebenenfalls diese Annotation in das Zielfile geschrieben und der Flag \$legalhit=1 gesetzt. Begann die Zeile mit einem anderen Zeichen, wurde überprüft ob \$legalhit=1 und ob das Feld den substring ‚xpect‘ enthielt. War dies der Fall, enthielt dieses Feld die zugehörigen Score und e-value Daten zu der annotierten Homologie. Der Wert des Feldes wurde dann ebenfalls in das Zielfile hinter die Annotation geschrieben und der Flag \$flag wurde gesetzt um diese (die innere) while()-Schleife zu verlassen. Die Variablen wurden hiernach wieder auf ihre Ausgangswerte gesetzt, @splitter aus dem Speicher gelöscht und der nächster Zyklus der äußeren Schleife erfolgte mit der nächsten Query-Sequenz.

Skript-4 (verwendet in 3.2.1):

```
#!/usr/bin/perl

# Script for creating a multi-FASTA file from the
# clipped sequences of all reads of the 'shotgun'-Genebank

use strict;
use warnings;

# generate filehandle for target file
my $new="c:/xxxxx/yyyyy.fas";

# generate pathhandle to sequence dir
my $genome="c:/aaaaaaa/bbbbbbb/";

# declare global temporal variable for holding filenames
my $file;

# open the directory with the sequence files
opendir DIR,$genome || die "Error can't open dir. $!\n";

##### Start outer while Loop #####
# Iterate through all files in the directory and assign the filename
# to $file in this while() loop

while ($file = readdir DIR)
{
my $checkfile;

# check for valid files
next if $file=~/\^\.\/;
next if $file=~\/err/ ;

#print "found file is $file\n"; # for cheking the filename

#concatenate pathhandle with actual filename to get absolute filepath
$checkfile = "$genome"."$file";

#print "found path is $checkfile\n"; # for checking the filepath

# open target file and print filename (= read name of the sequence run)
# as genename in FASTA format.
open OUTPUT,">>$new" || die "Error opening OUTPUT for append. $!\n";
print OUTPUT "\>";
print OUTPUT "$file\n";
```

```
close (OUTPUT);

# open the sequence-run file in line mode
open TFILE,$scheckfile || die "Error opening TFILE. $!\n";

#####      Start inner while Loop      #####

while (<TFILE>)
{
my $temp = $_;
my @tempseq;

# skip line if it contains no 4 whitespaces in a row
# or if it contains the substring 'SQ'
next if ( !($temp =~ /\s\s\s\s/) || ($temp =~ /SQ/) );

# assign the line to the tempseq array.
# each field contains a substring divided by whitespaces
@tempseq = split /\s+/, $temp;

# open the targetfile and print the first field of
# the array containing a string to the file. (= the second field)
open OUTPUT, ">>$new" || die "Error opening OUTPUT for append. $!\n";
splice (@tempseq,0,1);
print OUTPUT "@tempseq\n";
close (OUTPUT);

@tempseq = 0; # reset the array
}

#####      End of inner while Loop      #####

}

#####      End of outer while Loop      #####

#####      End of file      #####
```

Verwendete Variablen:

Alle lokal verwendeten Variablen wurden als Typ 'my' deklariert.

Filehandles:

\$new - Enthielt den Pfad zum Zielfile in welches die Ergebnisse gespeichert wurden.

Pathhandles:

\$genome - Enthielt den Pfad zu dem Verzeichnis (=path) in dem die files mit den Sequenzdaten abgespeichert waren.

Andere Skalare Variablen:

\$file - Global deklarierte Variable, die den Namen des aktuellen Sequenz-Files enthielt.

\$checkfile - Enthielt den absoluten Pfad zu dem File mit dem Namen \$file.

\$temp - Temporäre Variable in der inneren while()-Schleife. Sie enthielt den Inhalt der aktuellen Zeile des Sequenz-Files.

Arrays:

@tempseq - Enthielt in jedem Feld einen Substring einer Zeile des Sequenzfiles.

Funktionsweise des Programms:

Zuerst wurden die globalen Variablen deklariert und die Pfade zu den Files und Verzeichnissen, welche von dem Script benutzt wurden sollten, Variablen zugewiesen (=handles). Durch den opendir() Befehl wurde das Verzeichnis mit den Sequenzfiles geöffnet, und alle Files in diesem Verzeichnis wurden über eine while()-Schleife durchlaufen. In (\$file = readdir DIR) wurde durch den readdir()-Befehl bei jedem Zyklus der while()-Schleife der Name des aktuellen Files der Variablen \$file zugewiesen.

