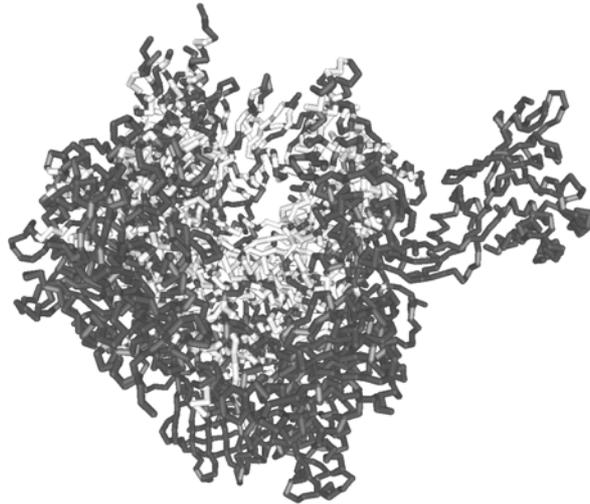


**Protein-Protein-Wechselwirkungen der Untereinheiten der RNA-Polymerase
von *Pyrococcus furiosus***



Dissertation

zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Bernd Goede

Kiel 2003

1 Inhaltsverzeichnis

1	<u>Inhaltsverzeichnis</u>	1
2	<u>Einleitung</u>	3
2.1	Anzahl und Benennung der RNA-Pol-Untereinheiten.....	4
2.2	Komponenten & Mechanismus der Transkription in <i>Archaea</i> , <i>Eukarya</i> , <i>Bacteria</i>	7
2.3	Ziel der Arbeit:	10
2.4	Anmerkungen zur Gliederung der Arbeit:.....	11
3	<u>Material und Methoden</u>	12
3.1	Bakterienstämme	12
3.2	Verwendete Plasmide	12
3.3	Chemikalien, Enzyme und Materialien	12
3.4	Klonierung der Untereinheiten der RNAP und Liste der verwendeten Primer	13
3.5	Herstellung kompetenter Zellen für die Transformation, CaCl ₂ -Methode	15
3.6	Far Western Blotting zur Analyse von Protein-Protein-Interaktionen	17
3.6.1	Radioaktive Markierung von Proteinen für Far Western Blotting	17
3.6.2	Plasmid pET-33b(+) (Novagen Inc.).....	18
3.6.3	Konstruktion einer C-terminalen Plasmid-Version: pET-33b(+)c2	19
3.6.4	Far Western Dot Blotting	22
3.6.5	Far Western Blotting.....	22
3.6.6	Zusammenfassung: Far Western Blotting in Stichworten	22
3.6.7	Blockierlösung für Far Western Blots	24
3.6.8	Beispiele für Markierungsansätze mit Proteinkinase A (PKA).....	24
3.7	Herstellung von PA-Gradientengelen (denaturierend oder nativ)	25
3.8	Bioinformatik: Verwendete Programme und (Online-)Tools	27
4	<u>Ergebnisse und Diskussion</u>	29
4.1	Untereinheiten der DNA-abhängigen RNA-Polymerase bei <i>Pyrococcus furiosus</i> .	29
4.2	Untereinheiten der RNA-Polymerase in archaellen Genomen	30
4.2.1	Zusammenstellung von Genomdaten für eine vergleichende Darstellung.....	32
4.2.2	Zuverlässigkeit und Aussagekraft von Genomdaten / bioinformatischen Analysen	32
4.2.3	Weitere Eckdaten der untersuchten archaellen Genome (Statistik).....	33
4.3	Auswertung von Alignments der drei <i>Pyrococcus</i>-Arten untereinander	34
4.3.1	Eigenschaften und konservierte Regionen der RNAP-Gene in <i>Pyrococcus</i>	37
4.3.1.1	B (rpoB) aus <i>P.furiosus</i> : Konservierte Regionen und Eigenschaften	39
4.3.1.2	A' (rpoA1) aus <i>P.furiosus</i> : Konservierte Regionen und Eigenschaften	40
4.3.1.3	A'' (rpoA2) und H (rpoH): Konservierte Regionen und Eigenschaften	41
4.3.1.4	D (rpoD) und L (rpoL): Konservierte Regionen bei <i>P.furiosus</i> :	41
4.3.1.5	E' (rpoE1), E'' (rpoE2) und F (rpoF): Konservierte Regionen bei <i>P.furiosus</i>	42
4.3.1.6	K (rpoK), N (rpoN) und P (rpoP): Konservierte Regionen bei <i>P.furiosus</i> :	44
4.3.1.7	Zusammenfassung: Konservierte Regionen	44
4.4	Zinkfinger-Motiv und katalytisches Zentrum in der A'-Untereinheit	45
4.4.1	Ist die hochkonservierte A'-Untereinheit geeignet für Stammbaum-Erstellung?	46
4.5	E' / E'': Überlappende Gene im Rahmen translationaler Kopplung von Genen	47
4.6	Zusammenfassende grafische Darstellung von homologen Genabschnitten	49
4.6.1	3D-Struktur der RNA-Pol II mit gegenüber <i>Archaea</i> konservierten Regionen	54
4.7	Liefert die Aminosäuresequenz Hinweise auf Hitzestabilität bzw. Löslichkeit?	59
4.7.1	Durchschnittliche Aminosäure-Zusammensetzung in den Proteingenen im Genom	65
4.8	Eine zweite Kopie von TFB im Genom von <i>Pyrococcus</i>	69
4.9	Klonierung der Untereinheiten der DNA-abhängigen RNA-Polymerase	72
4.10	Expression und Reinigung der RNA-Pol-Untereinheiten	72
4.11	Untersuchung von Protein-Protein-Wechselwirkungen	76
4.11.1	Bindungsstudien: Far Western Blotting-Experimente.....	76

4.11.1.1	Far Western Dot Blotting	76
4.11.1.2	Far Western Blotting.....	80
4.11.1.3	Zusammenfassung der Far Western Blotting-Ergebnisse und Diskussion	83
4.11.2	Bindungsstudien: Nativgel-Experimente	89
4.11.3	Bindungsstudien: Experimente mit Molekularsieb-Chromatographie.....	90
4.12	Reinigung der nativen RNA-Polymerase nach modifiziertem Reinigungsschema.	90
4.12.1	Nachweis von Untereinheit F im Western Blotting Experiment	92
4.12.2	Untersuchungen mit MALDI-TOF-MS	95
5	Zusammenfassung.....	97
6	Summary	98
7	Literaturverzeichnis	99
8	Anhang	107
8.1	Multiple Alignments	107
8.1.1	Vorbemerkung und Zusatzinformationen zu den Multiplen Alignments	107
8.1.2	Alignments der drei <i>Pyrococcus</i> -Arten untereinander.....	107
8.1.2.1	Untereinheit B (rpoB).....	107
8.1.2.2	Untereinheit A' (rpoA1).....	108
8.1.2.3	Untereinheit A'' (rpoA2).....	108
8.1.2.4	Untereinheit D (rpoD)	109
8.1.2.5	Untereinheit E' (rpoE1).....	109
8.1.2.6	Untereinheit E'' (rpoE2).....	109
8.1.2.7	Untereinheit F (rpoF).....	109
8.1.2.8	Untereinheit H (rpoH)	110
8.1.2.9	Untereinheit K (rpoK).....	110
8.1.2.10	Untereinheit L (rpoL)	110
8.1.2.11	Untereinheit N (rpoN)	110
8.1.2.12	Untereinheit P (rpoP).....	110
8.1.3	Alignments mit <i>Archaea</i> und <i>Eukarya</i>	111
8.1.3.1	Untereinheit B und B' / B'' (rpoB, rpoB1, rpoB2)	111
8.1.3.2	Untereinheit A' (rpoA1).....	115
8.1.3.3	Untereinheit A'' (rpoA2).....	118
8.1.3.4	Untereinheit D (rpoD)	120
8.1.3.5	Untereinheit E' (rpoE1).....	121
8.1.3.6	Untereinheit E'' (rpoE2).....	121
8.1.3.7	Untereinheit F (rpoF)	122
8.1.3.8	Untereinheit H (rpoH)	123
8.1.3.9	Untereinheit K (rpoK).....	123
8.1.3.10	Untereinheit L (rpoL)	124
8.1.3.11	Untereinheit N (rpoN)	124
8.1.3.12	Untereinheit P (rpoP).....	125
8.1.3.13	Aminosäure-Austausch: Bewertung der „Ähnlichkeit“ in Alignments	125
8.2	Liste verfügbarer 3D-Strukturen zum Thema RNA-Polymerase & Transkription .	126
8.3	Wie mache ich eine lokale BLAST-Suche?.....	128
8.4	„GenbankAnalyse“ – ein selbst geschriebenes Bioinformatik-Tool.....	130
8.4.1	Berechnung der theoretischen Nettoladung und des Isoelektrischen Punktes	131
8.5	Liste der mikrobiellen Genom-Dateien im Genbank-Format (Internet)	132
8.6	Accession-Nummern einiger in dieser Arbeit erwähnter Gene	133
8.7	Abkürzungen und Begriffe	134
8.8	Abbildungsverzeichnis	135
8.9	Danksagung.....	137
8.10	Lebenslauf	138
8.11	Erklärung.....	139

2 Einleitung

Wie schafft es ein hyperthermophiler Organismus, trotz extrem hoher Temperaturen seine strukturelle Integrität zu bewahren? Wie ist es möglich, dass Lebewesen selbst in kochend heißem Wasser in der Lage sind, komplexe und äußerste Präzision verlangende stoffwechselphysiologische Aufgaben zu gewährleisten, wie die korrekte Ablesung der Gene für die jeweils gerade benötigten Proteine (und Nicht-Proteingene wie tRNAs und rRNA)? Dieser „Transkription“ genannte Teilbereich des Zellstoffwechsels wird in der Arbeitsgruppe von Prof.Dr.M.Thomm intensiv u.a. an dem optimal bei 100°C wachsenden Archaeon (Archäobakterium) *Pyrococcus furiosus* (Fiala & Stetter, 1986) untersucht, wobei es möglich war, ein zellfreies Transkriptionssystem zu etablieren, mit dem eine spezifische Transkription bis zu einer Temperatur von 95°C möglich ist (Hethke *et al.*, 1996; Hethke *et al.*, 1999).

Natürlich ist *P.furiosus* nicht der einzige hyperthermophile, d.h. bei > 90°C wachsende Organismus (bzw. „extrem thermophil“, bei Wachstum >75°C), was sich leicht an dem folgenden Stammbaum demonstrieren läßt:

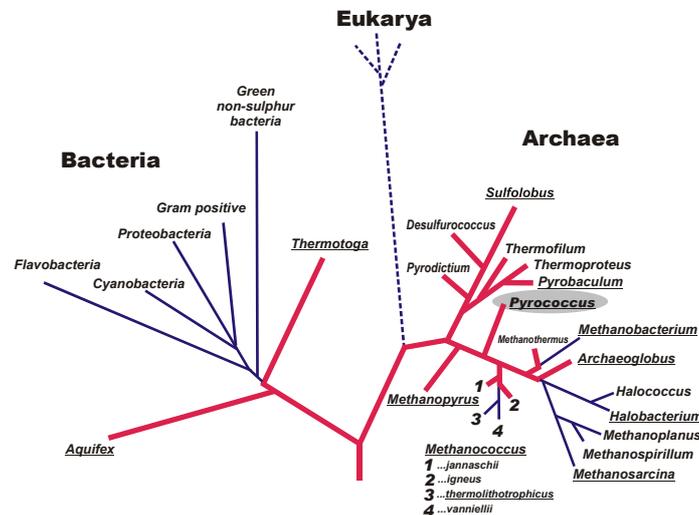


Abb.2-01: Stammbaum mit *Pyrococcus furiosus*, einem hyperthermophiler Vertreter der Archaea
 Modifiziert nach (Blöchl *et al.*, 1995). In diesem auf 16S rRNA-Analyse beruhenden Stammbaum sind hyperthermophile Arten rot und mesophile Organismen blau markiert. Im folgenden genauer untersuchte Arten sind unterstrichen. Die Gattung *Methanococcus* wurde mittlerweile in *Methanocaldococcus* umbenannt. *Methanobacterium* heißt jetzt *Methanothermobacter*.

Der Stammbaum zeigt eine zuerst von C.Woese postulierte (und inzwischen weitgehend anerkannte) Gliederung aller Lebewesen in drei große Gruppen (Domänen genannt), die *Archaea*, *Bacteria* und *Eukarya* (Woese & Fox, 1977; Woese *et al.*, 1990; Armache *et al.*, 2003). Die Kalkulation der phylogenetischen Verwandtschaftsverhältnisse auf Basis eines einheitlichen Kriteriums, nämlich der Struktur der 16S rRNA (bzw. 18S rRNA bei *Eukarya*) war revolutionär, weil die zuvor z.B. nach Organisationstyp und phänotypischen Eigenschaften (bei Bakterien oft Stoffwechseleigenschaften) aufgestellten Stammbäume oftmals nur in Teilbereichen für alle Merkmale kohärent waren. Das Muster der seltenen, aber regelmäßig im Laufe der Zeit auftretenden Mutationen im Gen eines hochkonservierten Marker-Moleküls wie der 16S rRNA kann dagegen mit computerbasierten Methoden in einen universellen Stammbaum umgerechnet werden.

Interessanter Weise zeichnete sich hier ein „heißer Ursprung“ des Lebens ab, d.h. in dem oben gezeigten Stammbaum erscheint die Mesophilie (also die Anpassung an moderate Temperaturen) jeweils als abgeleitetes Merkmal (selbst innerhalb einer Gattung, siehe *Methanocaldococcus*), während die hyperthermophilen Organismen die am tiefsten abzweigenden (und somit „ursprünglichsten“) Vertreter stellen. Allerdings existieren auch Daten, die gegen eine „primäre Hyperthermophilie“ sprechen, wie aktuelle Diskussionen zeigen (Islas *et al.*, 2003). Dass die

hyperthermophilen Gattungen *Aquifex* und *Thermotoga* wirklich an der Basis der *Bacteria* stehen, ist kürzlich erneut bestätigt worden (Di Giulio, 2003).

Während also bei *Bacteria* einige wenige und bei *Archaea* eine Vielzahl von Gattungen eine Anpassung an hohe Temperaturen zeigen, findet sich bei Eukarya (also den „höheren Lebewesen“ mit Tieren, Pflanzen und Pilzen) bisher kein einziges entsprechendes Beispiel.

Das Thema dieser Arbeit, nämlich „Protein-Protein-Wechselwirkungen der Untereinheiten der RNA-Polymerase von *Pyrococcus furiosus*“, deutet den komplexen Aufbau der neben den Transkriptionsfaktoren TBP und TFB und der Promotor-DNA mit TATA-Box (Thomm, 1996) wichtigsten Komponente für spezifische Transkription, der DNA-abhängigen RNA-Polymerase an. Da in dieser Arbeit ausschließlich DNA- und nicht RNA-abhängige RNA-Polymerasen untersucht und diskutiert werden (letztere kommen z.B. bei RNA-Viren vor), wird im folgenden meist nur noch von „RNA-Polymerase“ (ohne Zusatz) bzw. in Kurzform als „RNAP“ gesprochen.

Abgesehen von viralen RNA-Polymerasen, wie der (DNA-abhängigen) T7-RNA-Polymerase, die z.B. für die Überexpression rekombinanter Proteine in einigen *E.coli*-Expressionsstämmen (z.B. BL21(DE3)) verwendet wird, sind RNA-Polymerasen immer aus mehreren Untereinheiten aufgebaut, wobei die Komplexität in bezug auf deren Anzahl bei *Eukarya* und *Archaea* mit ca. 12 höher als bei *Bacteria* (5 Untereinheiten) ist. Bevor auf die domänen-spezifischen, grundlegenden Unterschiede des Mechanismus der Transkription eingegangen wird (z.B. erfolgt die Rekrutierung der RNA-Pol bzw. die spezifische Promotor-Erkennung bei *Bacteria* über alternative σ -Faktoren und nicht über TFB / TBP wie bei *Archaea* und *Eukarya*) zunächst ein Überblick:

2.1 Anzahl und Benennung der RNA-Pol-Untereinheiten

Wie auch der Tabelle unter [Abb.4-05] zu entnehmen ist, ist die Benennung der Gene der Untereinheiten der DNA-abhängigen RNA-Polymerase alles andere als einheitlich, was die „historische“ Entwicklung der Forschung auf diesem Gebiet widerspiegelt. Vor allem bei *S.cerevisiae* werden 4 – 5 Synonyme pro Gen gefunden, z.B. ist auch eine Bezeichnung wie „Rpa190“ für die größte Untereinheit bei Pol I üblich, bei der das ermittelte Molekulargewicht im Namen enthalten ist. Zumindest bei Pol II hat sich aber die Bezeichnung „Rpb + Zähler“ durchgesetzt, d.h. Rpb1 für die größte, und Rpb12 für die kleinste Untereinheit („Rpa“ und „Rpb“ dann entsprechend für Pol I und Pol III), wobei allerdings einige Untereinheiten in allen drei Polymerasen enthalten sind. Dies ist für Rpb5, Rpb6, Rpb10 und Rpb12 der Fall. Rpb19 kommt in Pol I und Pol II vor. Jedoch ist hier (und besonders bei Pol I und Pol III) manchmal nicht ganz eindeutig, wo von einer RNA-Pol-Untereinheit, und wo von einem RNA-Pol-assoziierten Faktor gesprochen werden muß. Gerade die eukaryotischen RNA-Polymerasen sind von einer Vielzahl von Hilfsproteinen und (z.B. die Transkriptionsaktivität modulierenden) Faktoren umgeben.

Die Untereinheit E'' ist in [Abb.2-02] bei *Pyrococcus* / *Archaea* noch nicht eingetragen, sollte aber (zumindest während bestimmter Funktionszustände der RNA-Pol) als Bestandteil des RNA-Pol-Komplexes nachgewiesen werden können (zur Überlappung des E''-Genes mit E' siehe [4.5]; E''/rpoE2 ist etwas größer als K/rpoK).

Innerhalb von *Archaea* hat sich die Benennung der Untereinheiten nach dem Schema „rpo + großgeschriebener Buchstabe“ durchgesetzt – es findet sich aber z.B. auch eine Bezeichnung wie „rpoB“ statt „rpoB1“ (*Halobacterium* sp.). Um hier keine Verwirrung aufkommen zu lassen, habe ich mich in dieser Arbeit an folgende Schreibweisen gehalten:

Die Gene der 12 Untereinheiten (bzw. 13 z.B. bei den methanogenen *Archaea*) der archaellen, DNA-abhängigen RNA-Polymerase werden mit „**rpoB (rpoB1 + rpoB2), rpoA1, rpoA2, rpoD, rpoE1, rpoE2, rpoF, rpoH, rpoK, rpoL, rpoN, rpoP**“ bezeichnet.

Bei den resultierenden Proteinen / Produkten wird von „**B (B' + B''), A', A'', D, E', E'', F, H, K, L, N, P**“ gesprochen, bzw. von der „B-Untereinheit der RNA-Polymerase“ und so weiter.

Weil letzteres ziemlich lang ist, die Bezeichnung ausschließlich mit einem einzigen Buchstaben aber irreführend sein kann, wurde in vielen Fällen die knappe und eindeutige Bezeichnung der Gene verwendet, obwohl eigentlich das Gen-Produkt gemeint ist. In der Regel ist dies aber eindeutig aus dem Kontext zu erschließen.

Dies gilt auch für die bei den Alignments und Stammbäumen verwendete Form der Art „*Pfu_rpoB*“ etc., bei der ich zusätzlich den ersten Buchstaben der Gattung in Verbindung mit den ersten beiden Buchstaben des Artnamens hinzugefügt habe. Diese Kombination ist in der Literatur eher ungebrauchlich (zumindest in Verbindung mit dem zusätzlichen Unterstrich) – es können auch die (teilweise sehr begrenzt aussagekräftigen) Bezeichnungen aus Tabelle [Abb.4-05] oder aber Umschreibungen evtl. unter Einbeziehung von Accession Number und ID aus der entsprechenden Protein-Datenbank verwendet werden (falls vorhanden).

Ein Namenskonflikt zeigt sich beim Vergleich der archaellen RNA-Pol-Gene mit denen bei *Bacteria*: Auch hier werden die RNA-Pol-Gene mit „RPOA“ etc. benannt – jedoch darf aufgrund desselben Buchstabens nicht auf Homologie geschlossen werden! (RPOD bezeichnet z.B. einen Sigma-Faktor, während *rpoD* aus *Archaea* zur Alpha-Untereinheit der RNA-Pol aus *Bacteria* homolog ist). Keine Verwechslungsgefahr besteht, wenn die Buchstaben des griechischen Alphabets verwendet werden, also das bekannte Schema „β'β+2x α+ ω,, (+ (alternative) σ - Faktoren).

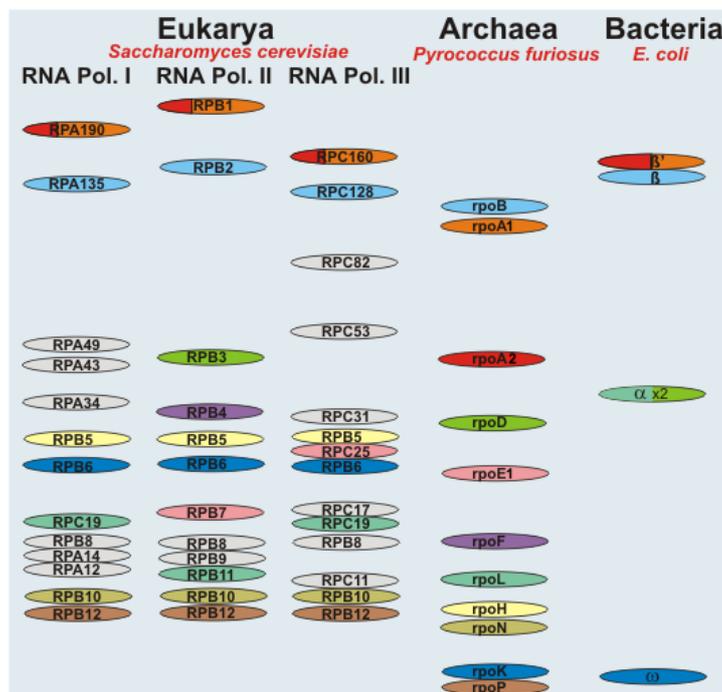


Abb.2-02: Anzahl und Größe der RNA-Pol-Untereinheiten bei Eukarya, Archaea und Bacteria

Sequenzhomologien zwischen den Untereinheiten der verschiedenen RNA-Polymerase-Typen wurden farblich markiert. In einigen Fällen besitzen zwei Untereinheiten Homologie zu einem einzigen Protein beim Vergleichsorganismus (*rpoA1* und *rpoA2* bei *Pyrococcus* gegenüber *Rpa190*, *Rpb1*, *Rpc160* und β' , sowie *Rpb3* und *Rpb11* bzw. *rpoD* und *rpoL* gegenüber der bakteriellen α -Untereinheit). Sofern es der Platz zuließ, wurden die Proteine ihrem Molekulargewicht entsprechend halblogarithmisch angeordnet. Die Gewichte reichen von 5757 Da für *rpoP* aus *P.furiosus* bis ca. 190 kDa für *Rpb1* aus Pol II.

Wie gut zu erkennen ist, folgen alle gezeigten RNA-Polymerasen vom Aufbau her prinzipiell dem Schema „Zwei große Untereinheiten, die durch mehrere kleine Untereinheiten ergänzt werden“ (Sonntag & Darai, 1995). Bei *Archaea* ist allerdings die größte Untereinheit mit dem katalytischen Zentrum (also das Homolog zu *Rpb1* der eukaryotischen RNA-Polymerase II (Pol II) und β' bei *Bacteria*) in die beiden Teile *rpoA1* und *rpoA2* gespalten. Bei einigen *Archaea* (u.a. den methanogenen *Archaea*, siehe [Abb.4-05]) ist zusätzlich *rpoB* in *rpoB1* und *rpoB2* gespalten, so dass in diesem Fall wieder die das katalytische Zentrum enthaltende Untereinheit *rpoA1* das größte Molekulargewicht besitzt. Das universell in DNA-abhängigen RNA-Polymerasen enthaltende Sequenzmotiv „NADFDGD“ ist verantwortlich für die Bindung des zentralen Mg^{2+} -Ions im katalytischen Zentrum in Form eines Chelates (Sonntag & Darai, 1995; Zaychikov *et al.*, 1996). Dies ist (zusammen mit weiteren hochkonservierten Abschnitten wie z.B. dem „RQP(TS)LH“-Motiv 26-31 Aminosäuren stromaufwärts vom Hauptmotiv (Sonntag & Darai, 1995)) ein starkes

Indiz für eine domänen-übergreifende Ähnlichkeit der RNA-Polymerasen sowohl in funktionaler als auch struktureller Hinsicht, was sich auch im Detail in einer Vielzahl weiterer charakteristisch konservierter Sequenzmuster zeigt (Bartlett *et al.*, 2000). Obwohl die bakterielle RNA-Pol ($\beta'\beta\alpha_2\omega$) in bezug auf Anzahl und Molekulargewicht deutlich von den archaeellen und eukaryotischen Polymerasen abzuweichen scheint, finden sich bei der Analyse der offensichtlich aus „Strukturmodulen“ aufgebauten großen Untereinheiten (Severinov *et al.*, 1996) viele Ähnlichkeiten.

Trotzdem stellt natürlich die eukaryotische Pol II (aus dem „Modellorganismus“ *Saccharomyces cerevisiae* = Bäckerhefe) die Hauptreferenz für Vergleiche mit der in dieser Arbeit untersuchten RNA-Polymerase aus *Pyrococcus furiosus* dar, u.a. weil hier zuerst eine 3D-Struktur für einen Komplex aus 10- bzw. kürzlich auch aus 12-Untereinheiten erstellt wurde (Cramer *et al.*, 2000; Cramer *et al.*, 2001; Armache *et al.*, 2003). Die zweite Hauptreferenz stellt die vor einiger Zeit aus rekombinanten Untereinheiten konstituierte archaelle RNA-Polymerase aus *Methanocaldococcus jannaschii* dar, für die spezifische katalytische Aktivität nachgewiesen wurde (Werner & Weinzierl, 2002). Allerdings existiert hier noch keine 3D-Struktur für den Gesamtkomplex. Die 3D-Struktur eines Komplexes aus den Untereinheiten rpoE1 und rpoF aus *M.jannaschii* (Todone *et al.*, 2001) wurde in die Interpretation der 3D-Struktur der E'- und F-Homologe Rpb4 und Rpb7 auf Basis der bisher detailliertesten eukaryotischen Kristallstruktur der Pol II mit einbezogen (Armache *et al.*, 2003).

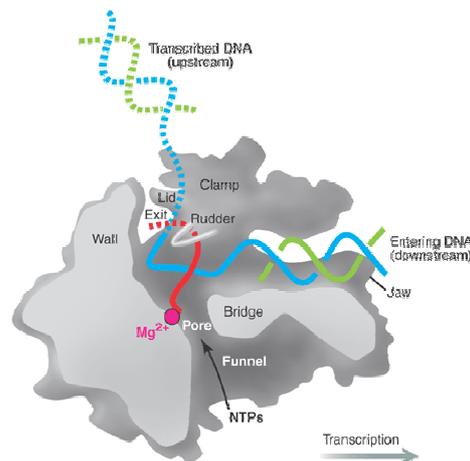


Abb.2-03: Schematische 3D-Darstellung: Prinzipielle Arbeitsweise einer RNA-Polymerase

Abbildung aus (Klug, 2001), siehe auch (Cramer, 2002). Es handelt sich um die eukaryotische RNA-Polymerase aus *S.cerevisiae*. Gezeigt wird die Elongationsphase der Transkription, bei der die RNAP sich bereits von den anderen, zur Initiation der Transkription benötigten Komponenten gelöst hat, und die DNA entlang wandert. Es handelt sich um einen Längsschnitt durch den Enzymkomplex entlang des großen Spaltes, an dessen Grund der Weg der DNA verläuft (grün-blaue Doppelhelix) und sich das katalytische Zentrum der RNAP mit dem komplexierten Mg^{2+} -Ion befindet (roter Kreis). Die Richtung der Transkription verläuft stromabwärts, in diesem Fall also nach rechts. Am anderen Ende der Spalte (in der Grafik oben schräg-links) verlassen die transkribierte DNA und die erfolgreich synthetisierte RNA (rote Linie) das Enzym (der grüne, nicht codogene DNA-Strang wurde der Übersichtlichkeit halber teilweise weggelassen). Die NTPs, als Substrat für die mRNA-Synthese gelangen über einen Trichter (Funnel), der in einer Pore mündet über einen besonderen Zugang direkt an das katalytische Zentrum. Eine Brücke (Bridge) genannte Proteinstruktur ist für die Auftrennung der beiden DNA-Stränge wichtig, die Voraussetzung für die "Ablösung der DNA" und die Ausbildung des DNA-RNA-Hybrids ist. Eine große, klammerartige Struktur (Clamp) deckt die Spalte nach einer durch die Anwesenheit von DNA und DNA-RNA-Hybrid induzierten, starken Konformationsänderung teilweise ab, was den Elongationskomplex stabilisiert.

Der geschilderte Mechanismus ist das Grundprinzip aller RNA-Polymerasen, wenngleich der Enzymaufbau nicht in allen Fällen so komplex ist (z.B. bei der T7-RNA-Polymerase aus einer einzigen Untereinheit). Die Abbildung wird als erster Überblick schon an dieser Stelle gezeigt, weil im folgenden recht lange von den einzelnen Untereinheiten, deren Details und beteiligten Faktoren die Rede sein wird, und erst in einem späteren Teil dieser Arbeit wieder auf die RNA-Polymerase und ihre Funktion als ganzes eingegangen wird.

2.2 Komponenten & Mechanismus der Transkription in *Archaea*, *Eukarya*, *Bacteria*

Diese Arbeit beschäftigt sich im wesentlichen mit den Eigenschaften und den Wechselwirkungen der einzelnen Untereinheiten der RNA-Polymerase aus *P. furiosus*. Der eigentliche Vorgang der spezifischen Transkription und seiner Regulation im Rahmen der Genexpression war dagegen (abgesehen von der Überprüfung der spezifischen Aktivität bei der Reinigung des nativen Enzyms aus *Pyrococcus*) nicht Thema der praktischen Experimente.

Um die im folgenden beschriebenen Unterschiede bei Anzahl, Struktur und Anordnung der RNA-Polymerase-Untereinheiten besser einordnen zu können, soll trotzdem kurz auf die Hauptunterschiede bei der archaeellen, eukaryotischen und bakteriellen Transkription eingegangen werden. Diese Einteilung stützt implizit gleichzeitig die genannte Einteilung aller Lebewesen in die drei oben diskutierten Domänen, weil sich in bezug auf den Transkriptionsapparat domänenspezifische (teilweise einzigartige) Besonderheiten finden:

Transkription bei *Bacteria*:

Wie bei *Archaea*, gibt es bei *Bacteria* nur eine DNA-abhängige RNA-Polymerase, die sowohl die mRNA für die Expression aller Proteingene, als auch rRNAs, tRNAs etc. generiert. Diese mRNAs können, wie bei *Archaea* eine polycistronische Struktur aufweisen. Die Core-Polymerase aus vier Untereinheiten (" $\beta'\beta\alpha_2$ ", d.h. die α -Untereinheit ist zweimal vorhanden) (Helmann & Chamberlin, 1988) wird durch die später entdeckte Omega-Untereinheit ergänzt, die in stöchiometrischen Mengen an die übrigen Untereinheiten assoziiert ist. Diese kleinste Untereinheit (s. Abb. 2) scheint wichtig für ein korrektes "Assembly" der RNA-Polymerase zu sein, und bindet offensichtlich an den N- und C-Terminus der größten Untereinheit β' , die das katalytische Zentrum enthält (Ghosh *et al.*, 2001), so dass sich das Untereinheiten-Muster " $\beta'\beta\alpha_2\omega$ " ergibt.

In Gegensatz sowohl zu *Archaea*, als auch *Eukarya* erfolgt die für eine gezielte Transkription nötige Erkennung der vor den jeweiligen Genen in einem bestimmten Abstand vom Transkriptionsstartpunkt stromaufwärts auf der DNA gelegenen Promotorsequenz über σ -Faktoren. Bei dem Promotor handelt es sich um die so genannte TATA-Box (Konsensus 5'-TATAAT-3'), auch "Pribnow-Schaller-Box" genannt (Pribnow, 1975; Schaller *et al.*, 1975) –10 bp vor dem Transkriptionsstart, sowie um die vor allem bei konstitutiv exprimierten Genen gefundene 2.Box bei Position –35 bp (Konsensus 5'-TTGACA-3'). Das Prinzip einer globalen Regulation ganzer Funktionsgruppen von Genen über "alternative σ -Faktoren", die nicht zuerst an den Promotor, sondern direkt an die RNA-Polymerase binden, um so deren Affinität zu bestimmten Promotor-Typen zu verstärken, scheint sich im Rahmen der Evolution als schlagkräftige Alternative zum (weiter unten beschriebenen) Mechanismus bei *Archaea* und *Eukarya* bewährt zu haben. Typisches Beispiel ist z.B. die so genannte "Hitzeschock-Antwort", bei der eine Stress-Situation durch eine Umstellung des Zellstoffwechsels auf ein anderes Expressionsmuster über den σ^{32} -Faktor erreicht wird (z.B. zur vermehrten Produktion von proteinstabilisierenden Chaperon-Proteinen). Darauf, dass es in bezug auf das Gen für den Standard- σ -Faktor σ^{70} bei *E. coli* und für die Untereinheit D der RNA-Polymerase bei *Archaea* eine Namensüberschneidung gibt (in beiden Fällen „rpoD“) wurde bereits hingewiesen. Obwohl man vermuten könnte, dass die bakteriellen Promotoren einem simplen Aufbau entsprechend den verschiedenen möglichen σ -Faktoren folgen, ist ihre Erkennung mit bioinformatischen Methoden auf Genom-Ebene aufgrund der beobachteten Vielfalt nach wie vor alles andere als trivial (Eskin *et al.*, 2003).

Transkription bei *Eukarya*:

Die Situation bei *Eukarya* ist schon deshalb komplizierter als bei *Bacteria* und *Archaea*, weil hier nicht eine, sondern gleich drei verschiedene DNA-abhängige RNA-Polymerasen existieren, die aber vom Prinzip her viele strukturelle (und funktionale) Ähnlichkeiten besitzen. So besitzen Pol I, Pol II und Pol III beispielsweise fünf gemeinsame Untereinheiten (Rpb5, Rpb6, Rpb8, Rpb10 und Rpb12). Während Pol I für die Synthese von 5,8S-, 18S- und 28S-rRNA zuständig ist, übernimmt die Pol III die Produktion von 5S-rRNA, tRNAs und snRNAs. Die Pol II synthetisiert alle für die Proteinexpression benötigten mRNAs, und ähnelt in Aufbau und Funktion am stärksten der archaeellen RNA-Polymerase. Aus diesem Grund (und vor allem, weil an dem eukaryotischen

"Modellorganismus *Saccharomyces cerevisiae*" die Transkription intensiv erforscht wird) stellt die Pol II das wichtigste Referenz-System bei der Erforschung von Details der RNA-Polymerase bei *Archaea* dar. Für die unterschiedlichen eukaryotischen Polymerasen gibt es jeweils typische Muster im Promotoraufbau, auf die hier nicht näher eingegangen wird. Interessanter Weise findet sich bei den Pol II-Promotoren oft ähnlich wie bei *Archaea* eine bei -25 bp gelegene TATA-Box (Konsensussequenz 5'-TATA^A/_TA^A/_T-3') (Breathnach & Chambon, 1981; Bucher, 1990), sowie weitere sogenannte CCAAT- und GC-Boxen (Dierks *et al.*, 1983; Kreidberg & Kelly, 1986), was insgesamt auch einen ähnlichen Erkennungsmechanismus von Seiten der beteiligten Faktoren und der RNA-Polymerase vermuten läßt. Grundsätzlich ähnlich (und ein prinzipieller Unterschied zu *Bacteria*) ist die Bildung eines Präinitiationskomplexes aus Promotor-DNA, TBP und TFB, an den der RNA-Polymerase-Komplex erst nachträglich ankoppelt, während die RNA-Polymerase bei *Bacteria* erst nach dem Binden des σ -Faktors an den Promotor bindet. Allerdings zeigt sich auch hier wieder die bei *Eukarya* sowohl gegenüber *Bacteria* als auch *Archaea* "kompliziertere" Situation: Während bei *Archaea* in vielen Fällen wirklich allein mit den 4 Komponenten (linearisierte, die Promotorregion enthaltende) DNA, TBP (aTBP = archaelles TATA-Box-Bindeprotein), TFB und der gereinigten RNA-Polymerase spezifische Transkription beobachtet werden kann (Hausner *et al.*, 1996; Qureshi *et al.*, 1997), ist dies bei *Eukarya* nur in Ausnahmefällen möglich. Pol II benötigt in der Regel ein Vielzahl von Transkriptionsfaktoren (mindestens 7, je nach Zählung) (Flores *et al.*, 1992; Orphanides *et al.*, 1996) und assoziierte Faktoren. So kommt beispielsweise TBP nicht allein, sondern integriert in einen Komplex aus mehr als 12 TAFs (TBP assoziierte Faktoren) vor. Unter anderem über diese Vielzahl beteiligter Proteine ergibt sich ein Netzwerk aus Wechselwirkungen und zusätzlichen (vermutlich teilweise redundanten) Regulationsmöglichkeiten, die in dieser Form bei *Archaea* nicht zu finden sind (Thomm, 2000).

Grundsätzlich ist der Ablauf der Initiation der Transkription bei *Eukarya* also etwa folgender: TFIID (so wird der Komplex aus TBP und TAFs genannt) bindet an den Promotor, wobei es der DNA sattelförmig aufsitzt und in dieser eine charakteristische Biegung verursacht. Anschließend tritt TFIIB (also TFB bei der Pol II) hinzu, wobei die Bindung von TFIIB als Coaktivator stabilisiert wird (Zawel & Reinberg, 1993). An diesen Präinitiationskomplex kann nun die, unabhängig vom bis hier geschilderten Ablauf zusätzlich mit TFIIE, TFIIF und TFIIH beladene Pol II ankoppeln und den Initiationskomplex bilden. Hierbei handelt es sich dann um die so genannte "aktivierte Transkription" im Gegensatz zu der, oben angedeuteten "basalen Transkription".

Nach erfolgreichem Start der Transkription können die am Promotor verbliebenen Faktoren TFIID und TFIIB entweder abdiffundieren, oder aber für weitere Transkriptionszyklen gebunden bleiben (das ist zumindest für TFIID aus energetischen Gründen wahrscheinlich (Nikolov & Burley, 1997)). Dass die mit dem beschriebenen Prozess verbundenen vielfältigen Protein-Protein-Wechselwirkungen nicht nach einem starren "Baukasten-Prinzip" ablaufen, sondern mit einer Kaskade von induzierten Konformationsänderungen einhergehen, zeigen z.B. Untersuchungen an TFIIB, welches erst durch Bindung zusätzlicher Aktivatoren und/oder Bindung an TFIID in die Lage versetzt wird, an die Pol II zu binden, weil sonst durch eine intramolekulare Bindung des N-Terminus an eine C-terminale Region von TFIIB die benötigten Binderegionen blockiert sind (Roberts & Green, 1994). Hierbei scheint auch der über TFIIF vermittelte Kontakt zwischen TFIIB und Pol II eine wichtige Rolle zu spielen (was auch ganz aktuelle Forschungsdaten belegen (Armache *et al.*, 2003)).

Transkription bei *Archaea*:

Bezüglich der RNA-Polymerase bei *Archaea* ist bekannt, dass sie zwar auch homolog zur bakteriellen RNA-Pol ist, jedoch insgesamt die meisten Ähnlichkeiten mit Pol II aus *Eukarya* aufweist (Langer *et al.*, 1995; Thomm, 2000; Bartlett *et al.*, 2003). Wie bei *Bacteria* ist sie sowohl für die Synthese der mRNA der Proteingene als auch für die der verschiedenen anderen RNA-Formen zuständig. Neben vieler Strukturähnlichkeiten der archaellen und eukaryotischen RNA-Pol (Pol II) ist vor allem die ähnliche Struktur und Funktion der beiden essentiellen archaellen Transkriptionsfaktoren TBP und TFB ein Hinweis darauf, dass auch der grundlegende

Mechanismus der Initiation der Transkription zwischen *Archaea* und *Eukarya* konserviert ist (Littlefield *et al.*, 1999; Tsai & Sigler, 2000; Bartlett *et al.*, 2003). In vielen Details scheint sich zu bestätigen, dass der vereinfachte Transkriptionsapparat der *Archaea* wahrscheinlich dem des gemeinsamen Vorläufers von *Archaea* und *Eukarya* ähnelt. (Kyrpides & Ouzounis, 1999; Thomm, 2000).

Ähnlich wie bei *Eukarya* bindet bei der Initiation der Transkription zuerst TBP spezifisch an die TATA-Box. Anschließend bindet TFB an TBP und über ein Helix-Turn-Helix-Motiv an das direkt stromaufwärts vor der TATA-Box befindliche "B-recognition element" (BRE), was entscheidend für die Orientierung des anschließend durch die Anlagerung der RNA-Polymerase komplettierten Initiationskomplexes ist. Auch hier werden TBP und TFB nicht mehr benötigt, nachdem der Transkriptionskomplex erfolgreich in die Elongationsphase übergegangen ist.

Aktuelle Forschung zum Thema "archaeelle Transkription" und "RNA-Polymerase":

Ein Schwerpunkt der Forschung sind Untersuchungen zur Regulation der Transkription, die bei *Archaea* im Gegensatz zum sonstigen Erscheinungsbild eher bakteriell als eukaryotisch geprägt scheint (obwohl es einzelne Gegenbeispiele gibt (Soppa, 1999)). Regulatorproteine vom bakteriellen Typ sind z.B. LrpA ("Leucine-responsive regulatory protein A") und der kürzlich charakterisierte Transkriptionsregulator Phr aus *P.furiosus* (Vierke *et al.*, 2003), der eine Rolle bei der Hitzeschock-Antwort spielt. Ptr1 und Ptr2 aus *M.jannaschii* sind weitere Beispiele für potentielle "bakterielle" Regulatorproteine (Transkriptionsaktivatoren) bei *Archaea* (Ouhammouch *et al.*, 2003).

Nach wie vor im Zentrum des Interesses steht das detaillierte Bindeverhalten der basalen Transkriptionsfaktoren TBP und TFB aneinander, an die verschiedenen Promotorelemente auf der DNA und an die RNA-Polymerase. Interaktionen können hier z.B. auch mit den oben genannten Regulatorproteinen oder zusätzlichen (nicht essentiellen) Faktoren wie Tip49 ("TBP interacting protein") auftreten. TFE (TFII Ea) wurde als weiterer, vielleicht stärker spezies-spezifischer basaler Transkriptionsfaktor vorgeschlagen (Soppa, 1999; Hanzelka *et al.*, 2001), wobei sich eine (insgesamt "aktivierende"), die Bindung von TBP und TFB an die TATA-Box stabilisierende Funktion herauszukristallisieren scheint (Bell *et al.*, 2001; Voss, 2003).

Aktuelle Ergebnisse zeigen, dass im Bereich der an der archaeellen Transkription beteiligten Faktoren noch mit Überraschungen zu rechnen ist: Erst vor kurzem wurde bei *P.furiosus* ein neues, unbekanntes Protein von ähnlichem Molekulargewicht wie TBP im stromaufwärts gelegenen Abschnitt der "Transkriptionsblase" entdeckt, das wahrscheinlich eine Funktion bei der Initiation der Transkription hat (Bartlett *et al.*, 2003).

Die Identifizierung und allgemeine Charakterisierung der Untereinheiten der verschiedenen RNAPs hat vor allem durch die Erstellung und die Analyse von 3D-Strukturen der gegenüber *Archaea* recht ähnlichen eukaryotischen Pol II auf der Basis von Röntgenstrukturanalyse (Cramer *et al.*, 2000; Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Cramer, 2002; Armache *et al.*, 2003), sowie durch die Rekonstitution der RNAP aus rekombinanten Untereinheiten und eine Kristallstruktur eines rpoE'-rpoF-Komplexes bei *Methanococcus jannaschii* große Fortschritte gemacht (Eloranta *et al.*, 1998; Werner *et al.*, 2000; Todone *et al.*, 2001; Werner & Weinzierl, 2002).

Für rpoE2 ist noch unklar, ob dieses Protein als Untereinheit im nativen RNAP-Komplex enthalten ist. Wenn ja, wäre eine Position im Umfeld von rpoE1 zu erwarten, da die beiden Untereinheiten in *Sulfolobus acidocaldarius* eine Einheit bilden.

Schließlich werden gegenwärtig Studien zur Aufdeckung des genauen Weges der DNA durch die große Spalte des RNAP-Komplexes vorbei am katalytischen Zentrum unternommen, wobei die spezifischen DNA-Protein-Kontakte zu bestimmten Untereinheiten eine wichtige Rolle für einen regelmäßigen und möglichst fehlerlosen Ablauf der Transkription spielen (Bartlett *et al.*, 2003). Diese Daten sind wichtig für die lückenlose Aufklärung des Mechanismus der Transkription.

Ein kritischer Vorgang ist der Übergang der Transkription von der Initiationsphase mit "open complex"-Bildung (Hausner & Thomm, 2001) hinüber in die Elongationsphase, für die insbesondere bei *Eukarya* eine starke Konformationsänderung gezeigt wurde, die unter anderem darin besteht, dass die "Klammer" (clamp) auf der einen Seite der quer durch den RNAP-Komplex

verlaufenden Spalte die am Grund der Spalte entlang laufende DNA bzw. das DNA-RNA-Hybrid teilweise einschließt, und somit den Komplex stabilisiert (Gnatt *et al.*, 2001).

Die Elongationsphase und hiermit verbunden die Korrekturlesefunktionen (mit nötigem "Backtracking", also einem teilweisen Zurücklaufen beim Auftreten von Fehlern), das Einlegen von Pausen, und schließlich die Termination der Transkription stellen weitere entscheidende Prozesse in der archaeellen Transkription dar (Hausner *et al.*, 2000; Lange, 2003), wobei der genaue Weg der RNA durch den RNAP-Komplex und ihr Schicksal zum Beispiel auch bei abortiven Prozessen von Interesse ist.

2.3 Ziel der Arbeit:

Das Thema dieser Arbeit fügt sich gut in einen der Forschungsschwerpunkte der Arbeitsgruppe von Prof.Dr.M. Thomm ein, nämlich die detaillierte Erforschung von Komponenten, Ablauf und Regulation der archaeellen Transkription an dem hyperthermophilen, archaeellen Modellorganismus *Pyrococcus furiosus* (Dahlke & Thomm, 2002; Bartlett *et al.*, 2003; Spitalny & Thomm, 2003; Vierke *et al.*, 2003).

Bisher kam die RNA-Polymerase als eine der essentiellen Komponenten eines zellfreien Transkriptionssystems (Hethke *et al.*, 1999) ausschließlich in nativer Form direkt aus *Pyrococcus* gereinigt zum Einsatz. Da es sich bei der RNA-Polymerase um einen dynamischen Komplex aus ca. 12 Untereinheiten handelt (die je nach stoffwechselphysiologischer Situation durchaus nicht immer alle im gleichen stöchiometrischen Verhältnis vorliegen) liegt es nahe, auch die Eigenschaften der einzelnen Untereinheiten und ihre Wechselwirkungen mit anderen Untereinheiten, Transkriptionsfaktoren und DNA/RNA genauer zu studieren. Außerdem ergeben sich durch die Verfügbarkeit der RNA-Polymerase-Untereinheiten in rekombianter Form und großer Menge / Reinheit neue Anwendungsmöglichkeiten, bis hin zu einer kompletten Rekonstitution des Gesamtkomplexes inklusive nachweisbarer Enzymaktivität, wie dies mittlerweile für *Methanocaldococcus jannaschii* gezeigt wurde (Werner & Weinzierl, 2002).

In dieser Arbeit wird durch die Identifizierung, Klonierung, Expression und Reinigung der meisten Untereinheiten der RNA-Polymerase die Basis für die (bereits jetzt erfolgende) Fortführung des Großprojektes "Rekonstitution der RNA-Polymerase bei *Pyrococcus furiosus*" gelegt.

Neben der Lösung vieler hiermit zusammenhängender praktischer Detailprobleme werden mit verschiedenen Methoden Protein-Protein-Wechselwirkungen der bereits verfügbaren Untereinheiten untersucht, was Informationen über die allgemeine Struktur des archaeellen RNA-Pol-Komplexes liefert. Eine zentrale Stellung nimmt hierbei die Methode des "Far Western Blottings" ein, bei der statt eines Antikörpers direkt das (radioaktiv markierte) Sonden-Protein mit dem Zielprotein interagiert, und so auf die Stärke der Wechselwirkung geschlossen werden kann.

Neben den praktischen Experimenten wird durch bioinformatische Analysen aller RNA-Polymerase-Untereinheiten auf Basis der aktuellen Daten der bis jetzt verfügbaren, komplett sequenzierten archaeellen Genome eine Gesamtübersicht erarbeitet. Mit den umfangreichen Alignments der einzelnen RNA-Pol-Gene werden Informationen über im Rahmen der Evolution besonders konservierte bzw. besonders stark abweichende Sequenz- bzw. Proteinabschnitte gewonnen. Diese Daten werden einerseits auf die besondere Situation bei *Pyrococcus furiosus* bezogen, andererseits aber auch ausführlich gegenüber der in vieler Hinsicht sehr ähnlichen RNA-Polymerase II (Pol II) aus *Saccharomyces cerevisiae* diskutiert.

Letzteres erwies sich als besonders ergiebig, da für die Pol II schon mehrere, aus Röntgenstruktur-Daten abgeleitete 3D-Strukturen verfügbar sind (Cramer *et al.*, 2000; Cramer *et al.*, 2001; Gnatt *et al.*, 2001; Armache *et al.*, 2003). Mit Hilfe geeigneter bioinformatischer Programme war es möglich, die in den Untereinheiten der eukaryotischen RNA-Polymerase gegenüber *Archaea* konservierten Regionen in der 3D-Struktur zu visualisieren. Die bei *Eukarya* identifizierten Struktur- und Funktionseinheiten innerhalb der RNA-Pol-Untereinheiten (Cramer *et al.*, 2001) werden im Detail in Beziehung mit dem Konsensus aus den oben genannten archaeellen Alignments gesetzt [8.1.3]. Die durch die Informationsfülle etwas unhandlichen Daten werden mittels selbst erstellter 3D-Grafiken und in selbst programmierten Diagrammen besser zugänglich gemacht.

Die resultierenden, beispielsweise durch das Fehlen oder die starke Abweichung bestimmter Strukturmodule bei *Archaea* erkennbaren Unterschiede gegenüber *Eukarya*, und sich hierraus ergebende Konsequenzen und neue Fragestellungen werden diskutiert.

2.4 Anmerkungen zur Gliederung der Arbeit:

Diese Arbeit ist grundsätzlich in einen theoretischen und einen praktischen Abschnitt gegliedert: [4.1] bis [4.8] behandeln die Eigenschaften der archaeellen und eukaryotischen RNA-Polymerase-Gene und ihr Vorkommen im Genom, wobei die umfangreichen Alignments im Anhang [8.1] eine wichtige Rolle spielen. Eine für *Saccharomyces cerevisiae* vorgenommenen, detaillierte Aufteilung der einzelnen RNAP-Untereinheiten in "Strukturmodule" (Cramer *et al.*, 2001) wurde auf *Pyrococcus* beziehungsweise auf *Archaea* übertragen. Auf diese strukturellen und funktionalen Regionen, die auch eine gute Basis für die Behandlung der Details der archaeellen RNAP darstellen, wird in folgenden Abschnitten eingegangen:

[4.3.1] Eigenschaften und konservierte Regionen der RNAP-Gene in *Pyrococcus*

[4.6] Zusammenfassende grafische Darstellung von homologen Genabschnitten

[8.1.3] Alignments mit *Archaea* und *Eukarya*.

Diese Kapitel bilden von der Aussage her eine Einheit, weshalb zum Beispiel die genannten Regionen auch in den Alignments unter [8.1.3] farbig markiert wurden. Da im Rahmen dieser Arbeit nicht alle Details erschöpfend diskutiert werden können, versteht sich die (mit einer großen Zahl von Grafiken versehene) Darstellung auch als Bestandsaufnahme, die durch die ausführlich aufbereiteten Daten eine Referenz für zukünftige Analysen darstellt.

Kapitel [4.6.1] ("3D-Struktur der RNA-Pol II mit gegenüber *Archaea* konservierten Regionen") liefert durch die Übertragung der zwischen *Eukarya* und *Archaea* bestehenden Sequenzähnlichkeiten der RNAP-Untereinheiten auf eine zusammenfassende 3D-Darstellung eine der Kernaussagen des theoretischen Abschnitts.

In den Kapiteln [4.9] bis [4.12] werden anschließend die in dieser Arbeit durchgeführten praktischen Experimente vorgestellt und diskutiert. [4.11.1.3] ("Zusammenfassung der Far Western Blotting-Ergebnisse und Diskussion") nimmt noch einmal mit vielen Grafiken ausgiebig Bezug auf die 3D-Strukturen und die erwähnten Strukturmodule. Durch den Bezug der Ergebnisse der Far Western Blotting-Experimente auf die im theoretischen Teil erarbeiteten 3D-Strukturdaten findet sich hier die Kernaussage des praktischen Teils und dieser Arbeit insgesamt.

Für das zügige Auffinden der vielen im Text zitierten Grafiken wurde unter [8.8] ein Abbildungsverzeichnis angelegt, welches sich auch auf Tabellen bezieht.

3 Material und Methoden

3.1 Bakterienstämme

Escherichia coli, Stamm JM109 (für "high copy"-Plasmid-Präparation)

Escherichia coli, (Expressions-)Stamm BL21(DE3)

mit den zusätzlichen Plasmiden pLys S / pLys E (Novagen/Merck)

pUBS520 (Brinkmann *et al.*, 1989)

Escherichia coli + "Codon Plus"-Plasmid mit zusätzlichen tRNAs für in *E. coli* seltene Codons:

Epicurian Coli[®] BL21-CodonPlus[™] (DE3)-RIL Competent Cells, #230245 (Stratagene)

Escherichia coli mit reduzierter RNase-Aktivität für langsameren Abbau der mRNA:

BL21 Star[™] (DE3) One Shot[®] Chemically Competent Cells, Cat.No. C6010-03 (Invitrogen)

(in der Praxis wurden durch die Transformation des Codon Plus-Plasmids in BL21 Star (DE3) die Eigenschaften der beiden letztgenannten Stämme vereinigt, um eine optimale Proteinmenge bei der Expression zu erreichen)

Pyrococcus furiosus DSM 3638 (für die Reinigung der nativen RNA-Polymerase)

3.2 Verwendete Plasmide

Die verwendeten Plasmide wurden von der Firma Novagen / Merck bezogen:

(<http://www.merckbiosciences.de/g.asp?f=NVG/pETtable.html>)

pET-17b (ohne Fusionsanteil)

pET-14b (mit 20 Aminosäuren langem N-terminalen Fusionsanteil inklusive His₆-Tag)

pET-15b (s.oben)

pET-33b(+) (mit 25 Aminosäuren langem N-terminalen Fusionsanteil mit His₆-Tag und PKA Site)

Hiervon abgeleitete, selbst modifizierte Varianten:

pET-33b(+)c2 (wie pET-33b(+), jedoch C-terminaler Fusionsanteil)

pET-17bHis (wie pET-17b, jedoch mit der "multiple cloning site" von pET-14b → mehr Restriktionsschnittstellen)

3.3 Chemikalien, Enzyme und Materialien

Die verwendeten Chemikalien und Enzyme waren, soweit nicht gesondert angegeben, die über den Fachhandel zu beziehenden, den aktuellen Standards entsprechenden Präparationen. In der Regel wurden Chargen der Reinheitstufe "PA" (zur Analyse) verwendet.

Radioaktivität für Western Blotting-Experimente:

[γ -³²P]-ATP (SCP-501; 6000 μ Ci/mmol): Hartmann Analytic GmbH, Braunschweig

Radioaktivität für TCA-Test bei der Reinigung der nativen RNA-Polymerase:

[α -³²P]-UTP (SCP-210): Hartmann Analytic GmbH, Braunschweig

Expressionsvektor mit PKA-Markierungsregion für Western Blotting-Experimente:

pET-33b(+): Novagen/Merck Cat.No. 69054-3

Proteinkinase A für Markierungsreaktionen:

PKAce[™] Kit von Novagen/Merck Cat.No. 70510-3

Weitere Quelle für Proteinkinase A (preisgünstig):

Sigma: Protein Kinase Catalytic Subunit (From bovine heart), Prod.No. P2645

→ in lyophilisierter Form; rücklösen mit folgender Lösung:

Ansetzen von 1ml Lösung zum Resuspendieren:

Das lyophilisierte Enzym enthält schon 19% Phosphat-Puffer, 80% Saccharose;

deshalb nur Zugabe von: 461 μ l H₂O + 500 μ l 100%iges Glycerin + 39 μ l 1M DTT

→ hiervon 100 μ l, um auf eine PKA-Endkonzentration von 10U/ μ l zu kommen

Wichtig: wegen hohem Glyceringehalt wie ein Restriktionsenzym behandeln (Aufbewahrung bei -20°C und nicht bei -70°C → Gefahr der Inaktivierung!)