Innerhalb dieser äußeren while()-Schleife wurde zuerst die lokale Variable \$checkfile deklariert und der Filename (= \$file) auf einen legitimen Namen hin überprüft. War der Name des Files legitim, wurde durch \$checkfile = "\$genome"."\$file"; der

absolute Pfad zu dem Sequenzfile der Variablen \$checkfile zugewiesen. Hierbei wurden die Variablen \$genome (Pfad zu dem Verzeichnis) und \$file (Name des Files) miteinander verknüpft. Es wurde das Zielfile, welches die entstandenen Daten enthalten sollte, geöffnet und der Name des Sequenzfiles in das Zielfile geschrieben. Der Name des Sequenzfiles wurde im FASTA-Format mit einem führenden ‚>‘ abgespeichert.

Anschließend wurde das Sequenzfile geöffnet und mit einer zweiten inneren while()-Schleife im Zeilenmodus eingelesen. In dieser inneren while()-Schleife wurden zuerst die lokalen Variablen \$temp und @tempseq deklariert. In diesem Schritt wurde der Variablen \$temp der Inhalt der aktuellen Zeile des Sequenzfiles zugewiesen. Als nächstes wurde überprüft, ob die Zeile (=\$temp) keine vier aufeinanderfolgenden ‚Whitespaces‘ (Leerzeichen, Returns, Tabs, ...) oder den Substring ‚SQ‘ enthielt. War eines hiervon der Fall, wurde durch next() zum nächsten Zyklus übergegangen. Ansonsten wurde die Zeile durch split() and ‚Whitespaces‘ getrennt, und die einzelnen Substrings in die Felder des @tempseq Arrays geladen. Dem @tempseq Array wurde durch splice() das erste Feld entfernt, und die anderen Felder in das Zielfile geschrieben. Zuletzt wurde der @tempseq Array auf den Wert 0 zurückgesetzt, für den nächsten Zyklus der while()-Schleife. Das am Ende erzeugte Zielfile enthielt nun alle Sequenzdaten mit dem ursprünglichen Filenamen als Sequenznamen.

Skript-5 (verwendet in 3.2.1):

```
#!/usr/bin/perl

# Script for renaming Sequences in a multi-FASTA file.
# The new names are in a .txt file with one name per line
# and have to be in the same order.

use strict;
use warnings;

# declare filehandles
my $names="c:/xxxxx/aaaaa.txt";
my $genome="c:/yyyyy/bbbbbb.fas";
my $new="c:/zzzzz/ccccc.txt";

# declare global variables
my $i=0;
my @genomematrix;
my @namematrix;

# read multiFASTA file in line mode and assign
# lines to @genomematrix array fields
open INPUT1,$genome || die "Error opening file for append. $!\n";
while (<INPUT1>)
{
    @genomematrix[$i]= $_;
    $i++;
}
close (INPUT1);

# recycle the array pointer
$i=0;

# read file with clone-names in line mode and assign
# lines to @namematrix array fields
open INPUT2,$names || die "Error opening file for append. $!\n";
while (<INPUT2>)
{
    @namematrix[$i]= $_;
    $i++;
}
close (INPUT2);

# recycle the array pointer again
$i =0;
```

```
# Iterate through all @genomematrix fields. Write the value to the target file
# with the appropriate name from @namematrix in multiFASTA format.
foreach $tempseq (@genomematrix)
{
  open (OUTPUT, ">>$new") || die "Error opening file for append. $!\n";
  if ($tempseq =~ /methano/)
  {
    print OUTPUT ">";
    print OUTPUT "$namematrix[$i]\n";
    $i++;
  }
  else
  {
    print OUTPUT "$tempseq\n";
  }
  close (OUTPUT);
}

#####          End of file          #####
```

Verwendete Variablen:

Alle lokal verwendeten Variablen wurden als Typ 'my' deklariert.

Filehandles:

\$names - Enthielt den Pfad zu dem .txt File in dem Namen der Klone abgespeichert waren.

\$genome - Enthielt den Pfad zu dem multiFASTA File, in die Sequenzdaten abgespeichert waren. Wurde von Skript 4 erzeugt.

\$new - Enthielt den Pfad zu dem File in dem Namen die Ergebnisse abgespeichert wurden sollten.

Andere Skalare Variablen:

\$i - Global deklarierte Variable, die in den Iteration über Arrays als Array-Pointer diente.

\$tempseq – Temporäre lokale Variable, mit dem aktuellen Substring des multi-FASTA Sequenzfiles.

Arrays:

@genomematrix - Enthielt in jedem Feld eine Sequenz des multi-FASTA Sequenzfiles.

@namematrix - Enthielt in jedem Feld einen Substring einer Zeile des Sequenzfiles.

Funktionsweise des Programms:

Zuerst wurden die Pfade zu den Files, welche von dem Script benutzt werden sollten, Variablen zu gewiesen (Filehandles) und die globalen Variablen deklariert. Die Inhalte der Files mit den Klontnamen und den Sequenzdaten wurden zeilenweise durch eine while()-Schleife in Felder von Arrays geladen. Hierbei wurden @namematrix die Klontnamen und @genommatrix die Sequenzen zugewiesen. Als Arraypointer diente die globale Variable \$i, die zwischen den Schritten auf 0 reinitialisiert wurde.

Anschließend wurde \$i auf 0 gesetzt und durch eine foreach()-Schleife über den @genommatrix Array iteriert. Bei jedem Zyklus wurde der Wert des aktuellen Feldes der temporären Variablen \$tempseq zugewiesen. Innerhalb der Schleife wurde überprüft, ob die Variable \$tempseq den Substring ‚methano‘ enthielt. War dies der Fall wurde in das Zielfile der aktuelle Name des Klons im FASTA-Format geschrieben. Der Name wurde aus dem Feld \$i des @namematrix Arrays entnommen und \$i wurde inkrementiert. Enthielt \$tempseq nicht den Substring ‚methano‘ wurde der Wert von \$tempseq in das Zielfile geschrieben. Hiernach erfolgte der nächste Zyklus der foreach()-Schleife. Das erhaltene Zielfile enthielt alle Sequenzläufe im multi-FASTA Format. Die Namen der Sequenzen waren nun die entsprechenden Klone der Genbank.

Skript-6 (verwendet in 3.2.1):

```

#!/usr/bin/perl

# Script for BLAST of Primer sequences from a tab delimited table as INPUT.
# Output is a tab delimited .txt table with the primer sequences, the start
# and end of the homolgy and the length and name of the orf.
# Input-Table:
# column1=name/number of gene, column2=for primer,
# column3=rev primer, column4 = gene length
# columns 5+ are ignored.

use strict;
use warnings;

# Declare filehandles
my $primer="C:/dddd/eeee.txt";
my $temp="c:/xxxx/yyyy.txt";
my $outputfile = "c:/aaaa/bbbb.txt";

# Declare global flag for found orientation
my $flag = 0;

#####
# uncomment needed Variable declarations
# and comment out the other ones.
#
#
# uncomment if forward Primer BLAST
#my $primerpos = 1;
#my $primerori = "for";
#
# uncomment if reverse Primer BLAST
my $primerpos = 2;
my $primerori = "rev";
#####

open PRIMER, $primer || die "Cannot open primer file for read :$!";

#####      Start outer while Loop      #####
while (<PRIMER>)
{
    my @foo = split /\t/; # array containing primer file columns in fields
    my @layerii;        # array for holding partial blast results
    my @splitter # array containing partial blast results 1 line per field
    my $k=0;        # array pointer for @splitter array
    my $goo;        # variable for holding chomp()ed primer sequence
    my $blaster;    # variable containing raw blast results
    my $flag2=0;    # flag for mismatch hit, used to compensate arraypointer increment