10x PKA-Puffer: Entweder aus PKAce™Kit (siehe oben); oder:
→ 200mM Tris-HCl pH 8.0, 1,5M NaCl, 200mM MgCl₂

Ponceau S-Färbung zum Anfärben von Proteinen im Rahmen von Far Western Blotting – Experimenten: Sigma, Product No. P7767 (Fertiglösung)

Farbsubstrat für die Enzymreaktion beim Western Blotting:
BM blue POD Substrat, präzipitierend (Best.Nr. 1 442 066, Roche)

3.4 Klonierung der Untereinheiten der RNAP und Liste der verwendeten Primer

Diese Arbeit stellt den erste Abschnitt eines umfassenderen Projektes dar, dem man den Titel "Untersuchungen zu und Rekonstitution der RNA-Polymerase aus den rekombinanten Untereinheiten aus *P.furiosus*" geben könnte.

Aus diesem Grunde mußten sehr viele Gene kloniert werden, so dass die folgende Primer-Liste recht umfangreich ist. Die Angaben beschränken sich hauptsächlich auf die Primer, deren Schnittstellen und weitere Eigenschaften, sowie eine kurze Angabe über Art des resultierenden Proteins und eventuelle Besonderheiten.

Die verschiedenen Schritte von der Auswahl geeigneter Primer, der PCR-Reaktion, die Reinigung der PCR-Produkte und anschließende Restriktionshydrolyse, die Ligation in das gewünschte Plasmid und die Transformation in *E.coli JM109* (beziehungsweise nach Überprüfung der Fehlerfreiheit der DNA durch Sequenzierung die Transformation in einen *E.coli*-Expressionsstamm) sind die heute üblichen molekularbiologischen Techniken, und wurden entsprechend den Angaben des jeweiligen Herstellers der verwendeten "Kits" durchgeführt.

Da es sich zum Teil um sehr lange DNA-Sequenzen handelte, wurde bei der PCR auf DNA-Polymerase zurückgegriffen, die eine hohe Korrekturlesefunktion mit guter (oder noch ausreichender) Prozessivität verbindet:

- 1.) PLATINUM[®] Pfx DNA Polymerase von Life Technologies™ / GibcoBRL[®] (Cat.No. 11708-013). Diese Polymerase ist bei normaler Temperatur durch einen Antikörper blockiert, und wird beim ersten Denaturierungsschritt aktiviert, so dass kein "Hot start" nötig ist.
- 2.) Expand™ High Fidelity PCR System, Cat. No. 1 732 641 (Roche)
- 3.) Taq-DNA-Polymerase von verschiedenen Herstellern (ausreichend bei kurzen DNA-Sequenzen)

Die Ligation der hydrolysierten PCR-Fragmente in die ebenfalls geschnittenen Plasmide wurde mit Hilfe der "T4-DNA-Ligase" Best.Nr. 481 220 (Roche) durchgeführt.

Für die Extraktion von DNA-Fragmenten aus Agarose-Gelen kam überwiegend das "DNA Extraction Kit #K0513" der Firma MBI Fermentas zum Einsatz.

Für die Transformation der Ligationsansätze in *E.coli JM109* wurde entweder die übliche CaCl₂-Methode angewendet, oder auf das "TransformAid™ Bacterial Transformation Kit" #K2711 der Firma MBI Fermentas zurückgegriffen.

Eingesetzte Primer und Plasmide:

Anmerkung: Die Primer wurden den üblichen Empfehlungen nach so gewählt, dass die Schmelzpunkte ähnlich, keine Komplementarität gegeben, und die Enden möglichst durch C oder G gebildet wurden. Bei den Expressionsprimern wurde eine zusätzliche, nicht komplementäre Sequenz mit der passenden Restriktionsschnittstelle angefügt, wobei am Ende je nach Enzym weitere Nukleotide angefügt wurden (Angaben hierzu z.B. im Katalog der Firma "New England Biolabs" unter der Überschrift: "Cleavage Close to the End of DNA Fragments").

Untereinheit B (rpoB) der RNAP aus *P.furiosus* (Gen: PF1564):

- Untereinheit B ist die einzige Untereinheit, die im Rahmen dieser Arbeit nicht kloniert wurde. Die geplanten Expressionsprimer wurden noch einmal modifiziert (persönliche Mitteilung, Naji, S.), so dass hier nur die zusätzlichen Sequenzierprimer aufgeführt werden:

pf_rpoB_sq1: 5'-CTCATGGTCAAAGAAAGCCCC-3' (21-mer, $T_m = 59,8^\circ\text{C}$)
pf_rpoB_sq2: 5'-GGGCTGGAGAGAGATAAGGA-3' (20-mer, $T_m = 59,4^\circ\text{C}$)
pf_rpoB_sq3: 5'-ACTCCATAAGGCCAGACGTTCT-3' (22-mer, $T_m = 60,3^\circ\text{C}$)
pf_rpoB_sq4: 5'-TTAGGCAAGACAGAAGAAGTGG-3' (22-mer, $T_m = 58,4^\circ\text{C}$)
pf_rpoB_sq5: 5'-TGGGCAAACCTTTAGAATTAGGG-3' (22-mer, $T_m = 56,5^\circ\text{C}$)
pf_rpoB_sq6: 5'-AGTAATAATAACCGAGACGGG-3' (21-mer, $T_m = 55,9^\circ\text{C}$)
pf_rpoB_sq7: 5'-TCAGAGGCTCCACCACATGG-3' (20-mer, $T_m = 61,4^\circ\text{C}$)

Untereinheit A' (rpoA1) der RNAP aus *P.furiosus* (Gen: PF1563):

- Die Klonierung erfolgte in das Plasmid pET-14b

- Anmerkung: die Analyse der Genomsequenz hat ergeben, dass rpoA1 drei Aminosäurepositionen früher als in PF1563 angegeben startet; die Primer wurden entsprechend gewählt, so dass das klonierte DNA-Fragment 2730 Nukleotide (ohne Stop-Codon), beziehungsweise das Genprodukt 910 Aminosäuren lang ist (statt 907aa).

Expressionsprimer:

PF_RPOA1b_FEX: 5'-CGAATTCCATATGCAATCCATGAAAAAAG-3' (**Nde I**, 29-mer, $T_m = 61,0^\circ\text{C}$)
PF_RPOA1b_REX: 5'-CGGGATCCTCACACCTTCGCCTTGTTATTTCC-3' (**Bam HI** 32-mer, $T_m = 70,8^\circ\text{C}$)

Sequenzierprimer:

PF_RPOA1_SQ1: 5'-GAAAAACCAACAATCTACTGGG-3' (22-mer, $T_m = 56,5^\circ\text{C}$)
PF_RPOA1_SQ2: 5'-CAATGAAGCTCCAGGAGTTCC-3' (21-mer, $T_m = 59,8^\circ\text{C}$)
PF_RPOA1_SQ3: 5'-CTCACAGAGTAAGGGTCATGC-3' (21-mer, $T_m = 59,8^\circ\text{C}$)
PF_RPOA1_SQ4: 5'-GTTAATGCCAGATCCGGAGG-3' (20-mer, $T_m = 59,4^\circ\text{C}$)
PF_RPOA1_SQ5: 5'-GTGCCGTTGCAGAGAAGTACC-3' (21-mer, $T_m = 61,8^\circ\text{C}$)

Untereinheit A'' (rpoA1) der RNAP aus *P.furiosus* (Gen: PF1562):

- Die Klonierung erfolgte in das Plasmid pET-14b

Expressionsprimer:

PF_RPOA2_FEX: 5'-GGAATTCCATATGGTCTCTCTTTCTACTAT-3' (**Nde I**, 30-mer, $T_m = 62,7^\circ\text{C}$)
PF_RPOA2_REX: 5'-CGGGATCCTCACACCTCCTCTTTCTTTTC-3' (**Bam HI**, 30-mer, $T_m = 69,5^\circ\text{C}$)

Sequenzierprimer:

PF_RPOA2_SQ1: 5'-CCATCAACCCCATGATGAC-3' (20-mer, $T_m = 59,4^\circ\text{C}$)
PF_RPOA2_SQ2: 5'-AGCAAGTCTTAAAGTCCCG-3' (20-mer, $T_m = 55,3^\circ\text{C}$)

Untereinheit D (rpoD) der RNAP aus *P.furiosus* (Gen: PF1647):

- Die Klonierung erfolgte in das Plasmid pET-14b, pET-33b(+), pET-33b(+)+c2

Expressionsprimer:

PF_RPOD_FEX: 5'-GGGACACCATATGGCCGGAATTGAAGTTCAG-3' (**Nde I**, 31-mer, $T_m = 69,5^\circ\text{C}$)
PF_RPOD_REX: 5'-CGGGATCCTCAAGAGGTCAATTTTTGAAG-3' (**Bam HI**, 29-mer, $T_m = 65,3^\circ\text{C}$)
(mit Stop-Codon, für Klonierung in pET-14b und pET-33b(+))
PF_RPOD_REX2: 5'-CGGGATCCAGAGGTCAATTTTTGAAGTTC-3' (**Bam HI**, 29-mer, $T_m = 65,3^\circ\text{C}$)
(ohne Stop-Codon, für Klonierung in pET-33b(+)+c2)

Untereinheit E' (rpoE1) der RNAP aus *P.furiosus* (Gen: PF0256):

PF_RPOE_FEX: 5'-GGAATTCCATATGTACAAGATAGTCACCGTAAAG-3' (**Nde I**, 34-mer, $T_m = 65,9^\circ\text{C}$)
PF_RPOE_REX: 5'-CGGGATCCTCACTTCTTACCCTCCTCTTTTC-3' (**Bam HI**, 32-mer, $T_m = 70,8^\circ\text{C}$)
(mit Stop-Codon, für Klonierung in pET-14b und pET-33b(+))
PF_RPOE_REX2: 5'-CGGGATCCCTTCTTACCCTCCTCTTTTC-3' (**Bam HI**, 29-mer, $T_m = 69,5^\circ\text{C}$)
(ohne Stop-Codon, für Klonierung in pET-33b(+)+c2)

Untereinheit E'' (rpoE2) der RNAP aus *P.furiosus* (Gen: PF0255):

PF_RPOE2_FEX: 5'-GGAATTCCATATGAGTGAAAAAGCCTGCAGACAC-3' (**Nde I**, 34-mer, $T_m = 68,3^\circ\text{C}$)
PF_RPOE2_REX: 5'-CGGGATCCTCAGCGCACCCCTTATGGCAT-3' (**Bam HI**, 28-mer, $T_m = 71,0^\circ\text{C}$)

Untereinheit F (rpoF) der RNAP aus *P.furiosus* (Gen: PF1036):

PF_RPOF_FEX: 5'-GGAATTCATATGATAGGGAGAGGGCAAACCTCG-3' (**Nde I**, 32-mer, $T_m = 68,2^\circ\text{C}$)

PF_RPOF_REX: 5'-CGGGATCCTCACTCTCTATACTTGTCAAGG-3' (**Bam HI**, 30-mer, $T_m = 68,1^\circ\text{C}$)

PF_RPOF_REX2: 5'-GCGGATCCCTCTCTATACTTGTCAAGGA-3' (**Bam HI**, 28-mer, $T_m = 66,6^\circ\text{C}$)

(ohne Stop-Codon, siehe oben)

Untereinheit H (rpoH) der RNAP aus *P.furiosus* (Gen: PF1565):

PF_RPOH_FEX: 5'-GCTATAGCATATGGCGGGGAAAAAGGAATTTAGC-3' (**Nde I**, 34-mer, $T_m = 68,3^\circ\text{C}$)

PF_RPOH_REX2: 5'-GCGGATCCTTAGTCTTCAACAACAACCC-3' (**Bam HI**, 28-mer, $T_m = 66,6^\circ\text{C}$)

PF_RPOH_REX3: 5'-GCGGATCCGTCTTCAACAACAACCCTATAG-3' (**Bam HI**, 30-mer, $T_m = 68,1^\circ\text{C}$)

(ohne Stop-Codon, siehe oben)

Untereinheit K (rpoK) der RNAP aus *P.furiosus* (Gen: PF1642):

PF_RPOK_FEX2: 5'-GGAATTCATATGTTCAAGTATACGAGG-3' (**Nde I**, 28-mer, $T_m = 62,2^\circ\text{C}$)

PF_RPOK_REX: 5'-CGGGATCCTCAGCTCGGTCTGATTACTG-3' (**Bam HI**, 28-mer, $T_m = 69,5^\circ\text{C}$)

PF_RPOK_FEX: 5'-CTAGCTAGCATGTTCAAGTATACGAGG-3' (**Nhe I**, 27-mer, $T_m = 63,4^\circ\text{C}$)

Untereinheit L (rpoL) der RNAP aus *P.furiosus* (Gen: PF0050):

PF_RPOL_FEX: 5'-GGAATTCATATGAAGATAGAAGTGATAAAGA-3' (**Nde I**, 32-mer, $T_m = 61,8^\circ\text{C}$)

PF_RPOL_REX: 5'-TTGGATCCTCAGCTCTTACC GCCTTCTCC-3' (**Bam HI**, 30-mer, $T_m = 70,9^\circ\text{C}$)

Untereinheit N (rpoN) der RNAP aus *P.furiosus* (Gen: PF1643):

PF_RPON_FEX: 5'-GGAATTCATATGGGGGCAAGTCCCGTGATTG-3' (**Nde I**, 32-mer, $T_m = 70,8^\circ\text{C}$)

PF_RPON_REX: 5'-CGGGATCCTCAATACACTCTGTAATGCA-3' (**Bam HI**, 28-mer, $T_m = 65,1^\circ\text{C}$)

Untereinheit P (rpoP) der RNAP aus *P.furiosus* (Gen: PF2009):

PF_RPOP_FEX: 5'-GGAATTCATATGGTTGAGGCCGTGTATAG-3' (**Nde I**, 30-mer, $T_m = 66,8^\circ\text{C}$)

PF_RPOP_FEX: 5'-CGGGATCCTAAATGGCCTTACCCTTC-3' (**Bam HI**, 27-mer, $T_m = 66,5^\circ\text{C}$)

TFB und TFB-2 aus *P.furiosus* (Gene: PF1377 und PF0687):

- Im Genom von *Pyrococcus furiosus* existiert eine zweite Kopie von TFB (siehe [4.8]), wobei allerdings noch nicht experimentell nachgewiesen wurde, dass die kodierte Proteinsequenz wirklich *in vivo* exprimiert wird. Ich habe zunächst den Namen TFB2C (2C für "second copy") verwendet. Neuerdings scheint sich aber die Bezeichnung TFB-2 (oder sogar TFB1 und TFB2 durchzusetzen, was aber leicht mit der Bezeichnung TFIIB für den homologen Transkriptionsfaktor bei *Eukarya* verwechselt werden kann).

TFB (TFB-1):

PWTF2B1: 5'-GGAATTCATATGAATAAGCAAAGGTTTGTGTC-3' (**Nde I**)

PF_TFBb_REX: 5'-CGGGATCCTCATGCTATAGGAACCTTAATC-3' (**Bam HI**, 30-mer, $T_m = 65,4^\circ\text{C}$)

PF_TFBb_REX2: 5'-CGGGATCCTGCTATAGGAACCTTAATCTTC-3' (**Bam HI**, 30-mer, $T_m = 65,4^\circ\text{C}$)

(ohne Stop-Codon, siehe oben)

TFB (TFB-2):

PF_TFB_C2FEX: 5'-GGAATTCATATGTCATCTACTGAGCCCGGAG-3' (**Nde I**, 32-mer, $T_m = 69,5^\circ\text{C}$)

PF_TFB_2CREX: 5'-GGAATTCCTTACCACCACAAGCTCTTTAAC-3' (**Eco RI**, 30-mer, $T_m = 65,4^\circ\text{C}$)

PF_TFB2Cb_REX: 5'-CGGGATCCTCACCACCACAAGCTCTTTAAC-3' (**Bam HI**, 31-mer, $T_m = 70,8^\circ\text{C}$)

3.5 Herstellung kompetenter Zellen für die Transformation, CaCl₂-Methode

Für einen Teil der im Rahmen dieser Doktorarbeit vorgenommenen Transformationen wurde das „TransformAid, Bacterial Transformation Kit #K2711“ von MBI Fermentas verwendet. Dies betrifft vor allem die Transformation der sequenzierten und fehlerfreien Plasmide in den jeweiligen Expressionsstamm. Für den primären Schritt der Transformation von Ligationsansätzen in *E.coli* JM109 wurde dagegen meist auf selbst hergestellte CaCl₂-kompetente JM109-Zellen zurückgegriffen, die nach folgendem Protokoll präpariert wurden: (siehe auch: MBI Fermentas) Animpfen von 5ml LB₀-Flüssigmedium mit Einzelkolonie (+ evtl. Antibiotikum); Inkubation über Nacht bei 37°C. Mit 4-5ml dieser Kultur ein 400ml Medium (im 2 Liter-Kolben) beimpfen und bis ca. OD₅₇₃ = 0.6 wachsen lassen. 20 min auf Eis; 15 min bei 5000rpm bei 4°C zentrifugieren; Überstand entfernen; in 200ml sterilem, eiskaltem TBS aufnehmen; (Hälfte des Kulturvolumens;

0.1 M NaCl, 5 mM Tris, 5 mM MgCl₂, pH 7,0); 10 min bei 5000 rpm bei 4°C zentrifugieren; Überstand entfernen; in 200 ml steriler, eiskalter CaCl₂-Lösung aufnehmen (0.1 M CaCl₂, 5 mM Tris, 5 mM MgCl₂, pH 7,0); 20 min auf Eis; 10 min bei 5000 rpm und 4°C zentrifugieren; Überstand entfernen; in 20 ml eiskalter CaCl₂-Lsg. resuspendieren; 1h auf Eis; Hinzufügen von 5ml Glycerin (86%, steril; Endkonz. 15-20%); in E-Caps portionieren und mit flüssigem Stickstoff im 4°C-Raum schockgefrieren. Lagerung bei -70°C.

3.6 Far Western Blotting zur Analyse von Protein-Protein-Interaktionen

Wie der Name schon andeutet, handelt es sich bei dem „Far Western Blotting“ um eine Technik, die von ihrem Mechanismus her stark dem klassischen Western Blotting mit (primärem und sekundären) Antikörper ähnelt. Statt der spezifischen Bindung des Antikörpers an das Zielprotein findet hier die Wechselwirkung direkt zwischen den beiden Proteinen statt. Besonders stark ist die Bindung bei Proteinen, die auch normalerweise eine starke Affinität zueinander haben - also beispielsweise den Untereinheiten eines Enzymkomplexes. In der Regel wird eines der Proteine an einer Membran immobilisiert, während das andere Protein frei in Lösung als Sonde dient. In der Praxis wird die Intensität einer solchen Protein-Protein-Interaktion wahrscheinlich meist schwächer als eine Antikörper-Protein-Bindung sein. Durch den Wegfall der „Sandwich-Ebenen“ von primärem und sekundären Antikörper (letzterer z.B. als Antikörper-POD-Konjugat, siehe [3.3] ist die Methode andererseits aber sehr direkt und einfach. Durch die radioaktive Markierung des Sonden-Proteins bleibt im Gegensatz zu „gekoppelten“ Methoden das Molekulargewicht gleich, was die Gefahr eines unerwünschten Ablösens verringert. Hiervon abgesehen, sind radioaktive Methoden aufgrund ihrer Sensitivität in der Regel ohnehin vorteilhaft.

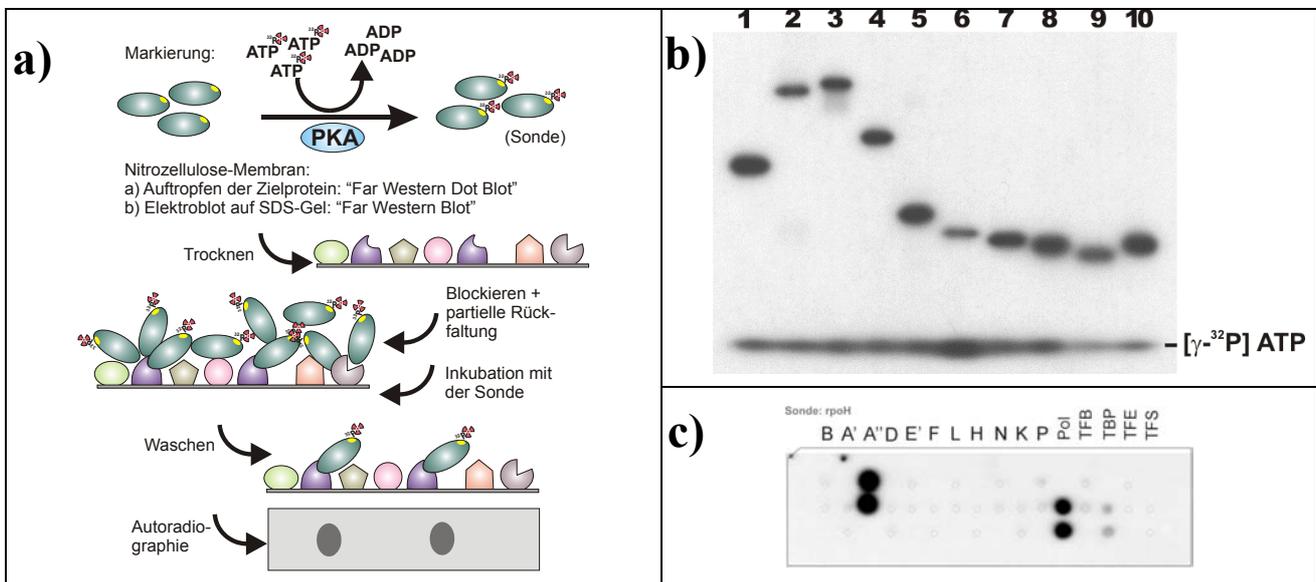


Abb.3-01: Prinzip der Far Western Blotting-Methode

a) Schematische Darstellung der Methode; b) Autoradiographie: Spezifisch über PKA-Site markierte Proteine, die in Far Western Blotting-Experimenten verwendet wurden (aufgetrennt mit 10-17% iger denaturierender PAGE, 0,1% SDS). 1) TBP_N 2) TFB_N 3) D_N 4) E'_N 5) F_C 6) L_N 7) H_C 8) N_N 9) K_N 10) P_N (3 - 10: Untereinheiten der DNA-abhängigen RNA-Polymerase; Tiefgestelltes N bzw. C: N- / C-terminaler Fusionsanteil mit His₆-Tag und PKA-Site für Markierung mit Proteinkinase A (PKA). Im unteren Teil der Abb. ist das nicht eingebaute, freie [γ-³²P]ATP zu erkennen); c) Beispiel für das Ergebnis eines Far Western Dot Blots (mit markierter Untereinheit H als Sonde).

3.6.1 Radioaktive Markierung von Proteinen für Far Western Blotting

Die rekombinanten Untereinheiten D, E', F, H, K, L, N, P der DNA-abhängigen RNA-Polymerase aus *Pyrococcus furiosus* sowie die Transkriptionsfaktoren TBP und TFB wurden mit Hilfe des Enzyms Proteinkinase A (PKA) spezifisch markiert. Dieses "Labeling" wird dadurch ermöglicht, dass die PKA nur beim Vorhandensein einer definierten Erkennungssequenz einen Phosphatrest mit ³²P auf das Zielprotein überträgt (Methode nach (Burgess *et al.*, 2000)). Der wichtigste Vorteil dieser Methode gegenüber anderen Vorschriften zur radioaktiven Markierung von Proteinen z.B. über eine unspezifische Alkylierungsreaktion besteht darin, daß hier evtl. noch in der Präparation enthaltene, kontaminierende Proteine nicht mit markiert werden. Auf diese Weise können unerwünschte Nebenbanden und der "Background" verringert beziehungsweise verhindert werden.

3.6.2 Plasmid pET-33b(+) (Novagen Inc.)

Am Anfang des Projektes wurden die RNA-Polymerase-Untereinheiten in die Plasmide pET-14b (rpoA1, rpoA2, rpoD, rpoE1 und rpoF) bzw. pET-15b (rpoA1 und rpoE2) kloniert. Bei den pET-Vektoren handelt sich um Expressionsvektoren, die eine Gewinnung großer Mengen des rekombinanten Proteins in einem geeigneten *E.coli*-Wirtsstamm wie z.B. BL21(DE3) ermöglichen. Die T7-RNA-Polymerase, die in das *E.coli*-Genom eingeschleust wurde, wird normalerweise nicht von der *E.coli*-RNA-Polymerase transkribiert, weil sich zwischen dem lac-Promotor und dem T7-Gen eine Bindesequenz für das lac-Repressor-Protein befindet, und folglich normalerweise blockiert wird. Bei der Induktion der Expression mit IPTG als Pseudosubstrat, ändert der lac-Repressor durch IPTG-Bindung seine Konformation und gibt den Promotor für das T7-RNA-Polymerase-Gen frei, so dass jetzt große Mengen T7-RNA-Polymerase produziert werden. Diese wiederum bindet an den T7-Promotor, der sich direkt vor dem klonierten Gen in dem pET-Vektor befindet und kann nun das Fremdgen ablesen. Die resultierende mRNA wird ganz normal von den *E.coli*-Ribosomen zur Herstellung von (rekombinantem Fremd-) Protein verwendet. Insgesamt wird der *E.coli*-Stoffwechsel so stark umgestellt, dass die Zellen hierdurch absterben würden, wenn die Induktion der Expression permanent und nicht erst bei einer definierten Zelldichte erfolgen würde. Dieses Basis-Prinzip wurde von der Firma Novagen/Merck in einer Vielzahl von kommerziell erhältlichen Plasmid-Varianten an alle möglichen Zwecke angepaßt. In unserer Arbeitsgruppe wurden beispielsweise häufig pET-17b (ohne zusätzlichen Fusionsanteil) und pET-14b (mit 20 Aminosäuren langem „His₆-Tag“) verwendet, wobei der Fusionsanteil normalerweise einen die Proteinreinigung per Affinitätschromatographie ermöglichenden „Tag“ also z.B. eine künstliche Sequenz aus 6 hintereinanderliegenden Histidin-Resten (His₆) für die Bindung an eine Nickel-Säule enthält.

Das meist deutlich vereinfachte und standardisierte Reinigungsprotokoll erkaufte man allerdings mit dem zusätzlichen, evtl. störenden Fusionsanteil des resultierenden Proteins. Aus diesem Grund besteht meist die Möglichkeit, einen Teil des Fusionsanteils anschließend zum Beispiel mit Thrombin- oder Enterokinase-Behandlung zu entfernen, was allerdings wieder ein zusätzlicher Aufwand ist (und glücklicherweise oft nicht nötig ist).

Der Einsatz von pET-33b(+) bot sich also an, weil Expressionsbedingungen, Wirtsstämme etc. beibehalten werden konnten. Neben der auch hier gegebene Möglichkeit zur Reinigung per Affinitätschromatographie / His₆-Tag, ist die PKA-Markierungsregion mit der Sequenz ArgArgAlaSerVal entscheidend. Die auch hier vorhandene Thrombin-Schnittstelle ist so angeordnet, dass nach erfolgter Hydrolyse nur der His₆-Tag entfernt wurde – die für die anschließende radioaktive Markierung nötige PKA-Erkennungssequenz verbleibt dagegen sinnvoller Weise als zusätzlicher Fusionsanteil am rekombinanten Protein. Da sich der restliche Fusionsanteil nicht meßbar auf die Markierungseffizienz auswirkt, wurde in den meisten Fällen auf eine Hydrolyse mit Thrombin verzichtet.

Allerdings sollte das Vorhandensein 25 zusätzlicher Aminosäuren am N- oder C-Terminus mit in die Überlegungen einbezogen werden, falls es zu experimentellen Problemen bei der Arbeit mit den Proteinen kommen sollte.

3.6.3 Konstruktion einer C-terminalen Plasmid-Version: pET-33b(+)_{c2}

Die Klonierung, Expression und Reinigung der mit der zusätzlichen PKA-Markierungsregion ausgestatteten Proteine unterschied sich nicht wesentlich von der bei den zuvor mit Hilfe von pET-14b und pET-15b erstellten Proteinen. Jedoch gab es gravierende Probleme bei der radioaktiven Markierung. Laut Protokoll (Burgess *et al.*, 2000) wurde die Abtrennung der bei der Markierungsreaktion nicht inkorporierten Nucleotide mit BIORAD „Bio-Spin 6“-Gelfiltrationssäulen empfohlen, was jedoch nur mit einigen *Pyrococcus*-Proteinen gut funktionierte. So ließ sich z.B. TBP sehr gut, TFB dagegen nur mit sehr schlechter Ausbeute markieren (Lange, 2001). Ähnliches galt für bestimmte Untereinheiten der RNA-Polymerase, wobei z.B. rpoD eines der Proteine war, beim dem diese Prozedur (im Gegensatz zu den meisten anderen Untereinheiten) gut funktionierte. Dies war einer der Gründe, die zu der Entscheidung führten, eine (kommerziell bei Novagen/Merck nicht verfügbare) C-terminale Version von pET-33b(+) zu konstruieren – in der Hoffnung, damit die offensichtlichen Markierungsprobleme zumindest bei einem Teil der Proteine durch eine Verlagerung des Fusionsanteils zu lösen. Auch wenn sich später zeigte, dass die Hauptursache der schlechten Markierung in einer unspezifischen Anlagerung der teilweise nicht so gut löslichen *Pyrococcus*-Proteine an die Säulenmatrix lag, stellt die C-terminale Variante eine nützliche Erweiterung der Anwendungsspektrums dar.

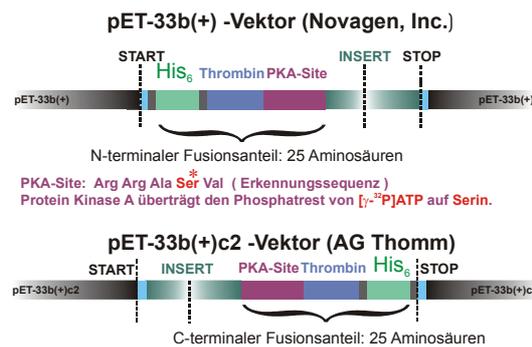


Abb.3-02: Schematische Darstellung des Vektors pET-33b(+)_{c2} und Variante pET-33b(+)

START: Translationsstartpunkt; **His₆:** His₆-Tag für Affinitätschromatographie z.B. mit Nickel-Säule; **Thrombin:** Erkennungsregion zur hydrolytischen Abtrennung eines Teils des Fusionsanteils (LeuValProArg^{*}GlySer); **PKA-Site:** Erkennungsregion für spezifische, radioaktive Protein-Markierung mit Proteinkinase A ([γ-³²P]ATP); **INSERT:** das jeweilige Gen, dessen Produkt exprimiert werden soll; **STOP:** Stop-Codon. **pET-33b(+)_{c2}:** Fusionsanteil C-terminal statt N-terminal wie bei pET-33b(+).

Konstruktion des Vektors pET-33b(+)_{c2}, einer C-terminalen pET-33b(+)-Variante:

Um Proteine nicht nur N-terminal sondern auch C-terminal radioaktiv markieren zu können, mußte der kommerziell erhältliche Original-Vektor von Novagen/Merck modifiziert werden. Im ersten Schritt wurde ein Teil der Multiple-Cloning-Site zwischen Xba I und Nde I durch die entsprechende Sequenz in pET-17b ersetzt, um die N-terminalen Fusionsanteile zu entfernen. Im zweiten Schritt wurde dann zwischen EcoRI und Xho I die Sequenz für Kinase-Site und Thrombin-Site eingefügt. Dadurch, dass die Anzahl der Restriktionsschnittstellen reduziert wurde, ist der Fusionsanteil des resultierenden Proteins etwas kleiner. Folgende Schnittstellen stehen für die Klonierung zur Verfügung: Nde I, Nhe I, Bam HI und Eco RI.

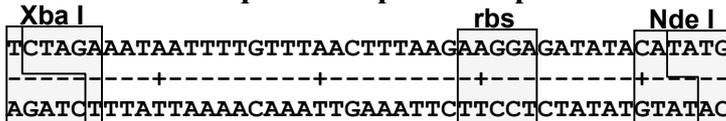
"Multiple Cloning Site" von pET-33b(+), Novagen Inc. (Original-Vektor):



Abb.3-03: "Multiple Cloning Site" von pET-33b(+)_{c2} im Detail

Xba I, Nde I, Nhe I, Bam HI, Eco RI, Sac I, Xho I: Restriktionsschnittstellen; rbs: Ribosomen-Bindestelle; Start: Translationsstart; Kinase Site: ArgArgAlaSerVal-Erkennungssequenz für Proteinkinase A zwecks spezifischer, radioaktive Markierung mit [γ -³²P]ATP.

Insert mit Teilsequenz entsprechend pET-17b:

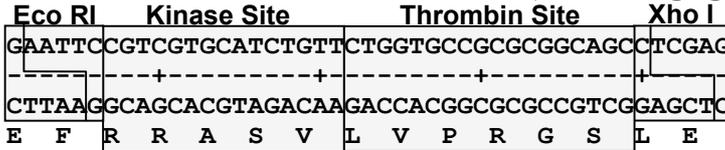


Verwendete Oligonucleotide:

Insert17bf3: 5'-CTAGAAATAATTTTGTTTAACTTTTAAGAAGGAGATATACA-3' (40n)

Insert17br: 5'-TATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTT-3' (38n)

Insert mit Thrombin und Kinase Site (Reihenfolge getauscht):



Verwendete Oligonucleotide:

InsertPKAf4: 5' -AATTCCGTCGTCGTCATCTGTTCTGGTGCCGCGCGGCAGCC-3' (39n)

InsertPKAr5: 5' -TCGAGGCTGCCGCGCGGCACCAGAACAGATGCACGACGG-3' (39n)

Resultierende "Multiple Cloning Site" von pET-33b(+)-c2:

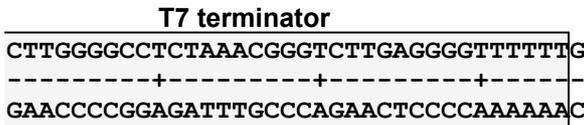
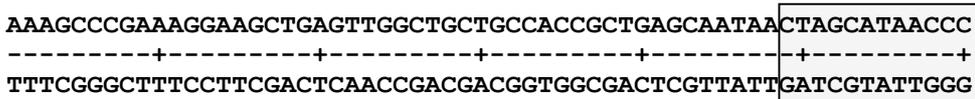
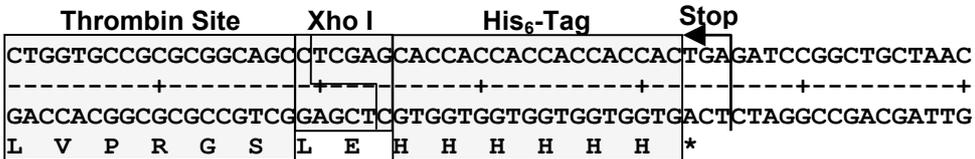
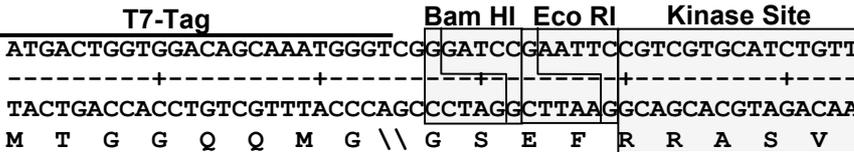
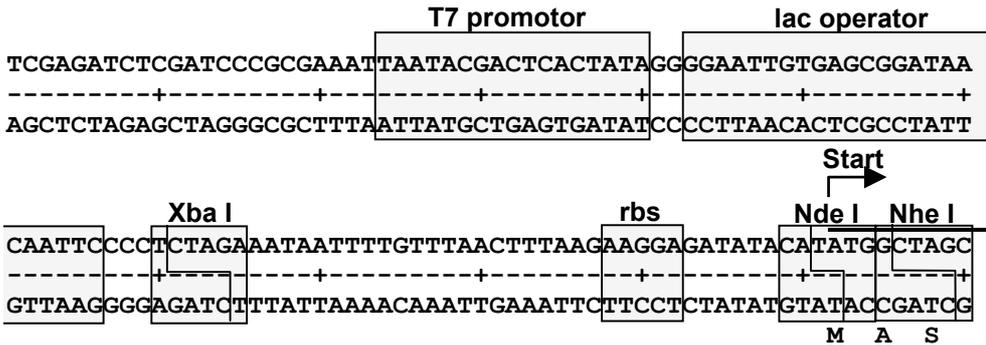


Abb.3-04: Konstruktion und "Multiple Cloning Site" von pET-33b(+)-c2 im Detail

Die C-terminale pET-33b(+)-Variante wurde in zwei Schritten erstellt: Zunächst wurde über die Schnittstellen Xba I und Nde I der in pET-33b(+) kodierte N-terminale Fusionsanteil entfernt und durch den entsprechenden Abschnitt aus pET-17b ersetzt. Anschließend erfolgte über die Schnittstellen Eco RI und Xho I das Einfügen des neuen C-terminalen Fusionsanteils, in dem His₆-Tag, "Thrombin Site" und "Kinase Site" in umgekehrter Reihenfolge angeordnet sind. Beide Schritte wurden mit zueinander komplementären Paaren von Oligonukleotiden realisiert (siehe oben).

3.6.4 Far Western Dot Blotting

Um später die Identifizierung der Positionen auf dem Film (Autoradiographie) zu erleichtern, wurden die Membranen auf einer Schablone getropft. Diese war auf einer Glasplatte befestigt und wurde von unten mit einer Lampe durchstrahlt. Eine spezielle Form der Schablone mit hohlen Punkten wurde bei der grafischen Auswertung halbtransparent in die Fotos der Autoradiogramme hineinkopiert, um die Interpretation zu erleichtern und eine falsche Zuordnung vor allem bei Membranen mit sehr wenig Signalen zu vermeiden.

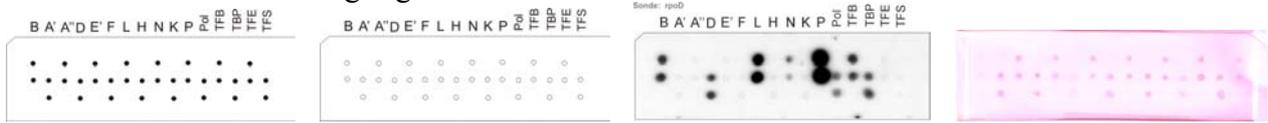


Abb.3-05: Beispiel für Far Western Dot-Blot

Von links nach rechts: Unterlegschaablone für das Auftropfen der Proteine auf die Nitrocellulose-Membran. Hohlschaablone, die halbtransparent über das Foto gelegt wird. Fertige Grafik mit den Signalen des Autoradiogramms. Ponceau S-gefärbte Membran, zur Kontrolle der Protein-Mengen.

3.6.5 Far Western Blotting

Beim Far Western Blotting erfolgt der Transfer der Proteine wie bei einem "klassischen" Western Blotting aus einem denaturierenden PA-Gel heraus per Elektrophorese auf die Nitrocellulose-Membran. Die folgenden Versuchsschritte sind dann identisch mit dem Protokoll für "Far Western Dot Blotting".

3.6.6 Zusammenfassung: Far Western Blotting in Stichworten

1.) Klonierung des gewünschten Gens in pET-33b(+) oder pET-33b(+)_{c2}

Die pET-33b(+) –Klonierungs- und Expressionsregion kodiert für einen zusätzlichen N-terminalen Fusionsanteil (bei pET-33b(+)_{c2} C-terminal). Die 25 Aminosäuren lange Sequenz enthält neben einem His₆-Tag zur einfachen Reinigung eine 5 Aminosäuren lange PKA-Site zur spezifischen radioaktiven Markierung des resultierenden rekombinanten Proteins. Die Erkennungssequenz ist ArgArgAlaSerVal.

2.) Expression und Reinigung: Vom Vorgehen her besteht kein Unterschied zu His₆-Tag-Proteinen, die in andere Vektoren wie zum Beispiel pET-14b kloniert wurden. Dadurch, dass der Fusionsanteil bei pET-33b(+) 25 Aminosäuren lang ist (20 bei pET-14b), kann es besonders bei kleinen Proteinen zu deutlichen Abweichungen im Laufverhalten bei der Gelelektrophorese kommen (d.h. die Proteine laufen aufgrund des höheren Molekulargewichtes langsamer).

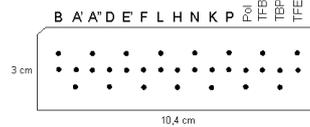
Wegen der Spezifität der Markierungsreaktion bzw. der Proteinkinase A (PKA, aus Herzmuskel) können auch Proteinpräparationen eingesetzt werden, die noch nicht völlig frei von kontaminierenden *E.coli*-Proteinen sind.

3.) Markierungsreaktion ca. 60min bei 30°C mit PKA: Das Pipettierschema ist einfach (detailliertes Schema siehe Anhang 1): Neben 10-fach PKA-Puffer sind als weitere Komponenten nur das zu markierende Protein, die Proteinkinase A und das [γ -³²P]ATP im Ansatz enthalten (und natürlich Wasser, um auf ein definiertes Volumen aufzufüllen).

3b.) Optional Abtrennung des im Markierungsansatz verbliebenen, nicht eingebauten [γ -³²P]ATP, z.B. durch Säulenchromatographie mit BIORAD "Biospin 6"-Säulen. Es hat sich gezeigt, dass dieser Schritt die Radioaktivität in den weiteren Versuchsschritten nur minimal verringert und, da die ATPs in den meisten Ansätzen nicht stören, ruhig weggelassen werden kann. Bei Proteinen mit begrenzter Löslichkeit kann die Verwendung die Menge resultierenden markierten Proteins sogar unnötig verringern. → das radioaktiv markierte Protein kann jetzt als Sonde verwendet werden!

4a) "Far Western Dot Blotting": Auftropfen der zu untersuchenden Zielproteine auf die Nitrocellulose-Membran und anschließende Trocknung (zur Unterscheidung von Artefakt-Signalen immer 2 Spots tropfen!). Praktisch ist für die spätere Auswertung das Auftropfen der Proteine auf

einem "Leuchttisch" und einer untergelegten Schablone, so dass das Licht von unten durch Schablone und Nitrocellulose-Membran hindurchscheint.



4b) "Far Western Blotting": Auftrennung der Zielproteine per Elektrophorese auf einem denaturierenden PA-(Gradienten)-Gel (0,1% SDS). Transfer der Proteine aus dem Gel auf eine Nitrocellulose-Membran per Elektroblothing. Die Durchführung unterscheidet sich hierbei nicht von der Vorgehensweise bei einem "normalen" Far Western Blotting-Experiment (Detektion per Antikörper). Trocknung der Membran. Um besonders bei den kleinen Untereinheiten ein "Durchblotten" der Proteine zu vermeiden, wurde "Protran Nitrocellulose Transfer Membran" von Schleicher & Schuell verwendet (PROTRAN BA75, 300mm x 3m Rolle, Porengröße 0,05µm, Order No. 10 402196). Zunächst wurde diese Membran auch für die Dot Blots verwendet – allerdings reicht hier auch die billigere "Optitran BA-S 85 Reinforced NC" (300mm x 3m Rolle, Porengröße 0,45µm, Ref.-No: 439196), die im späteren Verlauf zum Einsatz kam. Als Blot-Anlage wurde ein Semi-Dry-System von Biorad verwendet (in Verbindung mit Anodenpuffer I und II sowie Kathodenpuffer nach den üblichen Standardprotokollen (und einem Papierstapel aus insgesamt 18 Whatman-Papierstücken)). Geblottet wurde 2 Stunden am Gerätemaximum von 10Watt. Nach Beendigung des Elektrobloths wird das Gel fixiert und mit einer Coomassie-Färbung auf möglichst vollständigen Protein-Transfer hin untersucht.

Es ist empfehlenswert, die Membran gut zu trocknen (eventuell trocknen mit einem Fön), weil sonst die Gefahr besteht, dass sich Proteine während der Inkubationsschritte von der Oberfläche ablösen. Bisherige Praxis: Einen Tag bei Raumtemperatur (und/oder 30min im 37°C-Raum). Je nach Art der Proteine kann durch eine längere Trocknung, oder vorsichtige kurzzeitige Erwärmung erreicht werden, dass auch leicht lösliche Proteine fester auf der Membran haften. Allerdings darf hierdurch nicht die für die Bindung notwendige partielle Rückfaltung der Proteine verhindert werden.

5.) Einige Stunden Inkubation der Membranen mit Blockierlösung in einer Schale oder einer Folientasche auf dem Schüttler, um die gewünschte partielle Rückfaltung (Renaturierung) der Proteine zu gewährleisten (besser: über Nacht). In der Praxis ist das Arbeiten mit einer Schale, die knapp größer ist als die Membran deutlich angenehmer als mit einer Folientasche, da sich Kontaminationen mit radioaktiver Lösung leichter vermeiden lassen.

Rezept für die Blockierlösung: siehe [3.6.7]. Optimierung: Wird zu kurz inkubiert, besteht die Gefahr, dass die Membran nicht genügend blockiert wird, so dass später stärkere Hintergrundsignale sichtbar ist. Außerdem kann es zu ungenügender Rückfaltung und folglich zu einem geringeren Binden der radioaktiven Sonde kommen, was sich in verminderten Signalstärken äußern würde. Es wurde anfangs über Nacht blockiert, was aber für die gut löslichen Proteine ein Nachteil sein kann, weil hier ein komplettes Rückfalten im schlimmsten Falle zu einem kompletten Abdiffundieren von der Membran (und schließlich zu einem Fehlen von Signalen auf dem Röntgenfilm) führen kann. Um dies zu beurteilen ist die abschließende Ponceau S-Färbung besonders wichtig.

6.) Radioaktivität hinzugeben: Alte Lösung entfernen und Auffüllen mit neuer Blockierlösung, in die jetzt zusätzlich das spezifisch markierte Protein als Sonde gegeben wird. Inkubation für 2 Stunden bei 4 - 8°C auf einem Schüttler (evtl. an der Membran haftende Luftblasen entfernen).

7.) Waschen: Entsorgen der radioaktiven Markierungslösung und 2 - 3 Waschschritte frischer Blockierlösung (jeweils ca. 10min). In der Praxis hat sich gezeigt, dass 2 x Waschen meist ausreicht.

8.) Schnelles Trocknen der Membranen: Der Einsatz eines handelsüblichen Föns hat sich bewährt, um ein eventuelles Verlaufen von radioaktiv markiertem Protein zu verhindern.

9.) Detektion mutmaßlicher Protein-Protein-Wechselwirkungen durch Autoradiographie. Als praktisch hat es sich erwiesen, die Nitrocellulose-Membranen mit Tesa-Film auf eine passend zurechtgeschnittene Folie zu kleben, die dann in die Filmkassette gelegt wird. Auf diese Weise können mehrere Membranen aus unterschiedlichen Versuchen mit einem gemeinsamen Film ausgewertet werden. Bei der Auswertung der Experimente ist es sinnvoll, immer mindestens zwei Filme unterschiedlich lange auflegen).

10.) Überprüfung der auf der Nitrocellulose-Membran verbliebenen Proteinmengen mit Ponceau S-Färbung (siehe [3.3]) Dies ist wichtig beim Vergleich von Signal- beziehungsweise Bindungsstärken. Denn neben leichten Variationen in der Gesamtmenge der aufgetragenen Zielproteine spielt die verschieden gute Fixierung an die Nitrocellulose-Membran, also zum Beispiel die Neigung zum Ablösen während der Inkubation eine gewisse Rolle.

Außerdem kann so auf ausreichenden Transfer der Proteine aus dem SDS-Gel auf die Membran beim Elektrophoretogramm getestet werden: Ungeeignete Prozentigkeit oder Porengröße des SDS-Geles, aber auch ungünstige (positive) Proteinladung oder schlechte Löslichkeit können die resultierenden Proteinmengen verringern. Konkret bedeutet dies, dass für die großen Untereinheiten B und A' der RNAP ein 8%iges SDS-Gel geeignet ist, während für die kleinste Untereinheit P ca. 15% PA ein guter Richtwert ist (bei 12% PA werden P und K noch nicht aufgetrennt, sondern befinden sich in der Lauffront). Deshalb war für das Blotting einer Gesamtübersicht aller Untereinheiten der RNAP zwingend ein Gradientengel erforderlich (8 – 15%. Rezept s.Anhang).

3.6.7 Blockierlösung für Far Western Blots

Das genaue Rezept für den "Probing Buffer", der zugleich Blockier- und Waschlösung ist, wurde folgender Publikation entnommen: (Burgess *et al.*, 2000)

Pro Membran werden ca. 100ml Blockierlösung benötigt (für sehr große Membranen mehr).

Die Lösung ist 1-2 Tage bei 4°C haltbar.

Stammlösung	Final	100ml	200ml	300ml	400ml	500ml	1000ml
1 M Hepes, pH 7,2	20mM	2ml	4ml	6ml	8ml	10ml	20ml
2,5 M KCl	200mM	8ml	16ml	24ml	32ml	40ml	80ml
1 M MgCl₂·6H₂O	2mM	200µl	400µl	600µl	800µl	1ml	2ml
0,25 M ZnCl₂	0,1mM	40µl	80µl	120µl	160µl	200µl	400µl
1 M DTT	1mM	100µl	200µl	300µl	400µl	500µl	1ml
100% Tween 20	0,5%	500µl	1ml	1,5ml	2ml	2,5ml	5ml
Magermilchpulver	1%	1g	2g	3g	4g	5g	10g
50% Glycerin	10%	20ml	40ml	60ml	80ml	100ml	200ml
MilliQ Wasser		ad 100ml	ad 200ml	ad 300ml	ad 400ml	ad 500ml	ad 1000ml

Abb.3-06: Blockierlösung für Far Western Blotting-Experimente nach (Burgess *et al.*, 2000)

3.6.8 Beispiele für Markierungsansätze mit Proteinkinase A (PKA)

In den unten aufgeführten Protokollen wurde das Gesamtvolumen reduziert (als "Einfachansatz" wurden 5µl gesetzt). Auch die Menge PKA wurde reduziert: Die angegebene Menge ist völlig ausreichend. Die Ansätze werden direkt in PCR-Gefäße pipettiert, wobei die Radioaktivität ganz zuletzt im Isotopen-Labor zugegeben wird. Die Inkubation erfolgt bei 30°C im Thermocycler (entsprechend den Angaben in der Produktbeschreibung von Sigma; bei Novagen wird eine niedrigere Temperatur angegeben!). Möglicherweise würde schon eine Inkubationsdauer von 30min ausreichen; sicherheitshalber wurden 60-70min gewählt.

Die Ansätze wurden anschließend direkt als Sonde verwendet oder bei -20°C eingefroren. Die unter 3b) erwähnte optionale Abtrennung der freien [γ -³²P]ATP-Moleküle hat nur bei einigen der getesteten Proteine gut funktioniert (unter anderem TBP). Bei den meisten übrigen führte der Einsatz der empfohlenen BIORAD "Biospin 6"-Säulen aber zu drastischen Verlusten bei den

resultierenden Proteinmengen der nur begrenzt löslichen Proteine und wurde deshalb komplett weggelassen. Zwar erhöht sich hierdurch die Gesamtradioaktivität in den Far Western Blotting – Experimenten. Es wurde aber gezeigt, dass bei einer erfolgreichen Markierungsreaktion mindestens die Hälfte der eingesetzten $[\gamma\text{-}^{32}\text{P}]\text{ATPs}$ in die entsprechenden Proteine eingebaut wird (siehe [Abb.3-01,b]), so dass maximal mit einer Verdoppelung der Radioaktivität gerechnet werden muß. Dies bedeutet für die Praxis, dass (neben den üblichen und selbstverständlichen) keine zusätzlichen Schutzmaßnahmen nötig sind.

Die eingesetzten Proteinmengen sind bei Proteinen mit hohem Molekulargewicht groß und bei kleinen Proteinen deutlich geringer. Dies ist deshalb sinnvoll, weil die stöchiometrische Beziehung zwischen der Anzahl der PKA-Sites und den angebotenen $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ -Molekülen entscheidend ist – nicht aber die absolute Proteinmenge.

Beispiel:

→ Radioaktivität ist auf den 09.05.03 kalibriert, d.h. d.h. Tag 5 = Faktor 0,824 , d.h. in einem μl sind $8,24\mu\text{Ci}$ enthalten; da ich einen Doppelansatz mache ($10\mu\text{l}$), benötige ich $2 \cdot 10\mu\text{Ci}$, also $20\mu\text{Ci}$, d.h. ich muß $2,4\mu\text{l}$ einsetzen

(1): rpoE1/PKA-N (Frkt "17": ca. 1,6mg/ml)

1,0 μl 10x PKA-Puffer
 1,0 μl rpoE1/PKA-N (Frkt "17", ca. 1,6 μg Protein)
 4,6 μl H₂O
 2,4 μl [$\gamma\text{-}^{32}\text{P}$]ATP (20 μCi)
 1,0 μl PKA (10 U/ μl)

(2): rpoF/PKA-C, "Frkt.12" (ca. 5mg/ml); 1:5-Verdünnung mit H₂O auf 1mg/ml;

1,0 μl 10x PKA-Puffer
 1,0 μl rpoF/PKA-C (ca. 1 μg Protein)
 4,6 μl H₂O
 2,4 μl [$\gamma\text{-}^{32}\text{P}$]ATP (20 μCi)
 1,0 μl PKA (10 U/ μl)

→ 70min bei 30°C im Thermocycler; 5 μl von jedem Ansatz in die jeweilige Folientasche oder Schale; Gefäße mit Radioaktivität vor dem Öffnen immer kurz anzentrifugieren, um Kontaminationen zu vermeiden; 2 h Inkubation bei 4-8°C unter langsamem Schütteln; 2 x mit 20ml Blockierlösung waschen; Trocknen mit dem Fön; Aufkleben auf Folie und Auflegen eines Films über Nacht.

3.7 Herstellung von PA-Gradientengelen (denaturierend oder nativ)

Dieses Rezept bezieht sich auf ein 10x10cm Gel mit 0,8mm Spacern / Kamm. Verwendet wird ein einfacher Gradientenmischer mit einem Plexiglasblock, in den die beiden Pufferkammern gebohrt sind, und die über einen (mit einem Hebel verschließbaren) Kanal untereinander verbunden sind. In der vorderen Kammer befindet sich die schwere Lösung, die mit einem kleinen Rührfisch ständig gemischt wird (in die hintere füllt man die leichte Lösung). Über eine peristaltische Pumpe wird das Gel senkrecht stehend gegossen (sonst kann sich der Gradient nicht so gut ausbilden).

Hier wird von 6,5ml Gesamtvolumen ausgegangen (also leichte und schwere Lösung zusammen). Das genaue Gesamtvolumen muß an das jeweilige Elektrophoresesystem und die unterschiedlich breiten Spacer angepaßt werden. Nach dem Gießen des Gradientengels erfolgt die Übersichtung mit ca. 500 μl wassergesättigtem Isobutanol oder 1-Butanol.

Allgemein ist zu sagen, dass schnell gearbeitet werden sollte, damit das Gel nicht in den Schläuchen fest wird, oder Unregelmäßigkeiten beim Gießen auftreten. Außerdem ist ein sofortiges, mehrmaliges und sehr sorgfältiges Nachspülen des Gradientenmischers und der Schläuche wichtig.

Gradient 8% bis 15% / Gesamtvolumen 6,5ml #####**Leichte Lösung (hintere Kammer)**

867 µl Acrylamid-Lsg. (30%) → ergibt 8%

813 µl 4xTris/HCl/SDS pH 8,8

1677 µl H₂O_{MilliQ} (2x 838 µl)

11 µl 10% APS

2,4 µl TEMED

Schwere Lösung (vordere Kammer)

(als Volumen wegen Saccharose nicht 3,25ml sondern 2,5ml berechnen)

1219 µl Acrylamid-Lsg. (40% → ergibt 15% (2x 609 µl)

813 µl 4xTris/HCl/SDS pH 8,8

468 µl H₂O_{MilliQ}

0,73g Saccharose

11 µl 10% APS

2,4 µl TEMED

Sammelgel:

1,0 ml Acrylamid-Lsg. (30%)

2,0 ml 4xTris/HCl/SDS pH 6,8

4,7 ml H₂O_{MilliQ}

25 µl 10% APS

4 µl TEMED

Gradient 12% bis 17% / Gesamtvolumen 6,5ml #####**Leichte Lösung (hintere Kammer)**

1,3 ml Acrylamid-Lsg. (30%) → ergibt 12% (2x 650 µl)

813 µl 4xTris/HCl/SDS pH 8,8

1137 µl H₂O_{MilliQ}

11 µl 10% APS

2,4 µl TEMED

Schwere Lösung (vordere Kammer)

(als Volumen wegen Saccharose nicht 3,25ml sondern 2,5ml berechnen)

1380 µl Acrylamid-Lsg. (40% → ergibt 17% (2x 690 µl)

813 µl 4xTris/HCl/SDS pH 8,8

307 µl H₂O_{MilliQ}

0,73g Saccharose

11 µl 10% APS

2,4 µl TEMED

Sammelgel: siehe oben**Für Native-PAGE mit Gradienten-Gelen:**

→ Rezept siehe oben, jedoch die Tris-Puffer ohne SDS.