```

```
print "$foo[$primerpos]\n"; #print primer sequence to screen

chomp($goo=$foo[$primerpos]);# remove terminal newline from primer seq

#print primer sequence to temp file
open TEMP, ">$temp" || die "Cannot open footemp file for read :$!";
print TEMP "$goo"|| die "Cannot open footemp file for write :$!";
close (TEMP);

#make BLAST alignment with temp file as input, load output to @blaster
$blaster = `blastall -p blastn -d mmzreads2002.fas -i $temp `
||
die "Error in blast call :$!";

# Load partial raw BLAST results into @layerii fields
@layerii = split /\*\*\*\*/,$blaster;

# for raw BLAST data dumping
#open(OUTPUT2, ">>zxxxx.txt") || die "Error opening foobar.txt file for append.
#$!\n";
#print OUTPUT2 "\nblaster null ist:\n $blaster[0]\n layer null ist:\n $layer[0]\t" || die
#"Error writing to file. $!\n";
#close (OUTPUT2);

if ($layerii[0] eq /No hits found/) # do nothing if no hits are found
{
;
}
else
{
  @splitter = split /\n+/, $blaster; #split blast result at \n+

  undef $blaster; # free memory

  chomp($foo[3]=$foo[3]); # remove terminal newline from orf length

  # print primer orientation, ORF name , ORF length
  # and Primer sequence to target file.
  open(OUTPUT, ">>$outputfile") || die "Error opening foobar.txt file for
  append. $!\n";

  print OUTPUT "\n$primerori primer\t$foo[0]\t$foo[3]\t$goo\t" || die "Error
  writing to file. $!\n";

  close (OUTPUT);
```

```
##### start of inner while Loop #####
# iterate over results for getting result datas
while ($splitter[$k])
{
  if ( ($splitter[$k]=~/>/))
  {
    if ($splitter[$k+8]=~/ $goo/i)
    {
      print "FOUND$splitter[$k+8]\n"; # check on screen
      open(OUTPUT, ">>$outputfile") || die "Error opening foobar.txt file for
      append. $!\n";
      if($flag){ print OUTPUT "\n$primerori
      primer\t$foo[0]\t$foo[3]\t$goo\t$splitter[$k]\t";}
      else {print OUTPUT "$splitter[$k]\t";}
    }
    else {print "mismatch hit found. skipping.\n";$k=$k+9;$flag2=1;}
  }

  if ($splitter[$k]=~/Plus \ Minus/)
  {
    open(OUTPUT, ">>$outputfile") || die "Error opening foobar.txt file for
    append. $!\n";

    print OUTPUT "plus\minus\t";
    $flag=1;
  }

  if ($splitter[$k]=~/Plus \ Plus/)
  {
    open(OUTPUT, ">>$outputfile") || die "Error opening foobar.txt file for
    append. $!\n";

    print OUTPUT "plus\plus\t";
    $flag=1;
  }

  if ($splitter[$k]=~/bjct/)
  {
    open(OUTPUT, ">>$outputfile") || die "Error opening foobar.txt file for
    append. $!\n";

    print OUTPUT "$splitter[$k]\t";
  }
  if ($flag2==1){$k=$k-1;$flag2=0;}
  $k++;
}
##### End of inner while Loop #####

# reset variables
$flag=0;
$flag2=0;
close (OUTPUT);
}
}
```

```
##### End of outer while Loop #####  
#  
close (PRIMER);  
#  
##### End of file #####
```

Verwendete Variablen:

Alle lokal verwendeten Variablen wurden als Typ 'my' deklariert.

Filehandles:

\$primer - Enthielt den Pfad zu der .txt Tab-getrennten Tabelle mit den Primerdaten.

\$stemp - Enthielt den Pfad zu dem temporären Auslagerungsdatei für Primer Sequenzen

\$outputfile - Enthielt den Pfad zu dem File in dem Namen die Ergebnisse abgespeichert wurden sollten.

Andere Skalare Variablen:

\$flag - Global deklarierte Flag Variable. Wurde gesetzt wenn eine Orientierung gefunden wurde.

\$primerpos – Variable, mit der Spaltennummer der Primersequenz innerhalb der Primertabelle (Wert = 1 oder 2).

\$primeori – Variable mit der Orientierung des Primers in Relation zum Gen. (Wert = „for“ oder „rev“)

\$k – Array Pointer für @splitter

\$goo – Enthielt die Primer Sequenz ohne 'Linefeed'.

\$blaster – Enthielt die Rohdaten des BLAST-Abgleichs

\$flag 2 – Flag Variable. Wurde gesetzt falls eine nicht 100%ige Homologie gefunden wurde. Wurde benutzt, um den Array Pointer \$k zurückzusetzen.

Arrays:

@foo - Enthielt in jedem Feld den Wert einer Spalte der Primer Tabelle.

@layerii - Enthielt in jedem Feld einen Teil der BLAST-Rohdaten.

@splitter - Enthielt in jedem Feld eine Zeile der BLAST-Rohdaten.

Funktionsweise des Programms:

Dieses Skript ist eine modifizierte Version des Skript-3. Der generelle Ablauf und die Modifikationen zur Dokumentation sind den Beschreibungen von Skript-3 zu entnehmen. Im folgenden wird nur auf die Unterschiede zu Skript-3 eingegangen.