Puffer für Native-PAGE mit erhöhtem pH nach (Ehlers *et al.*, 2002):

"pH 9,4-System":

Trenngel-Puffer:

4x Puffer: 9,09g Tris → 750mM Tris/HCl pH 8,9

(ad 100ml mit H₂O_{MilliQ}, filtrieren und autoklavieren)

Sammelgel-Puffer:

4x Puffer: 3,03g Tris → 250mM Tris/HCl pH 7,5

(ad 100ml mit H₂O_{MilliQ}, filtrieren und autoklavieren)

5x Laufpuffer:

413mM Tris/HCl pH 9,4 + 165mM Glycin

3.8 Bioinformatik: Verwendete Programme und (Online-)Tools

NCBI: National Center for Biotechnology Information: Diese Seite ist die erste Anlaufstelle für Literatur-Recherche, Online-BLAST-Suche (nicht nur) in mikrobiellen Genomen, sowie Suche nach DNA- und Protein-Sequenzen aber auch kompletten Genomen. Alle in dieser Arbeit erwähnten, komplett sequenzierten Genome stammen aus dieser Quelle (siehe [8.5]).

<http://www.ncbi.nlm.nih.gov/> (allgemeine Startseite)

http://www.ncbi.nlm.nih.gov/genomes/static/a_g.htm (Liste komplett sequenzierter *Archaea*)

http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi (BLAST-Suche in mikrobiellen Genomen)

HUSAR Bioinformatics Lab (ein „Biocomputing Service“ des DKFZ Heidelberg). Hier hat sich einerseits das frei zugängliche **SRS** (Sequence Retrieval System) zum Durchsuchen der einschlägigen Datenbanken bewährt (besonders detaillierte Suchfunktionen). Daneben hat vor allem HUSAR („Heidelberg Unix Sequence Analysis Resources“) im engeren Sinne meine Arbeit begleitet. Es handelt sich um eine auf dem „GCG Wisconsin package“ beruhende, erweiterte Sammlung von bioinformatischen Programmen, die kostenpflichtig per Webinterface (oder per Telnet) bereitgestellt werden. Die meisten Alignments in dieser Arbeit wurden mit dem hier verfügbaren Programm „Malign“ erstellt. Einige Alignments wurden auch mit Clustal X (siehe unten) generiert. Die Ergebnisse der Funktion „Pepstats“ haben mir als Referenz (und Kontrolle) für die eigene Programmierung der Berechnung des theoretischen Molekulargewichtes, des theoretischen Isoelektrischen Punktes und der Netto-Ladung gedient (siehe 8.4.1)

<http://genome.dkfz-heidelberg.de/biounit/>

Bioedit 5.0.9: Ein Programm mit sehr großem Funktionsumfang (vor allem handelt es sich um einen Editor für multiple Alignments, im folgenden meist kurz als "Alignments" bezeichnet). Ich habe dieses Programm wegen seiner guten Export-Funktion im RTF-Format für die grafische Darstellung von Alignments benutzt. Außerdem läßt sich zum Beispiel eine lokale BLAST-Suche durchführen (siehe 8.3).

<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>

Clustal W 1.83 (Higgins *et al.*, 1996) bzw. die zugehörige grafische Oberfläche **Clustal X 1.83:** Mit diesem Programm können unabhängig von diversen, im Internet verfügbaren Online-Tools multiple Alignments auf dem eigenen Rechner durchgeführt werden.

<http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html>

TreeView 1.6.6: Einen Stammbaum auf der Basis von Alignments zeichnen. Es kann sich sowohl um Nucleotid- als auch um eine Proteinsequenzen handeln (Page, 1996).

<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>

Deep View - Swiss-Pdb Viewer 3.7: 3D-Darstellung von auf der Basis von Röntgenstruktur-Analyse und NMR gewonnenen Protein-3D-Daten. 3D-Modelle sind eine unschätzbare Hilfe zum Beispiel bei der Plausibilitätskontrolle von mit praktischen Methoden in Laborexperimenten gewonnenen Ergebnissen.

<http://us.expasy.org/spdbv/>

Die Hauptseite <http://us.expasy.org/> bietet den möglicherweise umfangreichsten Zugang zu 2D- und 3D-Protein-Tools, Informationen und Datenbanken.

Cn3D 4.1: Ein von NCBI (siehe oben) bereitgestelltes 3D-Programm, das ähnlich wie der Pdb-Viewer eine räumliche und intuitiv leicht zu erfassende Darstellung von Proteinen ermöglicht. Auch Protein-Komplexe und Kombinationen aus Proteinen und DNA- bzw. RNA-Fragmenten sind möglich. Das Programm arbeitet eng mit dem Webbrowser zusammen, und ermöglicht so eine schnelle Visualisierung der auf den NCBI-Seiten zur Verfügung gestellten bioinformatischen

Daten. Gelesen werden Daten im MMDB-Format (Molecular Modeling DataBase), wobei es sich um eine Teilmenge der aktuell verfügbaren Proteinstrukturen im PDB-Format handelt (die Daten wurden teilweise korrigiert und aufbereitet; original-PDB-Dateien können leider nicht geladen werden). Die Bildqualität und die Möglichkeiten zum Arrangieren der Proteine sind exzellent, weshalb für die 3D-Darstellungen in dieser Arbeit hauptsächlich auf dieses Programm zurückgegriffen wurde.

<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>

Protein Explorer 2 (Nachfolger von RasMol): Ein weiteres Tool zur Visualisierung der 3D-Struktur von Proteinen. Etwas problematisch bei der Browser-Kompatibilität.

<http://www.umass.edu/microbio/rasmol/>

<http://molvis.sdsc.edu/protexpl/frntdoor.htm>

<http://www.umass.edu/microbio/chime/registfrm/downlpew.htm>

Ultraedit 10: Die in dieser Arbeit gezeigten bioinformatischen Analysen beruhen auf einer Vielzahl von Einzeldateien im Textformat. Ohne einen komfortablen und leistungsstarken Texteditor wäre ein zügiges Arbeit stark erschwert worden. Z.B. lassen sich die großen Genom-Dateien im Genbank-Format (und die daraus von mir generierten Listen) nur vernünftig mit diesem oder einem ähnlichen Programm betrachten und manipulieren.

<http://www.ultraedit.com/>

4 Ergebnisse und Diskussion

Wie schon in der Einleitung erwähnt, besteht die RNA-Polymerase bei dem hier untersuchten Organismus *Pyrococcus furiosus* aus 11 – 12 Untereinheiten. Um für die Untersuchungen von Protein-Protein-Wechselwirkungen nicht nur die komplette, nativ (aus *Pyrococcus*) gereinigte RNA-Polymerase (siehe [4.12]) sondern auch die einzelnen Untereinheiten rekombinant in großer Menge und Reinheit zur Verfügung zu haben, mußten die entsprechenden Gene im Genom identifiziert, kloniert, in *E.coli* exprimiert und anschließend gereinigt werden.

Ohne die Genomprojekte, die mittlerweile auch für *Archaea* immer zahlreicher werden, wäre ein solches Projekt nicht denkbar. Man muß bedenken, dass erst 1996 das erste komplett sequenzierte archaeele Genom veröffentlicht wurde, und zwar das von *Methanocaldococcus (Methanococcus) jannaschii* (Bult *et al.*, 1996). Seit der Vollendung der Genom-Sequenzierung dreier Arten der Gattung *Pyrococcus*, nämlich *P.horikoshii* (Kawarabayasi *et al.*, 1998), *P.abysssi* (Robb *et al.*, 2001; Cohen *et al.*, 2003) und *P.furiosus* (Robb *et al.*, 2001) ergeben sich interessante Vergleichsmöglichkeiten (Maeder *et al.*, 1999; Lecompte *et al.*, 2001), worauf im späteren Verlauf noch detaillierter eingegangen wird.

Auch wenn das *P.furiosus*-Genom lange Zeit in einer vorläufigen, noch nicht so komfortabel zugänglichen Form vorlag, existierte seit 1998 eine wichtige Referenz, nämlich die bereits komplett sequenzierte Schwesterart *P.horikoshii*. Anhand dieser und anderen Daten konnten schnell einige der RNAP-Untereinheiten auch im Genom von *P.furiosus* lokalisiert und mit den praktischen Experimenten begonnen werden.

Die inzwischen in jeder Hinsicht sehr viel größere Datenbasis ermöglicht folgenden Überblick:

4.1 Untereinheiten der DNA-abhängigen RNA-Polymerase bei *Pyrococcus furiosus*

Die RNA-Polymerase aus *Pyrococcus furiosus* ist (wie bei den anderen *Archaea* auch) ähnlich komplex aufgebaut wie die eukaryotische Pol II (zum Beispiel aus *Saccharomyces cerevisiae*). E'' (rpoE2) wurde noch nicht mit in die unten gezeigte Übersicht aufgenommen, weil die Versuchsdaten hier noch spärlich sind. Allerdings besitzt das Gen ein hochkonserviertes Zinkfinger-Motiv (siehe [8.1.3.6]) und wird sehr wahrscheinlich im Rahmen gekoppelter Translation trotz geringfügiger Überlappung mit rpoE1 (siehe [4.5]) in *Pyrococcus* exprimiert. Weitere Experimente müssen zeigen, ob E'' wirklich permanenter Bestandteil des RNA-Pol-Komplexes ist.

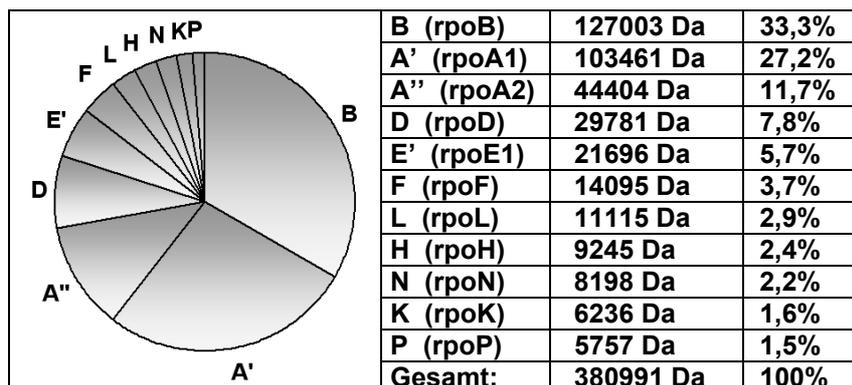


Abb.4-01: Anteil der einzelnen RNA-Pol-Untereinheiten am Gesamtkomplex

Das Molekulargewicht für Untereinheit A' ist aufgrund eines korrigierten Startpunktes geringfügig höher, als entsprechend dem annotierte Gen PF1563 zu erwarten wäre (siehe [8.1.2.2]). E'' ist in diese Berechnung nicht mit einbezogen worden, weil noch nicht sicher ist, ob dieses Protein tatsächlich im RNAP-Komplex enthalten ist.

Mit rund 381 kDa besitzt die *Pyrococcus*-RNAP eine ähnlich Größe wie beispielsweise *E.coli* (383 kDa, $\beta'\beta\alpha_2\omega$ -Komplex), und ist somit deutlich kleiner als die eukaryotische Pol II aus *S.cerevisiae* (Pol I: 593kDa, Pol II: 598kDa, Pol III: 640kDa). Das ist insofern überraschend, als *Bacteria* einen auf den ersten Blick stark abweichenden Aufbau der RNA-Pol gegenüber *Eukarya* und *Archaea*

zeigen (siehe [Abb.2-02]). Ursächlich für das geringere Molekulargewicht des archaeellen RNA-Pol-Komplexes ist das Fehlen von zu Rpb8 und Rpb9 homologen Proteinen, sowie von größeren Abschnitten verschiedener anderer Untereinheiten. Die theoretische Molekülmasse von Pol II aus *S.cerevisiae* beträgt ohne posttranslationale Modifikationen allerdings nur 514kDa. Folgendes Diagramm zeigt, auf welchen Untereinheiten die Molekulargewichtsdifferenz beruht:

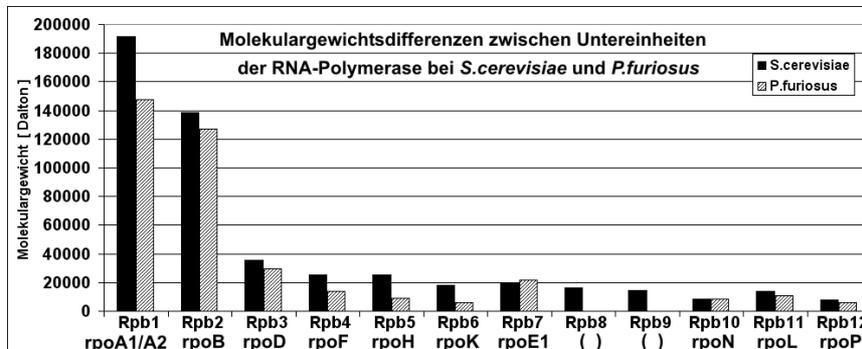


Abb.4-02: Größen der homologen RNA-Pol-Untereinheiten bei *P.furiosus* und *S.cerevisiae*

Bis auf Rpb7 und Rpb10 sind alle RNA-Pol-Untereinheiten aus *S.cerevisiae* kleiner als ihr Homolog in *P.furiosus* (RNA-Pol-Gene aus *S.cerevisiae* zur Berechnung des theoretischen MG siehe [8.6]).

Folgende Tabelle zeigt weitere Details zu den Genen der RNAP-Untereinheiten in *P.furiosus*:

UE	Gen	Genom	Position	aa	MW	Ladung	pl
B	rpoB	PF1564	(-) 1459639..1462992	1117	127004,2	- 10 (-5 / -5)	6.70
A'	rpoA1	PF1563	(-) 1456903..1459635	910	103461,3	- 6 (+4 -11)	6.83
A''	rpoA2	PF1562	(-) 1455699..1456892	397	44404,3	- 7 (-9 / +2)	5.86
D	rpoD	PF1647	(-) 1534790..1535575	261	29781,3	-14 (-13 / -1)	4.71
E'	rpoE1	PF0256	(-) 263488..264057	189	21696,5	0 (-8 / +8)	7.76
E''	rpoE2	PF0255	(-) 263306..263491	61	6953,0	-1 (-3 / +2)	7.20
F	rpoF	PF1036	(+) 991090..991452	120	14095,3	- 8 (+1 / -9)	4.66
H	rpoH	PF1565	(-) 1463013..1463261	82	9244,8	+2 (0 / +2)	9.69
K	rpoK	PF1642	(-) 1532854..1533027	57	6236,5	+2 (+4 / -2)	10.34
L	rpoL	PF0050	(-) 57460..57747	95	11114,8	- 5 (-5 / 0)	5.15
N	rpoN	PF1643	(-) 1533155..1533367	70	8198,6	- 1 (+1 / -2)	6.49
P	rpoP	PF2009	(+) 1857791..1857940	49	5757,0	+9 (+1 / +8)	10.73

Abb.4-03: Gene und allgemeine Eigenschaften der RNAP-Untereinheiten in *P.furiosus*

UE = Untereinheit der DNA-abhängigen RNA-Polymerase bei *Pyrococcus furiosus*; PFxxxx: fortlaufender Name der Gene im Genom (Locus; diese Bezeichnungen können in gleicher Weise wie Accession-Nummern zum Auffinden der genannten Gene in den einschlägigen Online-Datenbanken verwendet werden); (-) / (+) = Orientierung des Genes im Genom; aa = Anzahl der Aminosäuren (amino acids); pl = Isoelektrischer Punkt. Der Translationsstart von **rpoA1** (PF1563) wurde um 3 Aminosäurepositionen nach vorn korrigiert, so dass das Protein 910 statt 907 Aminosäuren lang ist (es werden die korrigierten Daten gezeigt). **Ladung**: Theoretische Nettoladung der Proteine bei einem Umgebungs-pH von 7,6 (die beiden Werte in Klammern sind die Ladung der N-terminalen und der C-terminalen Hälfte des Proteins als grober Richtwert für die Ladungsverteilung).

4.2 Untereinheiten der RNA-Polymerase in archaeellen Genomen

Während praktische Experimente im Rahmen dieser Arbeit ausschließlich mit der RNA-Polymerase aus *P.furiosus* (und deren Untereinheiten, sowie Transkriptionsfaktoren) durchgeführt wurden, diente die theoretisch-bioinformatische Analyse der entsprechenden Gene bei anderen Organismen einerseits der Überprüfung der gewonnenen Daten, und andererseits als übergreifender Kontext. Vor diesem Hintergrund können die speziellen Details bei *P.furiosus* diskutiert werden und Abweichungen erkannt werden.

Mit 17 komplett sequenzierten Genomen ist die Situation bei *Archaea* gegenwärtig noch recht übersichtlich, während bei *Bacteria* heute schon 127 Genome zugänglich sind (die Zahl der eukaryotischen Genom-Projekte ist wegen des viel größeren Umfangs und der komplexeren Struktur der Genome noch geringer). In dieser Arbeit wird zu einem Vergleich mit *Eukarya* zumeist

der Modell-Organismus *Saccharomyces cerevisiae* (Bäckerhefe) herangezogen, dessen aus 16 Chromosomen bestehendes Genom eines der bisher am genauesten untersuchten ist. (Eine Übersicht über bisher komplett sequenzierte archaelle und bakterielle Genome siehe auch: <http://www.ncbi.nlm.nih.gov:80/genomes/MICROBES/Complete.html>).

Archaea
Crenarchaeota
<i>Aeropyrum pernix K1</i> : (Desulfurococcales) (Kawarabayasi <i>et al.</i> , 1999)
<i>Sulfolobus solfataricus</i> : (Sulfolobales) (Zillig <i>et al.</i> , 1980)
<i>Sulfolobus tokodaii</i> : (Sulfolobales) (Kawarabayasi <i>et al.</i> , 2001)
<i>Pyrobaculum aerophilum</i> : (Thermoproteales) (Fitz-Gibbon <i>et al.</i> , 2002)
Euryarchaeota
<i>Halobacterium sp. NRC-1</i> : (Halobacteriales) (Ng <i>et al.</i> , 2000)
<i>Methanocaldococcus jannaschii</i> : (Methanococcales) (Bult <i>et al.</i> , 1996)
<i>Methanopyrus kandleri AV19</i> : (Methanopyrales) (Slesarev <i>et al.</i> , 2002)
<i>Methanosarcina acetivorans C2A</i> : (Euryarchaeota / Methanosarcinales) (Galagan <i>et al.</i> , 2002)
<i>Methanosarcina mazei Goel</i> : (Euryarchaeota / Methanosarcinales) (Deppenmeier <i>et al.</i> , 2002)
<i>Methanothermobacter thermoautotrophicus str. Delta H</i> : (Methanobacteriales) (Smith <i>et al.</i> , 1997)
<i>Pyrococcus abyssi</i> : (Thermococcales) (Cohen <i>et al.</i> , 2003)
<i>Pyrococcus furiosus DSM 3638</i> : (Thermococcales) (Robb <i>et al.</i> , 2001)
<i>Pyrococcus horikoshii</i> : (Thermococcales) (Kawarabayasi <i>et al.</i> , 1998)
<i>Thermoplasma acidophilum</i> : (Euryarchaeota / Thermoplasmatales) (Ruepp <i>et al.</i> , 2000)
<i>Thermoplasma volcanium</i> : (Euryarchaeota / Thermoplasmatales) (Kawashima <i>et al.</i> , 2000)
Nanoarchaeota
<i>Nanoarchaeum equitans Kin4-M</i> (Waters <i>et al.</i> , 2003); (das Phylum "Nanoarchaeota wurde kürzlich neu postuliert)
Korarchaeota
(auf RNA-Basis definiert; bisher keine kultivierbaren Vertreter, bzw. Mischkulturen)

Abb.4-04: Liste der gegenwärtig verfügbaren, komplett sequenzierten archaellen Genome

Die angegebenen Publikationen beziehen sich auf die Veröffentlichung des kompletten Genomes und nicht auf die Erstbeschreibung des jeweiligen Organismus.

Die Daten zu allen RNAP-Untereinheiten der hier aufgelisteten *Archaea* finden sich in der folgenden Tabelle. Die Quelle der Daten sind die für jede komplett sequenzierte Art erhältlichen Genomdateien im Genbankformat (siehe [8.5]), die mehr oder weniger die gleichen Daten enthalten, die auch in den einschlägigen Online-Gendatenbanken bereitgestellt werden. Um die Extraktion und Aufbereitung der Informationen über die RNAP-Untereinheiten und andere Daten aus diesen Genomdateien zu beschleunigen, wurde ein zu diesem Zweck nützliches Programm geschrieben (siehe [8.4]).

4.2.1 Zusammenstellung von Genomdaten für eine vergleichende Darstellung

Die folgende Tabelle zeigt die RNA-Pol-Gene der gegenwärtig 17 komplett sequenzierten *Archaea*. Die jeweilige Bezeichnung für das Gen entspricht den Angaben in den bei NCBI zugänglichen Genbank-Dateien der aufgeführten archaeellen Genome. Es kann mit dieser Bezeichnung ähnlich wie mit einer Accession-Nummer oder einer Protein-Datenbank-ID innerhalb von NCBI oder SRS (unter „Combined Sequence Databases“) gesucht werden.

	B / B'	B''	A'	A''	D	E'	E''	F	H	K	L	N	P
<i>Aeropyrum pernix</i>	APE1856 132379Da	--	APE1853 101741Da	APE1852 46025Da	APE1744 32058Da	APE0257 20712Da	APES005 6143Da	?	APES061 10368Da	APE0914 11437Da	APE0443 11930Da	APES058 8843Da	APES057 8692Da
<i>Archaeoglobus fulgidus</i>	rpoB1 67765Da	rpoB2 58544Da	rpoA1 97807Da	rpoA2 42422Da	rpoD 29568Da	rpoE1 21686Da	rpoE2 7037Da	AF1530 14026Da	rpoH 8510Da	rpoK 8870Da	rpoL 9757Da	rpoN 8695Da	AF0056 5063Da
<i>Halobacterium sp. NRC-1 *</i>	rpoB' 67276Da	rpoB'' 59397Da	rpoA1 108734Da	rpoC 43750Da	rpB3 27377Da	rpoE' 20796Da	rpoE'' 7339Da	VNG1169C 13648Da	rpoH 8507Da	rpoK 6654Da	rpoL 10010Da	rpoN 7443Da	VNG0237H 5179Da
<i>Methanocaldococcus jannaschii **</i>	MJ1041 72193Da	MJ1040 56487Da	MJ1042 152781Da	MJ1043 97087Da	MJ0192 21699Da	MJ0397 21216Da	MJ0396 6578Da	MJ0039 13390Da	MJ1039 9001Da	MJ0197 6292Da	MJ0387 11392Da	MJ0196 8695Da	RPC10 5452Da
<i>Methanopyrus kandleri AV19</i>	RpoB_1 72786Da	RpoB_2 61121Da	RpoC_2 108859Da	RpoC_1 48034Da	RpoA 31526Da	RpoE1 20755Da	RpoE2 7188Da	MK1659 14965Da	RPB5 10282Da	RpoZ 6235Da	RPB11 13024Da	RPB10 8423Da	RPC10 6317Da
<i>Methanosarcina acetivorans C2A ***</i>	rpoB1 66979Da	rpoB2 62716Da	rpoA1 98372Da	rpoA2 44157Da	(rpoD) 28749Da	rpoE1 21392Da	rpoE2 6844Da	MA1444 13677Da	rpoH 8811Da	rpoK 6619Da	rpoL 10374Da	rpoN 7163Da	rpoP 5379Da
<i>Methanosarcina mazei Goe1</i>	MM2272 67023Da	MM2273 60584Da	MM2271 98259Da	MM2270 45332Da	MM2158 28896Da	MM0596 21432Da	MM0597 6881Da	MM2537 12273Da	MM2274 9143Da	MM1759 10371Da	MM1880 10371Da	MM1758 7219Da	MM2626 5379Da
<i>Methanothermobacter thermautotrophicus</i>	MTH1050 67688Da	MTH1049 59191Da	MTH1051 98078Da	MTH1052 51004Da	MTH37 29395Da	MTH264 20695Da	MTH265 6735Da	MTH1324 13479Da	MTH1048 8808Da	MTH42 7043Da	MTH1317 9732Da	MTH40 6435Da	MTH680a 5809Da
<i>Nanoarchaeum equitans Kin4-M</i>	NEK170 67950Da	NEK188 56657Da	NEK547 100057Da	NEK465 42525Da	NEK492 26735Da	NEK401 20580Da	(rpoE2) 7216Da	?	NEK551 10794Da	NEK223 8036Da	NEK197 11180Da	NEK365 7459Da	(rpoP) 5294Da
<i>Pyrobaculum aerophilum</i>	PAE0665 126481Da	--	PAE0666 100119Da	PAE0667 42423Da	PAE0671 29291Da	PAE3562 21282Da	PAE3563 8297Da	PAE3335 12875Da	PAE0664 8331Da	PAE3271 12190Da	PAE1978 10678Da	PAE0675a 7717Da	PAE2258 6168Da
<i>Pyrococcus abyssi</i>	rpoB 127059Da	--	rpoA1 103057Da	rpoA2 44595Da	rpoD 29477Da	rpoE1 21456Da	rpoE2 6799Da	PAB0732 14136Da	rpoH 9284Da	rpoK 6271Da	rpoL 11081Da	rpoN 7787Da	PAB3072 5681Da
<i>Pyrococcus furiosus DSM 3638</i>	PF1564 127004Da	--	PF1563 103115Da	PF1562 44404Da	PF1647 29781Da	PF0256 21696Da	PF0255 6953Da	PF1036 14095Da	PF1565 9245Da	PF1642 6237Da	PF0050 11115Da	PF1643 8199Da	PF2009 5757Da
<i>Pyrococcus horikoshii</i>	PH1546 127089Da	--	PH1545 102977Da	PH1544 44508Da	PH1637 29886Da	PH1908 21512Da	PH1908a 6972Da	PH1127 14278Da	PHS044 9355Da	PH1631a 6271Da	PH0041a 11144Da	PH1632 13881Da	PHS056 8993Da
<i>Sulfolobus solfataricus ****</i>	rpoB1 73124Da	rpoB2 54035Da	rpoA1 99675Da	rpoA2 43805Da	rpoD 30294Da	rpoE1 20329Da	rpoE2 7341Da	rpoF 12811Da	rpoH 9670Da	rpoK 10858Da	rpoL 10285Da	rpoN 7591Da	rpoP 5597Da
<i>Sulfolobus tokodaii</i>	ST0277 126576Da	--	ST0276 100547Da	ST0275 44760Da	ST2067 29620Da	ST0379 20579Da	STS052 7372Da	ST0458 13328Da	STS038 11113Da	STS136 9816Da	STS238 10454Da	STS221 7619Da	ST0351 11790Da
<i>Thermoplasma acidophilum</i>	Ta0390 134691Da	--	Ta0391 100075Da	Ta0392 56946Da	Ta1030 30419Da	Ta1089m 21723Da	Ta1090 7330Da	Ta1298 12028Da	Ta0389 10384Da	Ta1160 6779Da	Ta1030 30419Da	Ta0431 8368Da	RPC10 5151Da
<i>Thermoplasma volcanium</i>	TVN1182 134557Da	--	TVN1181 100024Da	TVN1180 57506Da	TVN0565 30297Da	TVN0486 21446Da	TVN0485 7374Da	TVN0313 11636Da	TVN1183 9889Da	TVN1225 6791Da	TVN1435 13089Da	TVN1137 8483Da	TVN0310 5195Da

Abb.4-05: Tabelle mit den Namen der Gene und den theoretischen Molekulargewichten der RNA-Pol-Untereinheiten bei 17 komplett sequenzierten *Archaea*

* Das Genom von *N. equitans* liegt komplett sequenziert vor – seit dem 17.11.03 sogar in einer annotierten Version; evtl. handelt es sich aber noch um eine vorläufige Version – statt „gene“ wurde „locus_tag“ zur Bezeichnung der Gene verwendet (siehe [8.5]). rpoE2 wurde in dieser Version nicht als Gen erkannt. Mittels einer „lokalen BLAST-Suche“ (siehe [8.3]) ließ sich das Gen jedoch identifizieren. Dies gilt auch für rpoP.

** rpoA1 aus *M. jannaschii* enthält ein 451 Aminosäuren langes Intein (Position 461 bis 911; ohne Intein 100538Da). Auch rpoA2 enthält ein 471 Aminosäuren langes Intein (Position 76 bis 546; ohne Intein 43388Da)

*** Auch rpoD aus *M. acetivorans* ist bei der Annotierung des Genoms nicht als Gen erkannt worden. Die gegenüber *M. mazei* stark konservierte Sequenz konnte erst durch „lokale BLAST-Suche“ (s.o.) entdeckt werden. Evtl. wurde das Gen nicht erkannt, weil es eine, durch einen Sequenzierfehler bedingte, scheinbare Verschiebung des Leserasters enthält.

**** Die Gattung *Sulfolobus* besitzt normalerweise eine ungeteilte B-Untereinheit. Verschiedene Indizien deuten darauf hin, dass es sich in diesem Fall nicht um einen Sequenzierfehler, sondern tatsächlich um eine Besonderheit handelt, die in *S. solfataricus* unabhängig von den andern *Archaea* mit geteilter B-Untereinheit im Laufe der Evolution aufgetreten ist.

4.2.2 Zuverlässigkeit und Aussagekraft von Genomdaten / bioinformatischen Analysen

Identifizierung von Genen im Genom: Bis zu einem gewissen Grad hypothetisch!

Bei dem Inhalt der Genbank-Dateien mit dem Genom eines Mikroorganismus handelt es sich einerseits um die reine DNA-Sequenz, die (abgesehen von vereinzelt Sequenzierfehlern) zweifelsfrei feststeht. Andererseits handelt es sich um eine Liste mit identifizierten Genen, die jeweils mit Aminosäuresequenz, Name, Bemerkungen und Referenzen etc. versehen die Basis für Genom-Recherchen (sowohl lokal, als auch online) darstellen – beispielsweise über eine Stichwort- oder BLAST-Suche.

Es muß dem Anwender klar sein, dass es sich bei letztgenannten Daten um eine Interpretation handelt, die je nach der Kombination der verwendeten Programme (z.B. Glimmer, GeneScan, GeneMark), dem aktuellen Stand der großen Gen- und Protein-Datenbanken (also der Anzahl und Art der verfügbaren Datensätze), sowie abhängig von einer mehr oder weniger sorgfältige Nachbearbeitung unterschiedliche Ergebnisse liefern kann.

Prinzipiell sind zwei Schritte nötig: **1.** Korrekte Identifizierung eines ORF = "open reading frame" (durchaus keine triviale Angelegenheit, wenn keinerlei Vorinformationen vorhanden sind; jedes Programm verfolgt seine eigene Strategie, die etwas abweichende Ergebnisse liefern kann; neuere bioinformatische Ansätze orientieren sich beispielsweise auch an potentielle Ribosomen-Bindestellen und weiteren Indizien (Lukashin & Borodovsky, 1998; Aggarwal & Ramaswamy, 2002) etc.).

2. Zuordnung der gefundenen (und in Aminosäuresequenz übersetzten) ORFs zu den schon in den Datenbanken vorhandenen Proteinen (bzw. DNA bei Nicht-Proteingenen). Letztlich wird hier für jedes potentielle Gen eine BLAST-Suche durchgeführt. Je nach den gewählten Einstellungen erfolgt bei genügend eindeutiger Homologie der Eintrag in die o.g. Genbank-Datei.

Dies ist genau der Grund, weshalb sich bei *P.furiosus* bei 2065 Proteingenen in 1026 Fällen der Eintrag "hypothetical protein" findet. Allerdings muß bedacht werden, dass die Genbank-Dateien auch veralten können – evtl. sind durch weitere Forschung und manuelle Annotierung einige Gene nicht mehr „hypothetisch“.

Die u.a. in Tabelle [Abb.4-06] angegebene Gesamtzahl an Proteingenen ist also ein Richtwert. Ein Beispiel verdeutlicht den Sachverhalt:

Deinococcus radiodurans: Original Genbank-Datei AE000513.gbk mit 2579 Proteingenen. Hiervon abgeleitet die aktuelle, von NCBI bearbeitete Datei NC_001263.gbk mit 2629 Proteingenen.

Ein typischer Fehler ist auch die Angabe eines zu frühen oder zu späten Start-Punktes des Genes. Selbst beim Stop-Codon kann man nicht ganz sicher sein: In einigen Fällen wird das Codon "TGA" (bzw. "UGA" auf mRNA-Ebene) nicht als Stop-Codon, sondern als spezielles Codon mitten im offenen Leserahmen interpretiert, das für den Einbau einer seltenen Aminosäure, dem "Selenocystein" codiert. Entsprechende Beispiele sind sowohl für Organismen aus *Eukarya* und *Bacteria*, als auch für *Archaea* beschrieben (Wilting *et al.*, 1997; Rother *et al.*, 2001).

Insgesamt zeigt sich, dass viel "Handarbeit" bei der Analyse von Einzelfällen und Ausnahmen nötig ist, die sonst von den (aufgrund der immer größeren Datenmengen nötigen) automatischen Bearbeitungssystemen falsch zugeordnet werden.

Wichtig ist also, die vorliegenden Genomdaten permanent kritisch zu hinterfragen, um diese ab und zu auftretenden Fehler zu entdecken.

Folgende Tabelle enthält ergänzende Daten zu den *Archaea*, deren RNAP-Gene in [Abb.4-05] aufgelistet werden:

4.2.3 Weitere Eckdaten der untersuchten archaeellen Genome (Statistik)

	Dateiname	bp DNA	Gene	GC %	Start-Codon	Stop-Codon	ohne Überschneidung	mit Überschneidung
<i>Aeropyrum pernix</i>	NC_000854.gbk	1669695	1907 1841 66	56,3	ATG: 780(42%) GTG: 1041(57%) TTG: 17(1%)	TGA: 424(23%) TAA: 412(22%) TAG: 1005(55%)	851 / 508	237 / 310
<i>Archaeoglobus fulgidus</i>	NC_000917.gbk	2178400	2424 2420 4 ?	48,6	ATG: 1843(76%) GTG: 520(21%) TTG: 55(2%)	TGA: 1230(51%) TAA: 815(34%) TAG: 373(15%)	1047 / 499	662 / 215
<i>Halobacterium sp. NRC-1 *</i>	NC_002607.gbk	2014239	2149 2075 74	67,9	ATG: 2019(97%) GTG: 49(2%) TTG: 2(0%)	TGA: 1045(50%) TAA: 426(21%) TAG: 604(29%)	977 / 763	285 / 123
<i>Methanocaldococcus jannaschii *</i>	NC_000909.gbk	1664970	1772 1729 43	31,4	ATG: 1209(70%) GTG: 240(14%) TTG: 278(16%)	TGA: 222(13%) TAA: 1343(78%) TAG: 164(9%)	1047 / 458	208 / 58
<i>Methanopyrus kandleri AV19</i>	NC_003551.gbk	1694969	1733 1691 42	61,2	ATG: 502(30%) GTG: 588(35%) TTG: 517(31%)	TGA: 982(58%) TAA: 388(23%) TAG: 321(19%)	821 / 435	334 / 142

<i>Methanosarcina acetivorans</i> C2A*	NC_003552.gb	5751492	4550 4540 10	42,7	ATG: 2925(64%) GTG: 651(14%) TTG: 831(18%)	TGA: 2099(46%) TAA: 2217(49%) TAG: 224(5%)	2559 / 1477	470 / 43
<i>Methanosarcina mazei</i> Goe1	NC_003901.gb	4096345	3438 3371 67	41,5	ATG: 2372(70%) GTG: 441(13%) TTG: 558(17%)	TGA: 1445(43%) TAA: 1800(53%) TAG: 126(4%)	2061 / 1071	289 / 16
<i>Methanothermobacter thermautotrophicus</i>	NC_000916.gb	1751377	1920 1873 47	49,5	ATG: 1173(63%) GTG: 420(22%) TTG: 276(15%)	TGA: 889(47%) TAA: 554(30%) TAG: 430(23%)	1143 / 462	294 / 20
<i>Nanoarchaeum equitans</i> Kin4-M	NC_005213	490885	563 563 0 ?!	31,6	ATG: 471(84%) GTG: 56(10%) TTG: 36(6%)	TGA: 113(20%) TAA: 319(57%) TAG: 131(23%)	165 / 157	145 / 95
<i>Pyrobaculum aerophilum</i>	NC_003364.gb	2222430	2701 2605 96	51,4	ATG: 1618(62%) GTG: 884(34%) TTG: 103(4%)	TGA: 685(26%) TAA: 1140(44%) TAG: 780(30%)	1116 / 753	585 / 246
<i>Pyrococcus abyssi</i>	NC_000868.gb	1765118	1769 1769 0	44,7	ATG: 1317(74%) GTG: 287(16%) TTG: 161(9%)	TGA: 908(51%) TAA: 513(29%) TAG: 348(20%)	817 / 371	401 / 179
<i>Pyrococcus furiosus</i> DSM 3638	NC_003413.gb	1908256	2177 2065 112	40,8	ATG: 1739(84%) GTG: 276(13%) TTG: 50(2%)	TGA: 956(46%) TAA: 689(33%) TAG: 420(20%)	1011 / 456	481 / 228
<i>Pyrococcus horikoshii</i>	NC_000961.gb	1738505	1858 1801 57	41,9	ATG: 1291(72%) GTG: 505(28%) TTG: 4(0%)	TGA: 847(47%) TAA: 606(34%) TAG: 348(19%)	812 / 419	365 / 261
<i>Sulfolobus solfataricus</i>	NC_002754.gb	2992245	3027 2977 50	35,8	ATG: 1881(63%) GTG: 543(18%) TTG: 553(19%)	TGA: 875(29%) TAA: 1484(50%) TAG: 618(21%)	1275 / 873	638 / 240
<i>Sulfolobus tokodaii</i>	NC_003106.gb	2694756	2940 2826 114	32,8	ATG: 2170(77%) GTG: 655(23%) TTG: 1(0%)	TGA: 812(29%) TAA: 1532(54%) TAG: 482(17%)	1185 / 782	590 / 382
<i>Thermoplasma acidophilum</i>	NC_002578.gb	1564906	1530 1482 48	46,0	ATG: 1044(70%) GTG: 238(16%) TTG: 197(13%)	TGA: 953(64%) TAA: 331(22%) TAG: 198(13%)	717 / 512	237 / 63
<i>Thermoplasma volcanium</i>	NC_002689.gb	1584804	1548 1499 49	39,9	ATG: 1250(83%) GTG: 132(9%) TTG: 108(7%)	TGA: 575(38%) TAA: 673(45%) TAG: 251(17%)	757 / 485	241 / 64
<i>Escherichia coli</i> CFT073	NC_004431	5231428	5557 5379 178	50,5	ATG: 3641(68%) GTG: 927(17%) TTG: 810(15%)	TGA: 1736(32%) TAA: 3217(60%) TAG: 426(8%)	2576 / 1375	1060 / 545
* die Extrachromosomalen Elemente (Plasmide) von <i>M.jannaschii</i> und <i>Halobacterium</i> wurden nicht mit in die Übersicht einbezogen								

Abb.4-06: Tabellarische Darstellung statistischer Rahmendaten der archaeellen Genome

Genomdaten der gegenwärtig 17 komplett sequenzierten *Archaea* (*E.coli* als bakterielle Referenz); Quelle der Daten sind die von NCBI einheitlich nachbearbeiteten Genomdateien im Genbankformat (NC_XXXXX.gb). **bp DNA**: Länge des jeweiligen Genoms in Nukleotiden; **Gene**: Gene insgesamt / Proteingene / Nicht-Proteingene; **GC %**: Exakter GC-Gehalt auf Basis der Genomsequenz; **ohne Überschneidung**: Aufeinander folgende Gene ohne / mit Wechsel der Leserichtung; **mit Überschneidung**: Aufeinander folgende, sich partielle überschneidende Gene ohne / mit Wechsel der Leserichtung. Bei der Angabe der Anzahl der "Nicht-Proteingene" sind einige Zahlen unrealistisch klein – diese Werte wurden mit einem Fragezeichen markiert.

Die aufgeführten archaeellen Genome dienen als Basis für die im Anhang [8.1.3] gezeigten umfangreichen Alignments, sowie weitere Analysen. Die Auswertung der Häufigkeit von Start- und Stopcodons kann zum Beispiel nützlich bei der Identifizierung alternativer Translationsstartpunkte sein. Die überraschend hohen Werte für die Anzahl einander überlappender Gene bei *Archaea* und *Bacteria* bestätigen (selbst wenn ein Teil der Einträge durch fehlerhafte Zuordnung der Gene innerhalb der Genbankdateien verursacht würde) die in [4.5] dargestellte Analyse der "translationalen Kopplung" von Genen am Beispiel der Gene von rpoE1 und rpoE2.

4.3 Auswertung von Alignments der drei *Pyrococcus*-Arten untereinander

Da zum gegenwärtigen Zeitpunkt bereits die Genomdaten von drei komplett sequenzierten *Pyrococcus*-Arten verfügbar sind (siehe [8.5]), bietet sich ein Vergleich auf kurze evolutionäre Distanz an. Einerseits ist die Frage interessant, ob schon hier gravierende Unterschiede erkennbar sind. Andererseits läßt sich die Frage nach den Verwandtschaftsverhältnissen innerhalb der Gattung *Pyrococcus* diskutieren.

Zunächst einige Anmerkungen zu den Alignments, die im Anhang unter ([8.1.2]) zu finden sind:

Das Alignment mit den beiden anderen *Pyrococcus*-Arten legt bei *Pfu_rpoA1* einen alternativen Startpunkt bei dem Methionin drei Positionen früher als in PF1563 annotiert nahe. Auch die Analyse der potentiellen Ribosomen-Bindestelle spricht für diese Korrektur. Die Primer für die Klonierung dieser RNAP-Untereinheit in pET-14b wurden entsprechend diesem früheren Startpunkt ausgewählt. Das Valin an Position 4 bei *P.abysssi* und *P.horikoshii* kommt nicht als möglicher Start in Frage, weil er jeweils durch das DNA-Triplett „GTT“ (und nicht durch „GTG“ = alternatives Startcodon bei *Archaea* und *Bacteria*) kodiert wird. Die A'-Untereinheit enthält bei *Archaea* das katalytische Zentrum der RNA-Polymerase, also Aminosäuren, die besonders mit der DNA / RNA interagieren. Innerhalb der Sequenz „NADFDGD“ (Pos 456 ff.) (Dieci *et al.*, 1995; Sonntag & Darai, 1995) ist das Motiv D^{458,500,502} (also 3 Aspartatreste) für die Bindung des Mg²⁺-Ions im katalytischen Zentrum verantwortlich. Erwartungsgemäß sind die Aminosäuresequenzen der hier untersuchten 3 *Pyrococcus*-Arten in diesem Bereich identisch.

Die Untereinheit E'' (rpoE2) wurde längere Zeit für einen vermutlich in vivo nicht exprimierten, deaktivierten Teil der ehemals größeren Untereinheit E' gehalten, der durch eine einzelne Deletion im Leseraster verschoben wurde und um -1 Nucleotid-Positionen mit E' überlappt (-4, wenn man das Stop-Codon von rpoE1 mitzählt). Wie auch aus der unten gezeigten Tabelle zu entnehmen, ist die Mutationsrate aber für ein vermeintlich inaktiviertes Gen zu gering, was auf einen nach wie vor wirkenden, positiven Selektionsdruck auf das resultierende Genprodukt schließen läßt. *Pab_rpoE2* ist ohne Methionin am N-Terminus annotiert worden. Ein „ORF ohne Startcodon“ wäre ein gutes Argument für ein im Laufe der Evolution inaktiviertes Gen – allerdings ist das Codon für die direkt vorangehende Aminosäure „GTG“, welches als alternativer Start dienen kann und dann für Methionin kodiert. Dass bei einer Punktmutation des Startcodons ausgerechnet (als eine Art „stille Mutation“) das häufigste alternative Startcodon entsteht, könnte als weiteres Indiz für die Notwendigkeit der E''-Untereinheit für den Stoffwechsel des Organismus gewertet werden (allerdings dient bei *P.furiosus* in ca. 2% der Fälle auch „TTG“ als Start-Codon; siehe [4.2.3]). Erst das „große Alignment“ mit anderen Archaea räumt letzte Unsicherheiten in dieser Frage aus (siehe [8.1.3.6]).

Bei *Pab_rpoN* ist „TTG“ als Startcodon interpretiert worden – eigene Analysen zeigten, dass „TTG“ als alternatives Start-Codon bei Archaea weit verbreitet ist und teilweise (z.B. bei *M.jannaschii*) sogar häufiger als „GTG“ verwendet wird (siehe [Abb.4-06]). Das Valin an Position 3 kommt für einen alternativen Startpunkt dagegen nicht in Frage, weil das Codon „GTC“ lautet. Hier könnte man dann auch gleich den Start von *Pfu_rpoN* anzweifeln: Das Valin an Position 6 wird durch „GTG“ kodiert – hier startet auch *Pab_rpoN*. Für die Untersuchungen in dieser Arbeit habe ich mich an die oben dargestellte lange, annotierte Version gehalten. Dass *P.horikoshii* in bezug auf das ansonsten besonders konservierte rpoN über eine zusätzliche, 51 Aminosäuren lange, N-terminale Zusatzsequenz verfügen soll, kann getrost angezweifelt werden. Die normalerweise nur knapp 70 Aminosäuren lange Untereinheit N gehört bisherigen Forschungsergebnissen zufolge zu den kleinen Proteinen, die den DLNP-Komplex als eine Plattform für den weiteren Zusammenbau der RNA-Polymerase bilden (Werner & Weinzierl, 2002). Dass 51 zusätzliche Aminosäuren für diesen sensiblen, und für den Organismus überlebenswichtigen Vorgang einen erheblichen Störfaktor darstellen würden, ist sehr wahrscheinlich. Das Valin an Position 52 des *Pho_rpoN*-Genes (2 Aminosäuren vor dem Start des Genes bei *P.abysssi*) wird durch „GTG“ kodiert und repräsentiert wahrscheinlich den richtigen Startpunkt.

Pho_rpoP (PHS056) ist nicht korrekt annotiert: Die angegebene Aminosäuresequenz überschneidet sich um 25aa mit PHS057 („50S ribosomal protein L37“). Da das Valin an Position 17 durch „GTA“ kodiert wird, also als Transkriptionsstart nicht in Frage kommt, bleibt nur das Methionin an Pos. 19, wo auch die entsprechenden Gene von *P.furiosus* und *P.abysssi* beginnen.

Die Auswertung der beschriebenen Alignments führte zu folgender, zusammenfassenden Tabelle:

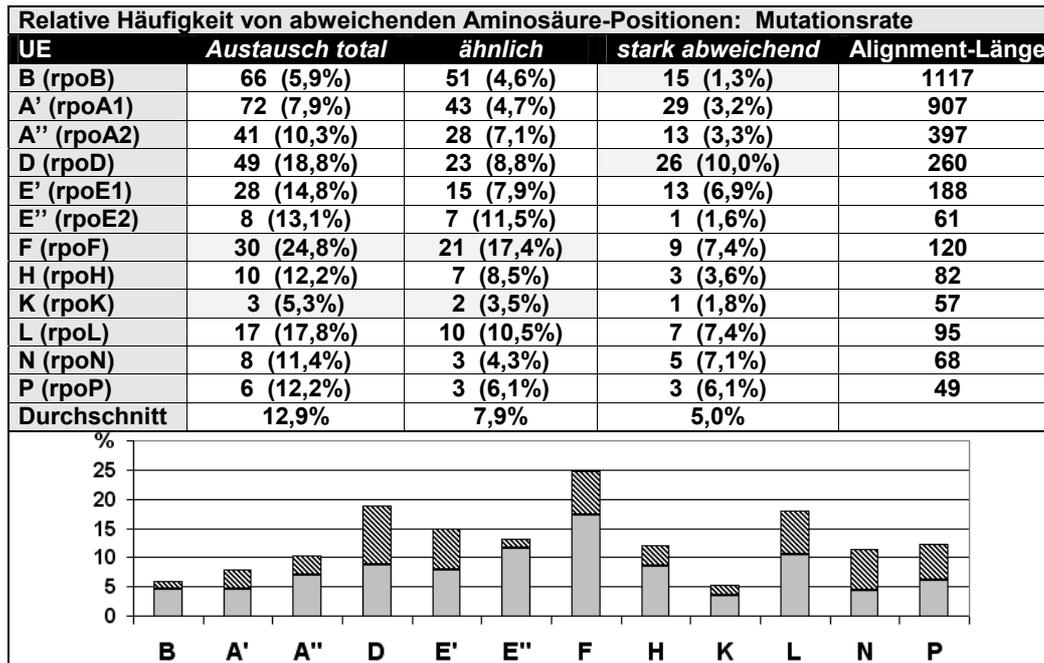


Abb.4-07: Aminosäure-Austausch in den RNAP-Genen innerhalb der Gattung *Pyrococcus*

Tabellarische Darstellung und grafische Darstellung der Häufigkeit von Aminosäure-Austausch. Ähnlich: hellgrau, stark abweichend: schraffiert. In der Tabelle wurde die niedrigsten und höchsten Werte grau hinterlegt.

Diese Tabelle gibt erste Antworten auf die Frage, ob auf den verschiedenen RNAP-Untereinheiten ein unterschiedlicher Selektionsdruck lastet, der sich in einer unterschiedlichen Mutationsrate äußert. Diese Übersicht liefert einen groben Überblick über Veränderung einzelnen Proteine im Ganzen, während in folgenden Abschnitten dieser Arbeit auch auf die stärker oder schwächer konservierten Regionen im Detail eingegangen wird

Anzahl der jeweils von den beiden anderen Arten abweichenden Aminosäure-Positionen:				
UE	<i>P.furiosus</i>	<i>P.abysssi</i>	<i>P.horikoshii</i>	Alignment-Länge
B (rpoB)	40	11	11	1117
A' (rpoA1)	48	5	14	907
A'' (rpoA2)	26	2	9	397
D (rpoD)	27	4	6	260
E' (rpoE1)	18	4	3	188
E'' (rpoE2)	5	2	1	61
F (rpoF)	15	3	8	120
H (rpoH)	5	1	3	82
K (rpoK)	3	-	-	57
L (rpoL)	6	4	6	95
N (rpoN)	-	1	2	68
P (rpoP)	2	1	2	49
Summe	195	38	65	3401
Summe %	5,7%	1,1%	1,9%	

Abb.4-08: Aminosäure-Austausch an Positionen, die bei den anderen Schwesterarten identisch sind

Es wurden alle Positionen mit einbezogen, an denen die Aminosäure im Alignment von denen bei den anderen beiden Arten abweicht. Die wenigen Stellen, an denen bei allen drei Arten unterschiedliche Aminosäuren zu finden sind, wurden ignoriert. Es wurde kein Unterschied gemacht, ob es sich um einen Austausch mit einer ähnlichen oder stark abweichenden Aminosäure handelt.

Die gezeigte Auswertung stimmt recht gut mit bereits bekannten Vergleichsdaten in Bezug auf die untersuchten drei *Pyrococcus*-Arten überein (Lecompte *et al.*, 2001): *P.furiosus* weicht mit 5,7% innerhalb der RNA-Pol-Gene am stärksten ab, während *P.abysssi* den beiden anderen Arten

offensichtlich am nächsten steht (1,1% Abweichung). Dies spiegelt u.a. die geographische Distanz der Fundorte wider: Während *P.abysyi* (Erauso *et al.*, 1993) und *P.horikoshii* (Gonzalez *et al.*, 1998) aus der Tiefsee des pazifischen Ozeans isoliert wurden, stammt *P.furiosus* aus einem marinen Solfataren-Feld, d.h. durch vulkanische Aktivität aufgeheizten Meeresgrund vor der Insel Vulcano im Mittelmeer vor Süd-Italien (Fiala & Stetter, 1986). Der theoretischen Stammbaum dieser Dreiecks-Konstellation zeigt also die beiden einander ähnlichen Arten *P.abysyi* und *P.horikoshii*, deren gemeinsamer Vorläufer sich schon vor etwas längerer Zeit von der zu *P.furiosus* führenden Linie abgetrennt hat (weite Details:(Lecompte *et al.*, 2001)). In den meisten Fällen scheint zu gelten, dass die Gene aus *P.furiosus* stärker denen aus *P.abysyi* als denen aus *P.horikoshii* ähneln. Diese Interpretation wird ergänzt durch die Protein-Stammbäume der Untereinheiten A' und L (siehe [4.4.1]), wo die drei *Pyrococcus*-Arten einen entsprechenden Cluster bilden, wobei nach dem Grad der Konservierung des jeweiligen Proteins die Detail-Relationen etwas unterschiedlich ausfallen.

4.3.1 Eigenschaften und konservierte Regionen der RNAP-Gene in *Pyrococcus*

Ein Teilaspekt der in diesem Kapitel gezeigten Grafiken wurden auf der Basis umfangreicher Alignments der homologen RNAP-Gene innerhalb von *Archaea* und im Vergleich mit *Saccharomyces cerevisiae* als Vertreter der *Eukarya* gewonnen. In einzelnen Fällen wurden Gensequenzen auch von noch nicht komplett sequenzierten *Archaea* herangezogen (z.B. *Sulfolobus acidocaldarius*). Während sich die *S.cerevisiae*-Gene recht gut in die Multiplen Alignments fügen, wurde auf die Integration der stärker abweichenden bakteriellen Sequenzen in den Konsensus von vornherein verzichtet um die Darstellung nicht zu unübersichtlich werden zu lassen.

Ziel der weiter unten gezeigten Darstellungen zu jeder RNAP-Untereinheit bei *P.furiosus* ist es, durch die Lokalisierung und Eingrenzung der besonders konservierten Positionen einen Hinweis auf für die Enzymaktivität der RNA-Polymerase besonders wichtigen Domänen (und auch einzelne Aminosäuren-Positionen) zu erlangen. Denn der Mechanismus der Transkription und der Aufbau der RNA-Polymerase bei *Archaea* ähnelt zwar dem bei *Eukarya*, weist aber auch charakteristische Unterschiede hierzu auf, wie man aus der unterschiedlichen Größe und streckenweise stark abweichenden Sequenz und 3D-Struktur der beteiligten Untereinheiten und Faktoren ableiten kann.

Das Argument, dass bei *Eukarya* viele Prozesse komplizierter (reguliert) sind als bei *Archaea*, und dass sich das zumeist in der größeren Anzahl der beteiligten Komponenten äußert, gilt ausnahmsweise nicht für die Anzahl der Untereinheiten, aus denen sich die RNAP bei *Eukarya* und *Archaea* zusammensetzen: Mit Ausnahme von Rpb8 und Rpb9 sind zu allen Untereinheiten der eukaryotischen Pol II archaeelle Homologe gefunden wurden (siehe [Abb.2-02]). Da einige Untereinheiten bei *Archaea* in zwei Teilen vorliegen (A' + A'', E' + E'') kommt man bei den methanogenen *Archaea*, die als Besonderheit zusätzlich eine gespaltene B-Untereinheit aufweisen (B' + B'') rechnerisch sogar auf 13 Untereinheiten – also eine mehr, als bei *Eukarya* (Pol II: 12 Untereinheiten). Bezieht man aber die bemerkenswert große Zahl der (bisher bekannten) an der Transkription beteiligten eukaryotischen Faktoren und RNA-Pol-assoziierten Hilfsproteine hinzu, so ergibt sich ein anderes Bild, so dass grundsätzlich weiter die Arbeitshypothese aufrechterhalten werden kann, dass die archaeelle RNA-Polymerase als Modellsystem für die komplexere eukaryotische RNA-Polymerase dienen kann, weil sie zwar in vieler Hinsicht sehr ähnlich, andererseits aber auch deutlich weniger komplex (und damit experimentell leichter zugänglich) ist. Allerdings sind auch für *Archaea* typische (und einzigartige) Struktur- und Funktionsmerkmale zu erwarten – z.B. als Anpassung an die oftmals extremen Standorte / Lebensräume.

Da in diesem Kapitel Multiple Alignments und die Diskussion eine Rolle spielt, wo eine Proteinsequenz "ähnlich" und konserviert, und wo sie "stark abweichend" ist, wird kurz und sehr allgemein darauf eingegangen, wie diese Parameter in der Bioinformatik behandelt werden, und was für Möglichkeiten sich für den Anwender ergeben.