Die Eingabe für dieses Programm ist eine Tab-getrennt .txt Tabelle, die mit Excel erzeugt wurde. Die Tabelle mußte wie folgt aufgebaut sein:

Spalte 1 = Name/Nummer des Gens

Spalte 2 = Sequenz des Vorwärts-Primers

Spalte 3 = Sequenz des Rückwärts-Primers

Spalte 4 = Länge des Gens

Die Spalten 5 und höher wurden von dem Skript ignoriert.

Als Query (=Eingabe Sequenz) für den BLAST-Abgleich wurde eine Primer Sequenz benutzt und in ein temporäres File geschrieben. Dieses File wurde wie in Skript-3 gegen die BLAST Datenbank abgeglichen. Als Datenbank wurde in diesem Skript das multiFASTA-File aus Skript-5 verwendet.

In der inneren while()-Schleife wurde zusätzlich zu Skript-3 die zugehörige Orientierung und die Sequenz der Homologie über 3 if()-Klauseln abgefragt. War die Sequenz der Homologie nicht identisch mit der Primer Sequenz wurde \$flag2=1 gesetzt, das Ergebnis verworfen und weiter gesucht. Ansonsten wurden in das Zielfile folgende Werte mit Tabs getrennt geschrieben.

- Genname,
- Genlänge,
- Primersequenz,
- Name des Klons mit der homologen Sequenz,
- Homologiesequenz mit Start und Stop

Das erhaltene File erhielt in jeder Zeile diese Werte für jeweils eine Homologie.

Danksagungen

Herrn Priv. Doz. Dr. U. Deppenmeier danke ich für die Überlassung des Themas, und der Möglichkeit zum selbständigen Arbeiten. Besonders dankbar bin ich ihm für die kritische und sorgfältige Durchsicht des Manuskripts dieser Arbeit und für sein Engagement um dieses Projekt zu ermöglichen.

Herrn Prof. Dr. G. Gottschalk danke ich dafür, dass er sich als Korreferent zur Verfügung gestellt hat.

Allen ehemaligen (Holger Brüggemann und Sebastian Bäumer) und derzeitigen (Sabine Lentes, Kai Hofmann, Marc Hoffmeister, Frank Fallinski, Tina Prust, Hauke Schmidt, Marcell Grapp) Mitarbeitern im Labor 206 danke ich für das nette Arbeitsklima, aber auch für schöne Mittagspausen in der Mensa.

Meiner Frau Joanna danke ich für Ihre Geduld und seelische Unterstützung während der Promotionsphase und meinem Sohn Melvin dafür, dass er mich mit seinem Lächeln immer wieder ermutigte.

Sabine Lentes danke ich besonders für die Hilfe bei der Erstellung der Plasmidbank, dem Grundstock für die DNA-Microarray-Chips.

Für die Korrekturarbeiten und sein Wissen über die neue deutsche Rechtschreibung sei ganz herzlich meinem lieben Freund Martin Schindel gedankt.

Außerdem gilt mein Dank denjenigen, die ein „Dankeschön“ verdient haben, hier aber unerwähnt bleiben.

So bleibt noch der herzliche Dank an meine Eltern, die mir dieses Studium inklusive Promotion ermöglicht haben und mir nicht nur mit Rat, sondern auch oft mit tatkräftiger Unterstützung zur Seite standen.

Lebenslauf

- 11.06.1974 Geboren in Göttingen
- 1980 – 1984 Besuch der Erich Kästner Grundschule in Göttingen
- 1984 – 1986 Besuch der Jahnschule Orientierungsstufe in Göttingen
- 1986 – 1990 Besuch der Voigt Realschule in Göttingen
- 1990 – 1991 Besuch des Fachgymnasiums Wirtschaft (BBS I) in
Göttingen
- 1991 – 1993 Besuch des Hainberg-Gymnasiums in Göttingen
- 1993 Abitur am Hainberg-Gymnasium
- Oktober 1993 Immatrikulation an der Georg-August-Universität in
Göttingen für das Fach Biologie auf Diplom
11. April 1996 Vordiplom in den Fächern Genetik, Zoologie, Chemie und
Physikalische Chemie
06. Juli 2000 Diplom in den Fächern Genetik (Hauptfach), Mikrobiologie
und Organische Chemie. Die Diplomarbeit trug den Titel:
„Funktionelle Konsequenzen der Mutagenese des
membrangebundenen Chemokinrezeptors CCR5“
- Oktober 2000 Immatrikulation an der Georg-August-Universität in
Göttingen für das Fach Angewandte Informatik
- April 2001 Beginn der experimentellen Arbeiten zur vorliegenden
Dissertation