Folgende Grafik verdeutlicht, welche Aspekte die Algorithmen zur Generierung von Alignments und zur Aufdeckung von Ähnlichkeit und Sequenzhomologie beachten müssen:

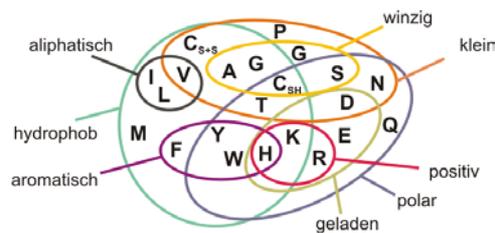


Abb.4-09: "Protein-Scoring-Systeme": Gruppierung ähnliche und unähnlicher Aminosäuren

Nach einer Folie, die von Rainer König vom DKFZ Heidelberg (HUSAR Biocomputing Service Group) im Rahmen eines Bioinformatik-Workshops in Kiel zur Veranschaulichung von „Protein-Scoring-Systemen“ gezeigt wurde.

Wie man sieht, wurden die 20 proteinogenen Aminosäuren in derartiger Weise angeordnet, dass ähnlich wie in der Mengenlehre Kreise und Ellipsen sich vielfach überschneidende Teilmengen bilden (das gezeigte Schema erhebt keinen Anspruch auf Vollständigkeit). Offensichtlich wird die Situation dadurch kompliziert, dass einige Aminosäuren an bis zu 5 Eigenschaften-Mengen teilhaben, wobei dann zwangsläufig je nach aktueller Fragestellung eine unterschiedliche Gewichtung derselben notwendig ist (beispielsweise darf Histidin gleichzeitig als „hydrophob“, „aromatisch“, „polar“, „geladen“, „positiv geladen“ betrachtet werden).

Nach dem „Dogma der Biochemie“ ist die Primärstruktur der Aminosäuresequenz allein zur Ausbildung der finalen 3D-Struktur des Proteins und somit seiner spezifischen Eigenschaften ausreichend (allerdings handelt es sich oftmals um ein durch Chaperone unterstütztes „assisted assembly“). Wenn man davon ausgeht, dass auf den Eigenschaften eines Organismus (und im Detail auf den Eigenschaften seiner Proteine und sogar auf Eigenschaften einzelner Aminosäuren an ganz spezieller Position im Protein) ein Selektionsdruck lastet, so wird unmittelbar klar, dass ein Aminosäure-Austausch mit einer sehr ähnlichen Aminosäure von der Evolution weniger hart „bestraft“ wird, als ein Austausch, der die Form des Proteins grundlegend verändert und z.B. die Inaktivierung eines Enzyms zur Folge hat, weil sein Substrat nicht mehr in das katalytische Zentrum paßt (natürlich kann auch mal ein "Glückstreffer" dabei sein, der zu einer schlagartig verbesserten biologischen "Fitness" des Organismus führt – das bewährteste Prinzip der Evolution aber scheint noch immer die Konservierung oder höchstens vorsichtige Veränderung zu sein – schließlich hat sich die jeweilige Struktur meist schon über unzählige Generationen hinweg bewährt).

In gewisser Weise ist also der Aminosäure-Code ähnlich „degeneriert“ wie der „genetische Code“, das heißt, ein Austausch des aliphatischen Leucins mit Isoleucin oder Valin hat (fast?) keinen Einfluß auf die resultierende 3D-Form bzw. spezifische Eigenschaften. Typisch ist auch der Austausch ähnlich geladener Aminosäuren wie Aspartat mit Glutamat oder Arginin mit Lysin.

Mag sich dies einerseits erschwerend bei der Identifizierung eines Proteins auswirken, so kann dieses Phänomen andererseits auch nützlich bei dem Vergleich von Verteilungsmustern zur Aufdeckung phylogenetischer Zusammenhänge sein.

Der Austausch von Aminosäuren kann im Laufe der Evolution so weit fortschreiten, dass das Auffinden eines homologen Gegenstücks in einem anderen Organismus allein aufgrund der Aminosäure-Sequenz nicht mehr greift, obwohl die für das Protein charakteristische allgemeine Gestalt und Funktion noch erhalten ist. Hier kommen dann häufig fortgeschrittene Techniken der Bioinformatik zum Einsatz, die sich bei der Suche nach Homologen eher auf die Analyse von 2D- und 3D-Strukturen stützen. Die rasante Entwicklung der Rechnerkapazitäten und die Fortschritte beim Verständnis der Faktoren und Kräfte, die die Form eines Proteins bestimmen, liefert gegenwärtig immer leistungsfähigere Werkzeuge (Computerprogramme), die teilweise online zugänglich sind.

Im folgenden wird grundsätzlich vorausgesetzt, dass auf den für die Funktion essentiellen „Merkmale“ eines Proteins ein positiver Selektionsdruck lastet, was ihre „Konservierung“ im Lauf der Evolution erzwingt. Die Faustformel „Konservierte Region = Hinweis auf Funktion“ kann als grober Anhaltspunkt für das Auffinden besonders wichtiger und interessanter Regionen im Protein

dienen. Im Gegenzug können auffällige Abweichungen vom Konsensus bei bestimmten Arten unter Umständen als für den speziellen Organismus (und für sein Überleben in einem bei *Archaea* oft extremen Habitat) förderliche Modifikationen diskutiert werden (oder vielleicht auch als Folge von Rekombinationsereignissen).

Ein Problem ist oftmals, dass sich umfangreichen Alignments seitenlang hinziehen und keinen intuitiv zu erfassenden Überblick, sondern eher Detailinformationen liefern. Dies gilt auch für die Proteinsequenz der großen Untereinheiten der RNAP aus den 17 komplett sequenzierten *Archaea*. Aus diesem Grund habe ich die unten gezeigten Grafiken selbst programmiert (mit Borland Delphi 5, siehe [8.4]), um jeweils bestimmte Eigenschaften der untersuchten Proteine komprimiert in einer Zeile darstellen zu können.

Die erste Zeile zeigt jeweils die aus den diskutierten Alignments ([8.1.3]) abgeleiteten konservierten Regionen in den RNA-Pol-Genen aus *Pyrococcus furiosus*. Die Zusatzgrafiken mit speziellen Aminosäuren (z.B. „geladene Aminosäuren“) oder mit Sekundärstrukturvorhersagen stellen eine Ergänzung dar, und können schon an dieser Stelle einige Indizien für strukturelle Besonderheiten innerhalb der konservierten Regionen liefern, bevor dann mit weiteren bioinformatischen Tools (bzw. Labor-Experimenten) Voraussagen bezüglich 2D- und 3D-Struktur (oder über die Wahrscheinlichkeit von Protein-Protein-Interaktionen mit einer anderen RNAP-Untereinheit oder einem Transkriptionsfaktor) konkretisiert werden.

Ein Sonderfall stellt die Untereinheit B (rpoB) dar: Bei ca. der Hälfte der untersuchten *Archaea* ist sie in zwei Teile gespalten (N-terminal B'' (rpoB2) und C-terminal B' (rpoB1); [4.2.1]). Weil hier (ähnlich wie beim Hinzufügen von der Sequenz von Rpb2 aus *S.cerevisiae*) der Konsensus erheblich reduziert wird, wurden zwei separate Alignments durchgeführt (einerseits B + B' und andererseits B + B'') und in einem einzigen Konsensus bezogen auf *P.furiosus* kombiniert (siehe Anhang [8.1.3.1]).

4.3.1.1 B (rpoB) aus *P.furiosus*: Konservierte Regionen und Eigenschaften

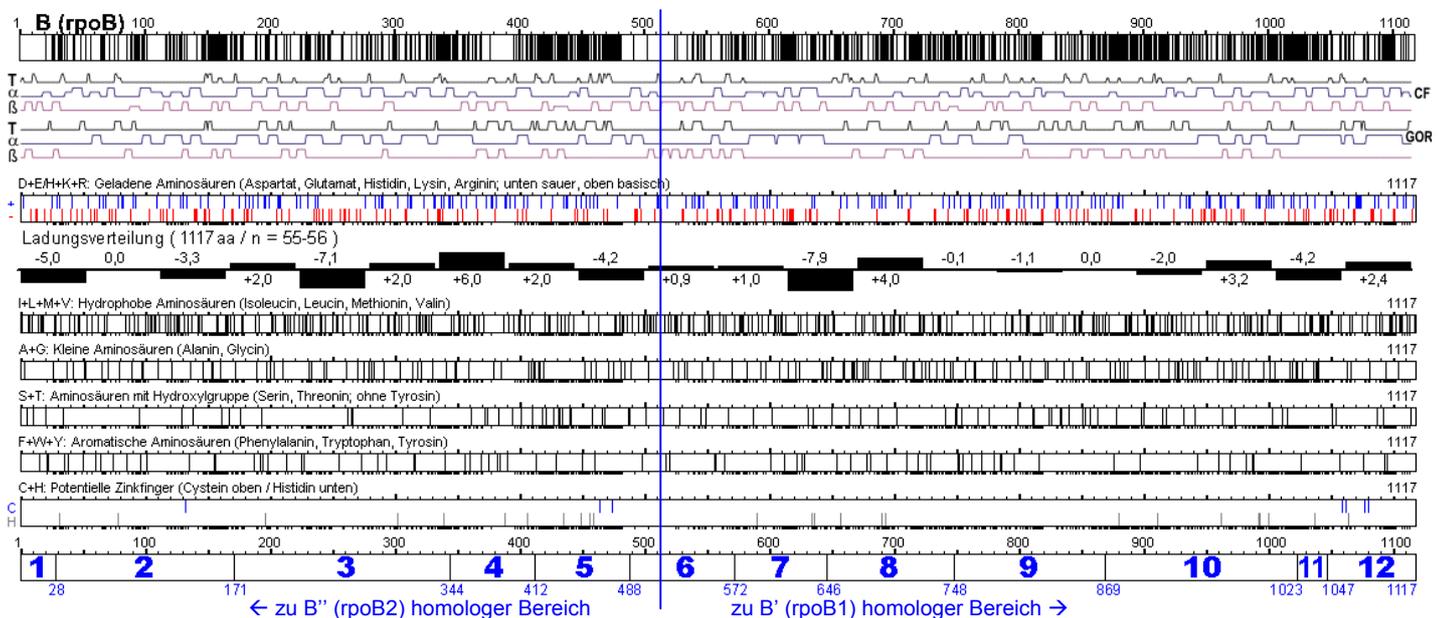


Abb.4-10: Eigenschaften von rpoB aus *P.furiosus*

Erste Zeile mit in *P.furiosus* konservierten Aminosäure-Positionen der Untereinheit B entsprechend Konsensus (schwarz = konserviert; siehe [8.1.3.1]). T = „turn“, α = α -Helix, β = β -Faltblatt, 1. Sekundärstrukturvorhersage nach Chou-Fasman (CF) (Chou & Fasman, 1978), 2. Vorhersage nach Garnier, Osguthorpe, Robson (GOR) (Garnier *et al.*, 1978). Die Kombination dieser beiden Methoden kann wichtige Hinweise liefern, bleibt aber hypothetisch (Kyngas & Valjakka, 1998). Saure Aminosäuren/negativ geladen: **rot**. Basische Aminosäuren/positiv geladen: **blau**. In der unteren Zeile Cystein- (oben) und Histidin-Positionen (z.B. für potentielle Zinkfinger-Motive). Die Zeile mit den blauen Zahlen stellt die Übertragung der für die eukaryotische RNAP aus *S.cerevisiae* vorgenommene Einteilung in Struktur- und Funktionsabschnitte (Cramer *et al.*, 2001) auf die Verhältnisse bei *Pyrococcus furiosus* dar ([Abb.4-21, Rpb2]).

Die angegebene Trennlinie für die entweder zu rpoB2 oder rpoB1 homologen Abschnitte bei Position 515 entspricht dem relativ einheitlichen Startbereich von rpoB1 bei den meisten *Archaea*

mit geteilter B-Untereinheit. *Sulfolobus solfataricus* fällt hier durch seinen zum N-Terminus hin verschobenen Start deutlich aus dem Rahmen, was an anderer Stelle ausführlicher diskutiert wird. Die Länge des C-Terminus von rpoB2 ist bei den o.g. *Archaea* dagegen ziemlich variabel.

Wie schon für die RNAP aus *E.coli* gezeigt wurde (Severinov *et al.*, 1996), kann bei Rekonstitutionsexperimenten mit künstlich in mehrere Einzelproteine aufgeteilten Untereinheiten durchaus mit korrektem „Assembly“ und teilweise wiederhergestellter katalytischer Aktivität des RNAP-Gesamtkomplex gerechnet werden. Gegenwärtig wird die größte Untereinheit B im Rahmen der Fortführung des *P.furiosus*-RNA-Polymerase-Projektes nicht nur als ganzes, sondern auch entsprechend der oben angegebenen Trennlinie in zwei Teilproteinen untersucht (persönliche Mitteilung, Naji,S.).

4.3.1.2 A' (rpoA1) aus *P.furiosus*: Konservierte Regionen und Eigenschaften

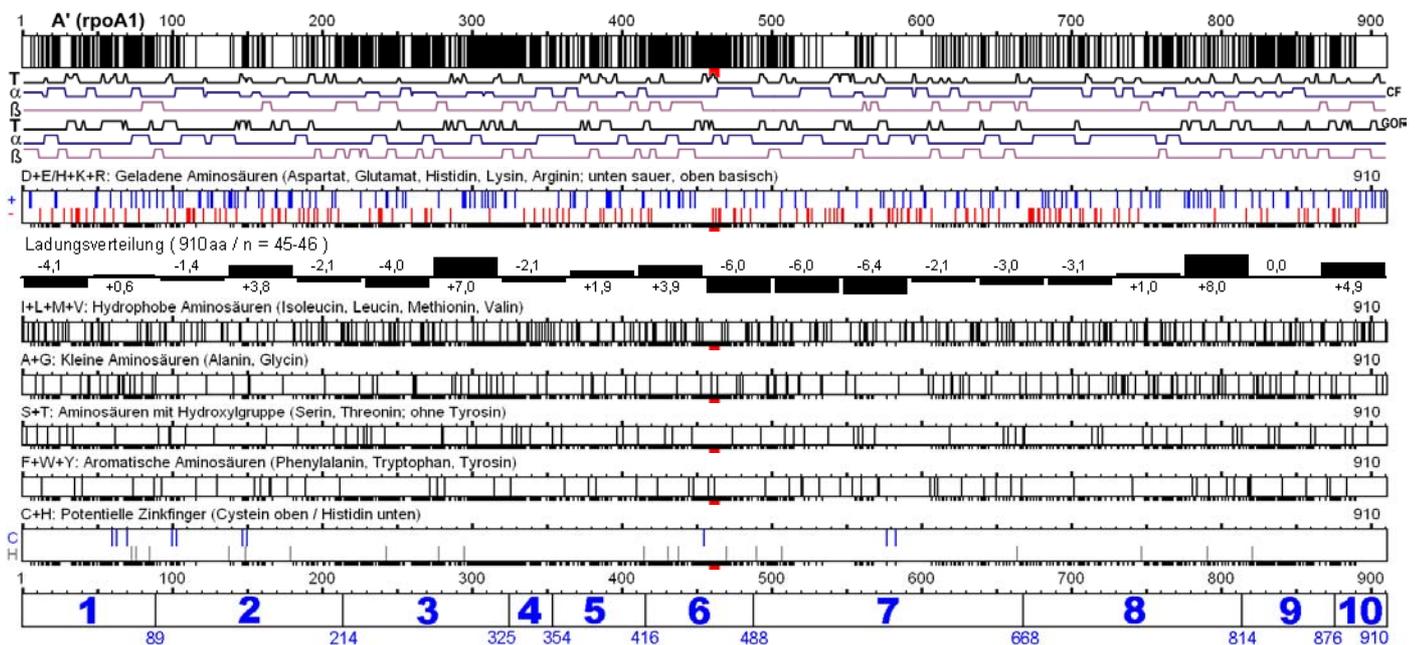


Abb.4-11: Eigenschaften von rpoA1 aus *P.furiosus*

(Das hochkonserviertes „NADFDGD“-Motiv (Dieci *et al.*, 1995; Sonntag & Darai, 1995; Zaychikov *et al.*, 1996) des katalytischen Zentrums im Bereich von Pos.460 wurde rot unterstrichen).

Erste Zeile mit in *P.furiosus* konservierten Aminosäure-Positionen der Untereinheit A' entsprechend Konsensus (schwarz = konserviert; siehe [8.1.3.2]). T = „turn“, α = α -Helix, β -Faltblatt, CF und GOR: siehe Legende von [Abb.4-10]. Saure Aminosäuren/negativ geladen: rot. Basische Aminosäuren/positiv geladen: blau. In der unteren Zeile Cystein- (oben) und Histidin-Positionen. Zeile mit den blauen Zahlen: Siehe Legende zu [Abb.4-10] und Angaben unter [Abb.4-20, Rpb1].

Diese (bei *Pyrococcus* zweitgrößte) Untereinheit ist die am stärksten konservierte RNA-Pol-Untereinheit, und enthält das katalytische Zentrum der RNA-Polymerase. Die universale Verbreitung des NADFDGD-Motivs ist nicht ganz zufällig – schließlich ist dieser Bereich an der Komplexierung des zentralen Mg^{2+} -Ions beteiligt (Zaychikov *et al.*, 1996). Das o.g. Motiv scheint so essentiell zu sein, dass es sich nicht nur bei allen untersuchten *Archaea* und *S.cerevisiae* findet, sondern auch bei *E.coli* mit seiner, auf den ersten Blick (scheinbar) ziemlich stark abweichenden RNA-Pol-Struktur. Sogar viele cytoplasmatische DNA-Viren enthalten dieses Motiv (Sonntag & Darai, 1995). Zusätzlich scheint für die korrekte Funktion dieser ohnehin sehr stark konservierten Untereinheit die Sequenz RQP(TS)LH von entscheidender Bedeutung zu sein, die bei *P.furiosus* 27 Aminosäure-Positionen stromaufwärts vom genannten Hauptmotiv liegt.

4.3.1.3 A'' (rpoA2) und H (rpoH): Konservierte Regionen und Eigenschaften

Die Untereinheit H bindet im Far Western Blot-Experiment sehr spezifisch an A'', worauf auch die verfügbaren 3D-Daten hinweisen. Deshalb werden die beiden RNAP-Untereinheiten hier gemeinsam betrachtet.

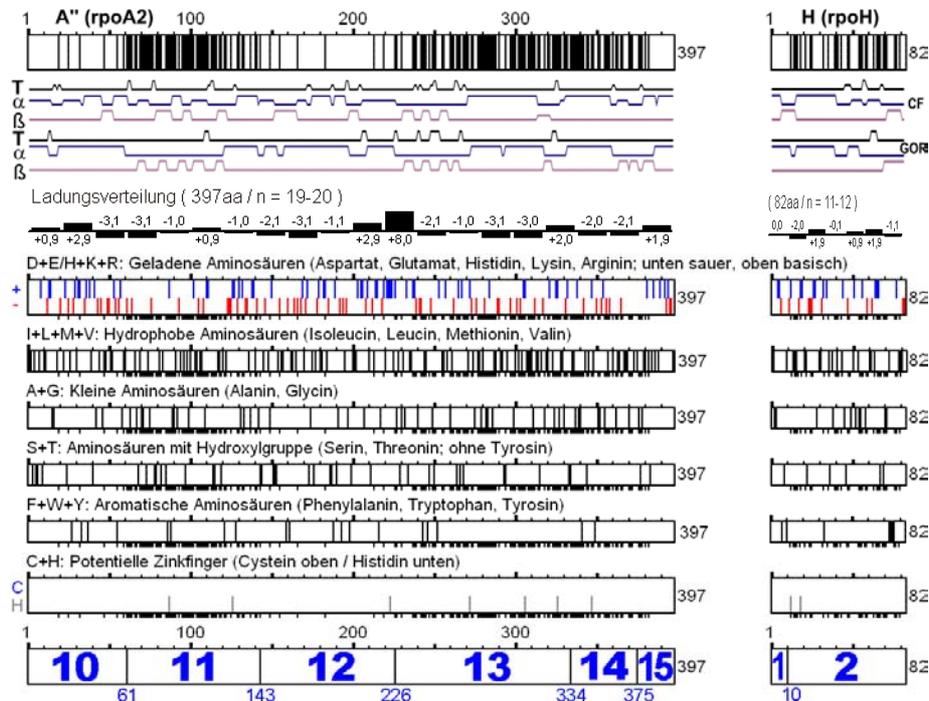


Abb.4-12: Eigenschaften von rpoA2 und rpoH aus *P.furiosus*

Links die 397 Aminosäuren lange, drittgrößte Untereinheit A''; rechts die 82 Aminosäuren lange Untereinheit H. Erste Zeile mit in *P.furiosus* konservierten Aminosäure-Positionen entsprechend Konsensus (schwarz = konserviert; siehe [8.1.3.3]). T = „turn“, α = α -Helix, β -Faltblatt, CF und GOR: siehe Legende von [Abb.4-10]. Saure Aminosäuren/negativ geladen: rot. Basische Aminosäuren/positiv geladen: blau. In der unteren Zeile Cystein- (oben) und Histidin-Positionen. Zeile mit den blauen Zahlen: Siehe Legende zu [Abb.4-10] und die Angaben unter [Abb.4-20, Rpb1] und [Abb.4-24, Rpb5].

Bei A'' (rpoA2) lassen sich grob zwei voneinander getrennte konservierte Bereiche erkennen: 1. ca. Pos. 60 – 140; 2. ca. Pos. 235 – 380. Die drei *Pyrococcus*-Arten verfügen über eine Gruppe von 4 aufeinander folgende Tyrosin-Reste am C-Terminus von rpoH. Die hierzu homologe Untereinheit Rpb5 aus *S.cerevisiae* zeigt im C-terminalen Drittel recht hohe Sequenzähnlichkeit zur H-Untereinheit der *Archaea*, ist aber insgesamt deutlich größer.

Untereinheit B, A', A'' (und H) bilden den "katalytischen Kern" der RNA-Polymerase. Aktuelle Untersuchungen mit der *P.furiosus*-RNAP haben gezeigt, dass es diese vier Untereinheiten sind, die den direkten Kontakt zu der DNA herstellen, die in der großen Spalte am katalytischen Zentrum vorbei längs durch den RNAP-Komplex hindurch verläuft (Bartlett *et al.*, 2003).

4.3.1.4 D (rpoD) und L (rpoL): Konservierte Regionen bei *P.furiosus* :

Wie aus der Literatur für *M.jannaschii* bekannt ist (Eloranta *et al.*, 1998; Werner & Weinzierl, 2002), bilden die Untereinheiten D und L einen Komplex, der sehr wahrscheinlich (gemeinsam mit N (rpoN) und evtl. P (rpoP) die gemeinsame „Assembly Plattform“ für die übrigen Komponenten des RNAP-Komplexes bildet. Auch bei Far Western Blot- und Native-PAGE- und Molekularsieb-Chromatographie-Experimenten im Rahmen dieser Arbeit zeigte sich die starke Affinität der beiden Proteine zueinander, weshalb sie hier gemeinsam betrachtet werden.

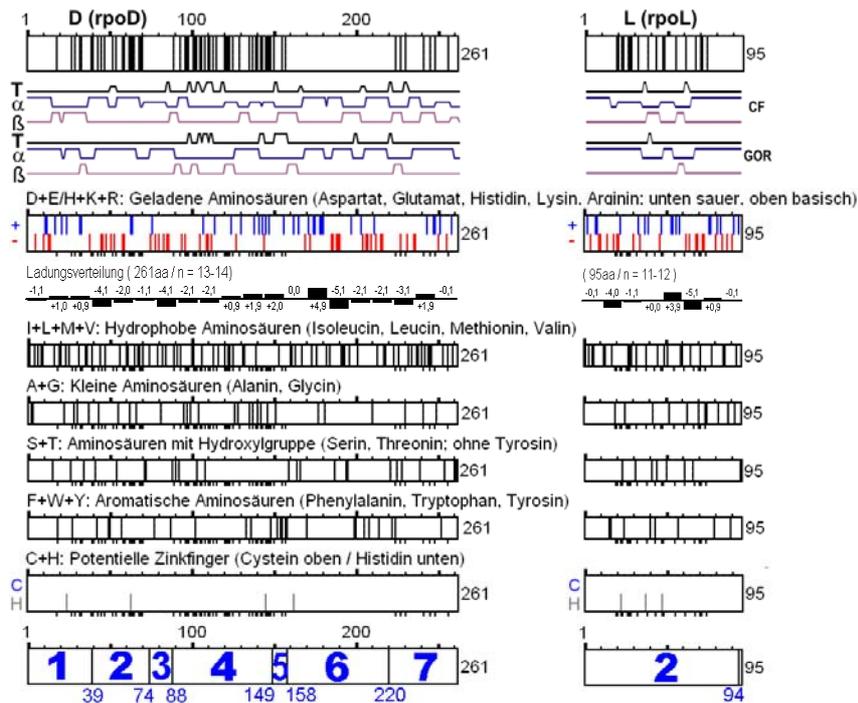


Abb.4-13: Eigenschaften von rpoD und rpoL aus *P.furiosus*

Links die 261 Aminosäuren lange Untereinheit D; rechts die 95 Aminosäuren lange Untereinheit L. Erste Zeile mit in *P.furiosus* konservierten Aminosäure-Positionen entsprechend Konsensus (schwarz = konserviert; siehe [8.1.3.4] und [8.1.3.10]). T = „turn“, α = α -Helix, β -Faltblatt, CF und GOR: siehe Legende von [Abb.4-10]. Saure Aminosäuren/negativ geladen: rot. Basische Aminosäuren/positiv geladen: blau. In der unteren Zeile Cystein- (oben) und Histidin-Positionen. Zeile mit den blauen Zahlen: Siehe Legende zu [Abb.4-10] und Angaben unter [Abb.4-22, Rpb3 und Rpb11]. Insgesamt ist ein geringer Grad an Konserviertheit feststellbar. Das in rpoD/Rpb3 bei *Saccharomyces* und einigen *Archaea* vorhandene „Zink loop“ (3) fehlt hier komplett und die „Loop-Region“ (6) ist auffällig variabel.

4.3.1.5 E' (rpoE1), E'' (rpoE2) und F (rpoF): Konservierte Regionen bei *P.furiosus*

Far Western Blots in dieser Arbeit haben gezeigt, dass die RNAP-Untereinheit F sehr spezifisch und ausschließlich an E' bindet, wobei die „Kreuzreaktionen“ mit anderen RNA-Pol-Untereinheiten minimal sind (siehe [Abb.4-47]). In der Literatur deutet sich an, dass diese beiden Proteine zur Bildung eines Heterodimers neigen, das nicht unbedingt zu jedem Zeitpunkt in stöchiometrischem Verhältnis im RNA-Pol-Komplex gefunden werden muß, sondern möglicherweise eher locker an die übrigen Untereinheiten gebunden sind (Todone *et al.*, 2001).

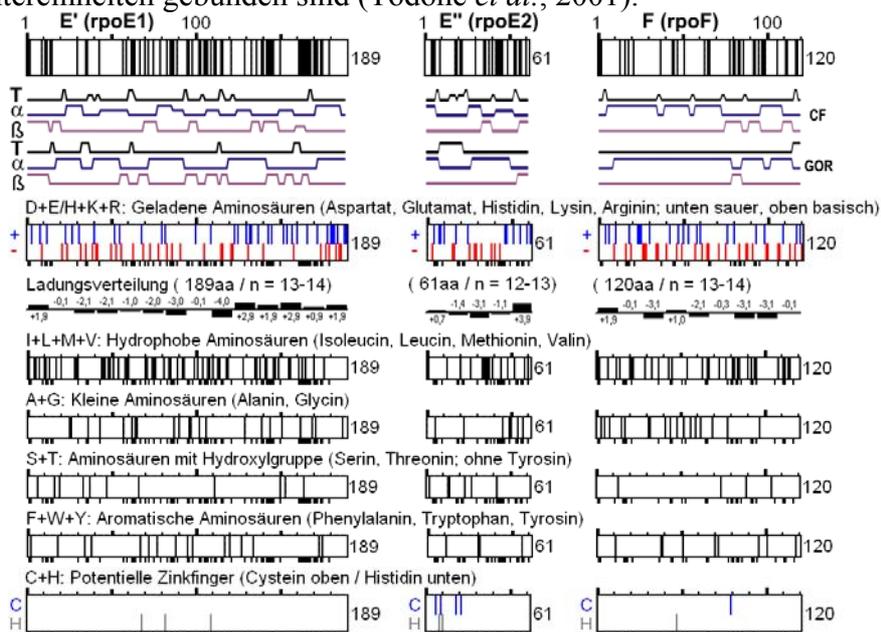


Abb.4-14: Eigenschaften von rpoE1, rpoE2 und rpoF aus *P.furiosus*

Links die 189 Aminosäuren lange E'-Untereinheit; in der Mitte die 61 Aminosäuren lange E''-Untereinheit; rechts die 120 Aminosäuren lange Untereinheit F. Erste Zeile mit in *P.furiosus* konservierten Aminosäure-Positionen entsprechend

Konsensus (schwarz = konserviert; siehe [8.1.3.5, [8.1.3.6]) und [8.1.3.7]. T = „turn“, α = α -Helix, β -Faltblatt, CF und GOR: siehe Legende von [Abb.4-10]. Saure Aminosäuren/negativ geladen: rot. Basische Aminosäuren/positiv geladen: blau. In der unteren Zeile Cystein- (oben) und Histidin-Positionen. Zeile mit den blauen Zahlen: Siehe Legende zu [Abb.4-10] und Angaben unter [Abb.4-23, Rpb7 und Rpb4].

Abgesehen von *S.acidocaldarius* liegt die Untereinheit E bei allen hier untersuchten *Archaea* in zwei Teilen (E' und E'') vor.

Bei E' (rpoE1) fällt bei einigen *Archaea* ein stark geladener C-Terminus auf, der fast ausschließlich aus sauren und basischen Aminosäuren besteht (siehe [8.1.3.5]). Dieser Bereich fällt allerdings nicht in die direkt davor liegende, etwa 16 Aminosäuren lange hochkonservierte Region.

Die nur 61 Aminosäuren lange E''-Untereinheit (rpoE2) zeigt insgesamt keine auffällig verringerte Ähnlichkeit zum Konsensus, wie es bei einem (durch eine einzelne Deletion und damit verbundenen Rasterverschiebung) inaktivierten Gen der Fall sein würde. Interessant ist jedoch, dass gerade im N-terminalen Bereich in der Umgebung des hochkonservierten Zinkfinger-Motivs die Aminosäure-Sequenz auffällig schlecht konserviert ist (die vier Cystein-Reste werden mit blauen Strichen angedeutet, bzw. im Alignment rot markiert [8.1.3.6]). Ein weiteres Indiz für besondere Eigenschaften dieses Proteins deutet sich in komplett entgegengesetzt geladenen Regionen an: Dem Zinkfinger-Motiv, das grob das N-terminale Drittel der Gesamtsequenz beansprucht, folgt ein negativ geladener Mittelteil mit sauren Aminosäuren, der schließlich C-terminal in einen Bereich mit positiv geladenen, basischen Aminosäuren in Verbindung mit einer Anhäufung kleiner Aminosäuren mündet (beide Abschnitte knapp ein Drittel der Gesamtsequenz).

Die Aminosäuresequenz von F (rpoF) zeigt auf den Blick keine besonderen Auffälligkeiten, abgesehen vielleicht von einer insgesamt ziemlich geringen Übereinstimmung mit dem Konsensus aus dem Alignment. Die BLAST-Suche hat gleich bei mehreren *Archaea* keinen sicheren Hinweis auf ein entsprechendes Protein geliefert (bzw. falsche Treffer), so dass hier nur 14 statt 17 bzw. 18 *Archaea* in das Alignment integriert werden konnten (siehe [8.1.3.7]). Je weniger Sequenzen von weiter entfernten Arten verwendet werden, desto mehr gewinnen die drei einander sehr ähnlichen *Pyrococcus*-Arten an Gewicht, was im Alignment eine gewisse Schiefelage verursacht (allerdings hat jede Auswahl Vor- und Nachteile – entscheidend ist die gegenwärtige Verfügbarkeit archaeller Genome; auch *Sufolobus* ist ja überdurchschnittlich in den Alignments enthalten – mit *S.solfataricus*, *S.tokodaii* und in einigen Fällen zusätzlich *S.acidocaldarius*).

4.3.1.6 K (rpoK), N (rpoN) und P (rpoP): Konservierte Regionen bei *P.furiosus* :

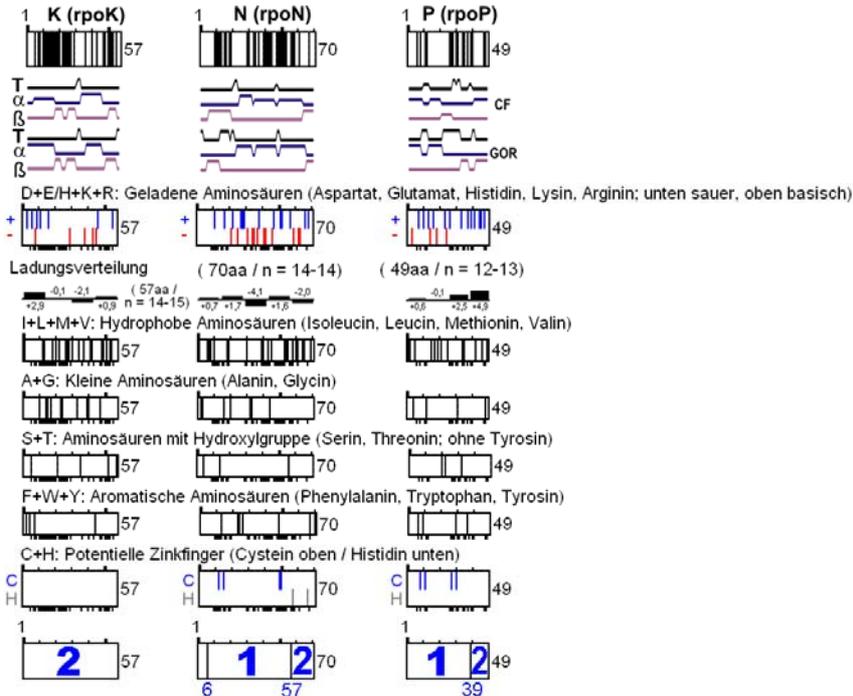


Abb.4-15: Eigenschaften von rpoK, rpoN und rpoP aus *P.furiosus*
Links die 57 Aminosäuren lange Untereinheit K; in der Mitte die 70 Aminosäuren lange Untereinheit N; rechts die nur 49 Aminosäuren lange Untereinheit P. Erste Zeile mit in *P.furiosus* konservierten Aminosäure-Positionen entsprechend Konsensus (schwarz = konserviert; siehe [8.1.3.9], [8.1.3.11] und [8.1.3.12]). T = „turn“, α = α -Helix, β -Faltblatt, CF und GOR: siehe Legende von [Abb.4-10]. Saure Aminosäuren/negativ geladen: rot. Basische Aminosäuren/positiv geladen: blau. In der unteren Zeile Cystein- (oben) und Histidin-Positionen. Zeile mit den blauen Zahlen: Siehe Legende zu [Abb.4-10] und Angaben unter [Abb.4-24, Rpb6] und [Abb.4-25, Rpb10 und Rpb12].

Neben rpoB, rpoA1, rpoN und rpoE2 besitzt auch rpoP ein Zinkfinger-Motiv. Ähnlich wie bei rpoE2 ist auch hier der Grad der Konserviertheit direkt in der Umgebung des C4-Typ-Zinkfinger-Motivs ziemlich gering. Die zweite Hälfte des nur 49 Aminosäuren langen Proteins ist auffällig positiv geladen (+8 gegenüber +1 am N-Terminus), was sich auch in der extremen Gesamtladung (+9) bzw. einem entsprechend hohen Isoelektrischen Punkt (pI 10,7) äußert (siehe auch [Abb.4-03]).

4.3.1.7 Zusammenfassung: Konservierte Regionen

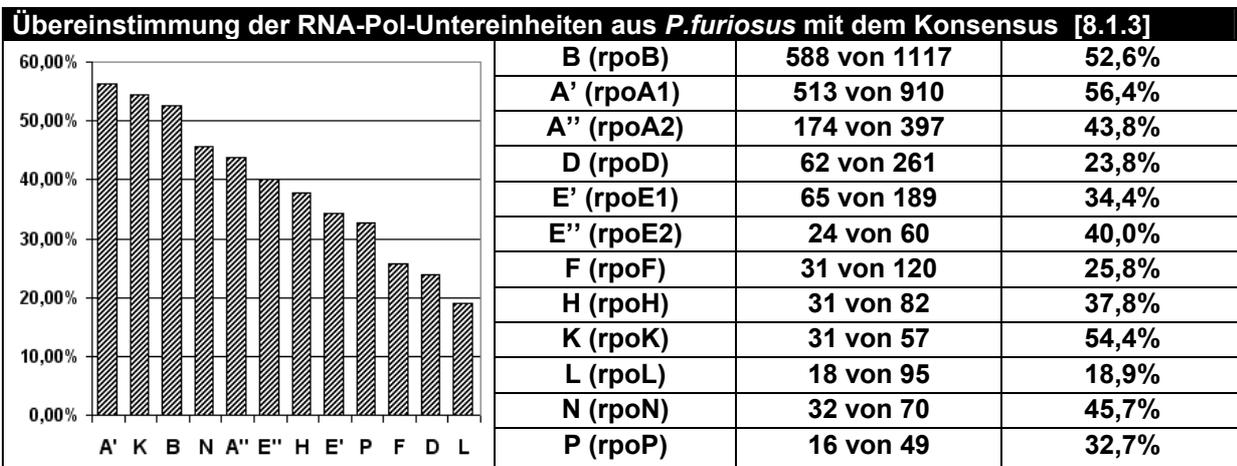


Abb.4-16: Übereinstimmung der RNA-Pol-Untereinheiten aus *P.furiosus* mit dem Konsensus
Die „großen“ archaeellen Alignments bestätigen die Ergebnisse der *Pyrococcus*-internen Alignments: A' mit dem katalytischen Zentrum ist am stärksten konserviert, während sich D, F und L am weitesten vom archaeellen Konsensus entfernt haben (die Differenz zwischen A' und L beträgt immerhin 37,5%).

4.4 Zinkfinger-Motiv und katalytisches Zentrum in der A'-Untereinheit

Das Motiv „NADFDGD“ des katalytischen Zentrums der A'-Untereinheit der RNA-Polymerase ist ohne Ausnahme komplett in allen untersuchten Protein-Genen enthalten – selbst bei *S.cerevisiae* und *E.coli* (Dieci *et al.*, 1995; Sonntag & Darai, 1995; Zaychikov *et al.*, 1996). Aber auch die direkt flankierenden Aminosäure-Positionen sind hoch konserviert: Mit Ausnahme von *E.coli* und *M.kandleri* findet sich das erweiterte Motiv „PYNADFDGDENM“ bei allen anderen hier untersuchten Organismen. Zusammen mit der Position der Zinkfinger-Motive (bzw. der Positionen von Cystein- bzw. Histidin-Resten) zeigt das folgende „grafische Alignment“ auch in der globalen Proteinstruktur konservierte Muster („NADFDGD-Motiv“ rot markiert):

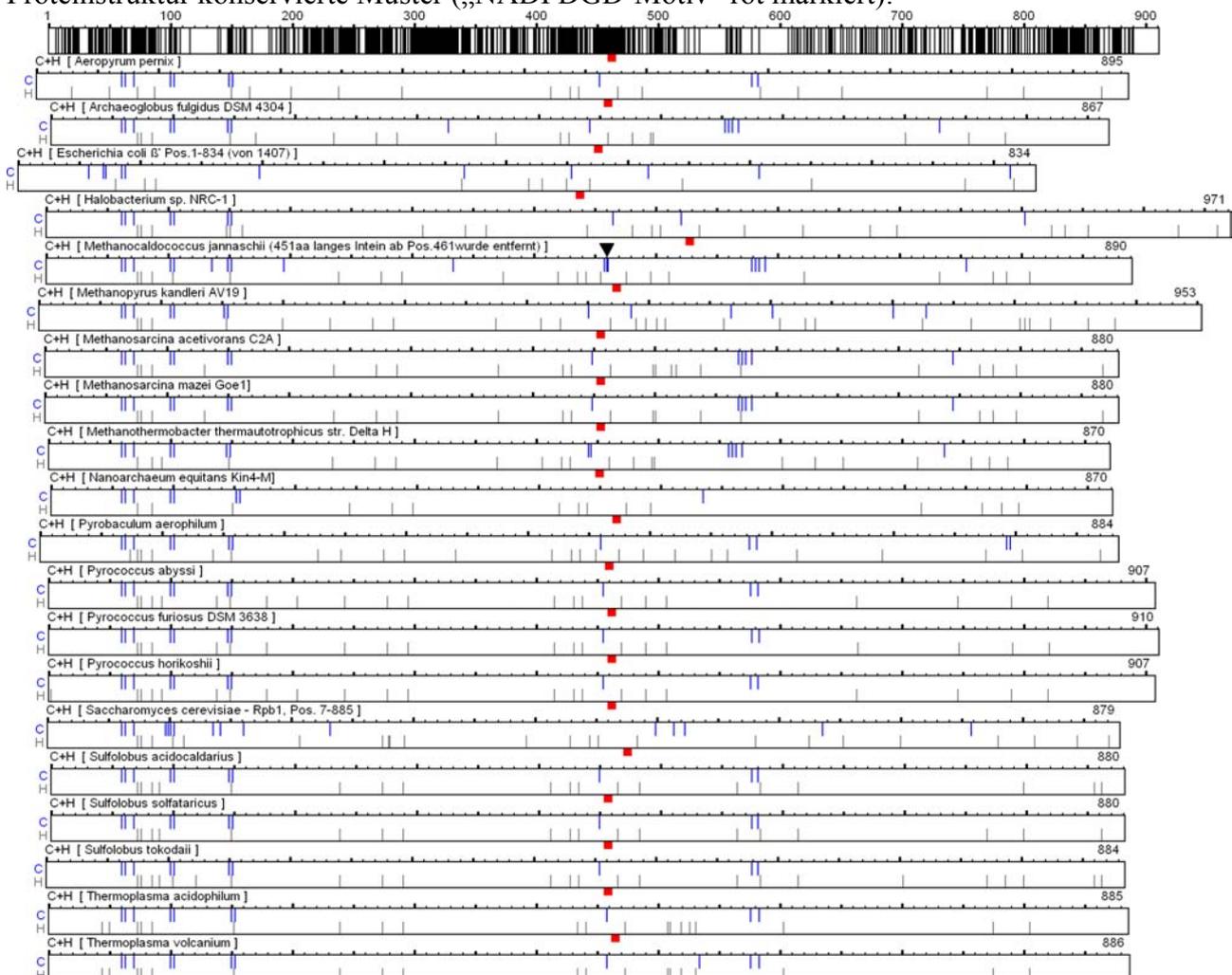


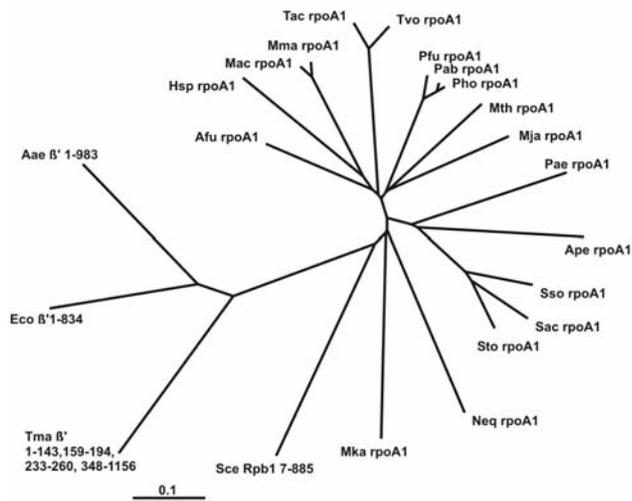
Abb.4-17: Vergleichende grafische Darstellung von rpoA1 (Rpb1, β')

"Grafisches Alignment" der am stärksten konservierten RNAP-Untereinheit A'. Rote Markierung: Katalytisches Zentrum. Blaue Striche: Cystein-Reste, graue Striche: Histidin-Reste. Da sowohl Cystein als auch Histidin selten vorkommende Aminosäuren sind, deren Position aber meist besonders konserviert ist, ergibt sich ein selbst über die Domängengrenzen hinweg erstaunlich regelmäßiges Muster an Positionen und Abständen, was ein starkes Indiz für die besondere funktionale Bedeutung dieser Positionen und der Untereinheit A' an sich ist.

An den Anfang wurde der auf *P.furiosus* bezogene Konsensus aus dem Alignment [8.1.3.2] gestellt, um die Verteilung besonders konservierten Regionen mit einzubeziehen. Das Intein an hochsensibler Position direkt vor dem "katalytischen Zentrum" bei *M.jannaschii* wurde mit einem schwarzen Dreieck markiert. Die Grafiken der einzelnen Sequenzen wurden so justiert, dass die beiden Cystein-Reste des ersten Zinkfinger-Motives untereinander stehen. Wie man erkennen kann, liegt das hypothetische Intein bei *Halobacterium* zwischen diesem Motiv und dem katalytischen Zentrum. Auch wenn es nicht den Anschein hat: Die *E.coli*-Sequenz ist korrekt unter den homologen Cystein-Positionen angeordnet – nicht völlig unerwartet zeigen sich bei diesem Vertreter der *Bacteria* auch sonst die größten Abweichungen vom (archaeell dominierten) Konsensus.

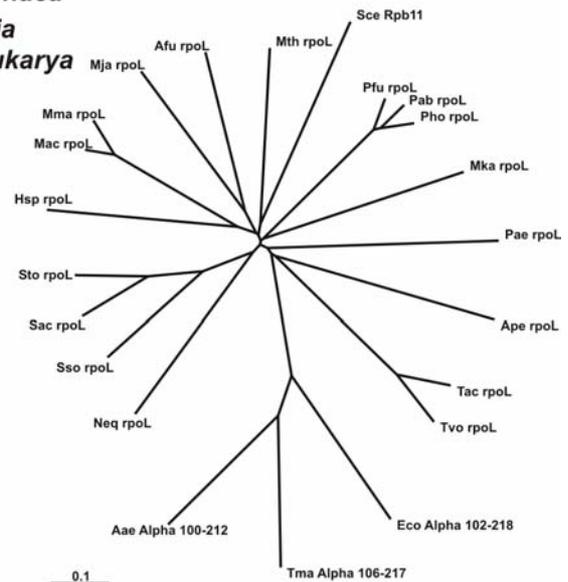
4.4.1 Ist die hochkonservierte A'-Untereinheit geeignet für Stammbaum-Erstellung?

RpoA' bei *Archaea*
 β' bei *Bacteria*
 Rpb1 bei *Eukarya*



stark konserviert

RpoL bei *Archaea*
 α bei *Bacteria*
 Rpb11 bei *Eukarya*



schwach konserviert

Abb.4-18: Proteinstammbäume für rpoA1 (Rpb1 / β') und rpoL (Rpb11 / α).

Ape = *Aeropyrum pernix*, **Tac** = *Thermoplasma acidophilum*, **Tvo** = *Thermoplasma volcanium*, **Mka** = *Methanopyrus kandleri* AV19, **Mja** = *Methanocaldococcus jannaschii*, **Mth** = *Methanothermobacter thermautotrophicus*, **Mac** = *Methanosarcina acetivorans* C2A, **Mma** = *Methanosarcina mazei* Goe1, **Pfu** = *Pyrococcus furiosus* DSM 3638, **Pab** = *Pyrococcus abyssi*, **Pho** = *Pyrococcus horikoshii*, **Afu** = *Archaeoglobus fulgidus*, **Hsp** = *Halobacterium* sp. NRC-1, **Sac** = *Sulfolobus acidocaldarius*, **Sso** = *Sulfolobus solfataricus*, **Sto** = *Sulfolobus tokodaii*, **Pae** = *Pyrobaculum aerophilum*, **Neq** = *Nanoarchaeum equitans* Kin4-M., **Eco** = *Escherichia coli*, Positionen 1-834 der β'-Untereinheit der RNA-Polymerase (von insgesamt 1407aa), Alpha 102-218 = Positionen der α-Untereinheit, **Sce** = *Saccharomyces cerevisiae*, Positionen 7-885 von Rpb1 der Pol II (von insgesamt 1733aa), **Aae** = *Aquifex aeolicus*, homologe Abschnitte der α- bzw. β'-Untereinheit, **Tma** = *Thermotoga maritima*, α- bzw. β'-Untereinheit, wobei Tma gegenüber Aae über 3 zusätzliche Inserts verfügt, die entfernt wurden. Die phylogenetische Distanz wird durch die Länge der Äste veranschaulicht, wobei die Angabe „0,1“ des Maßstabes für „0,1 Austausch pro Aminosäure-Position“ steht. Multiple Alignment: Clustal X 1.83; grafische Darstellung des Stammbaumes: TreeView 1.6.6 (siehe [8.3]).

Aquifex und *Thermotoga* wurden dieser Darstellung hinzugefügt, um neben *E.coli* zusätzlich die hyperthermophilen *Bacteria* mit ins Spiel zu bringen. Für die Erstellung von Protein-Stammbäumen wurde die am stärksten konservierte RNA-Pol-Untereinheit A' gewählt, der als Kontrast ein Stammbaum der geringst-konservierten Untereinheit L entgegengestellt wird.

Entgegen meinen Erwartungen ist aber in diesen beiden Stammbäumen keine auffällige Nähe zum potentiell „heißen“ Ursprung und bestimmten hyperthermophilen *Archaea* erkennbar. Wie aus der Literatur bekannt, eignen sich stark konservierte Gene wie rpoA1 besser für eine Auflösung phylogenetischer Relationen über große Distanzen hin, während geringer konservierte Gene eine bessere Differenzierung z.B. innerhalb einer Gattung ermöglichen.

Recht plausibel ist bei dem rpoA1-Stammbaum die Nähe von *N.equitans* (*Nanoarchaeota*) zum bekanntermaßen sehr tief im Stammbaum abzweigenden (und möglicherweise besonders „ursprünglichen“) *M.kandleri*. Passend ist auch, dass in diesem Umfeld die Linie entspringt, die zu *Eukarya* und *Bacteria* führt. Bei dem Vergleichsstammbaum mit Untereinheit L ist dieser Zusammenhang nicht mehr so gut erkennbar. Dafür scheinen die Abstände zwischen nah verwandten Arten besser aufgelöst. Beim Vergleich dieser beiden Stammbäume muß bedacht werden, dass rpoA1 viel größer als rpoL ist und aufgrund der Vielzahl an überprüften Aminosäurepositionen zwangsläufig ein statistisch ausgewogeneres und weniger von „Zufällen“ abhängiges Ergebnis liefert.

Interessant ist die Konstellation der drei *Pyrococcus*-Arten untereinander: Vor dem Hintergrund der Daten aus den Dreier-Alignments (siehe [8.1.2]) scheint die Abweichung von *P.abyssii* und *P.horikoshii* einerseits und *P.furiosus* andererseits bei Untereinheit A' etwas extrem zu sein, während sie bei Untereinheit L (etwas zu) gering erscheint. Wahrscheinlich würde ein Stammbaum

mit einem RNA-Pol-Gen, das in seiner Mutationsrate zwischen A' und L liegt den realen phylogenetischen Relationen am nächsten kommen.

Als Fazit bestätigt sich, dass Protein-Stammbäume neben den 16S-rRNA-Stammbäumen durchaus ähnliche (und ergänzende) Ergebnisse bezüglich mutmaßlicher Phylogenie liefern können. Ideal wäre es wahrscheinlich, von allen homologen Proteinen von Organismen Stammbäume zu erstellen, wodurch eine Schiefelage z.B. durch horizontalen Gentransfer weniger verhängnisvoll wäre. Bei den momentanen raschen Fortschritten auf dem Gebiet der Bioinformatik wäre dies durchaus denkbar.

4.5 E' / E'': Überlappende Gene im Rahmen translationaler Kopplung von Genen

Die Untereinheit E' (rpoE1) ist eines der mittelgroßen RNAP-Proteine, deren Vorkommen in der nativen RNAP schon lange bekannt ist. Von *Sulfolobus acidocaldarius* ist bekannt, dass die hier vorkommende E-Untereinheit größer als bei anderen *Archaea* ist (Lanzendörfer *et al.*, 1994) und über einen zusätzlichen C-terminalen Anteil mit einem charakteristischen Zinkfinger-Motiv verfügt. Zumindest in allen anderen momentan zugänglichen, komplett sequenzierten archaeeellen Genomen ist dieser Anteil in einem separaten Gen enthalten, das meist rpoE2 genannt wird. Im Gegensatz zur E'-Untereinheit ist über die Funktion von E'' nicht viel bekannt. Da das Gen in der Regel direkt hinter rpoE1 liegt und mit diesem (unter Verschiebung des Leserasters überlappt), wurde es lange Zeit nicht als eigenständiges Gen annotiert (In dieser Arbeit wurde zu einem frühen Zeitpunkt die gemeinsame Klonierung und Testexpression des sowohl rpoE1, als auch rpoE2 umfassenden DNA-Abschnittes vorgenommen um zu klären, ob es sich um einen Sequenzierfehler handelt. Das resultierende Protein zeigte jedoch dasselbe Molekulargewicht wie rpoE1 allein). Mittlerweile findet sich in fast allen hier ausgewerteten Genomdateien (im Genbank-Format) ein entsprechender Verweis auf rpoE2.

Das Alignment mit den übrigen *Archaea* und *S.cerevisiae* unter [8.1.3.6] zeigt den, zumindest für das mutmaßliche Zinkfingermotiv erstaunlich hohen Grad an Konserviertheit, was die Betrachtung von rpoE2 als einen durch Rastermutation abgetrennten, inaktivierten Teil von rpoE extrem unwahrscheinlich erscheinen lässt. Bleibt also die zum Beispiel auch bei *Bacteria* bekannte "gekoppelte Translation", die in vielen Fällen auf Genomebene schwierig zu erkennen ist, vor allem wenn die Gene überlappen und keine der üblichen, stromaufwärts vom Translationsstart gelegene Motive wie die Ribosomen-Bindestelle (rbs) vorhanden sind. Der Mechanismus ist grundsätzlich der, dass die Ribosomen nach dem Erreichen des Stop-Codons des vorangehenden Genes nicht sofort dissoziieren, sondern in der Lage sind, die folgende intercistronische Region (über deren maximale Länge die Angaben etwas auseinander gehen) zu überbrücken und beim Erreichen eines Start-Codons mit der Translation des nachfolgenden Genes zu beginnen, ohne dass eine zusätzliche rbs benötigt wird. Als Sonderfall gibt die "negative intercistronische Region", bei der die aufeinander folgenden Gene partiell überlappen. Auch hier erfolgt der Translationsstart vom nächsten verfügbaren, stromaufwärts gelegenen Start-Codon.

Da ich mit meinem Programm "GenbankAnalyse" (siehe [8.4]) Zugriff auf alle Gene in den im Genbank-Format (*.gbk) vorliegenden mikrobiellen Genomen hatte, lag es nahe zu überprüfen, ob die gekoppelte Translation auch bei *Pyrococcus* weit verbreitet ist (was die Hypothese bezüglich einem aktiven, *in vivo* exprimierten rpoE2-Gen wahrscheinlicher erscheinen lassen würde) oder eher die Ausnahme ist.

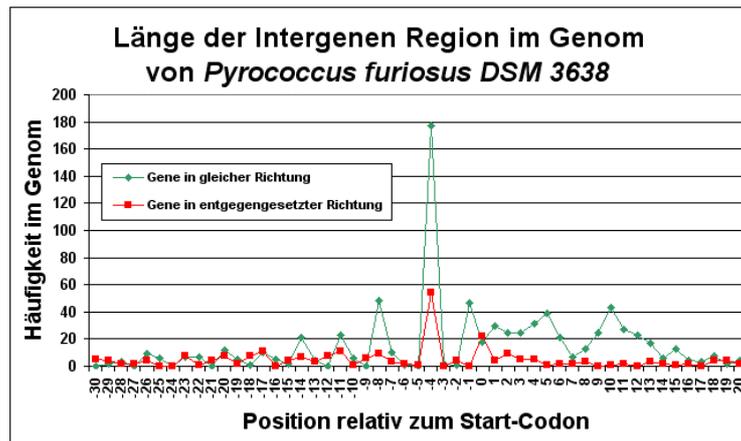


Abb.4-19: Länge der intergenen (intercistronischen) Regionen im Genom von *P.furiosus*

Die auf der X-Achse aufgetragenen Zahlen stehen für Nukleotidpositionen auf DNA-Ebene. 0 bedeutet, dass die Gene direkt ohne Verschiebung des Leserasters aneinander anschließen (also Stop- und Start-Codon direkt hintereinander liegen). Datenbasis sind die (natürlich teilweise hypothetischen) 2065 in der Genom-Datei von *Pyrococcus furiosus* annotierten Proteingene (siehe [8.5]). Positionen einschließlich Start- und Stop-Codon. Die grüne Linie zeigt die Häufigkeit, mit der die beiden hintereinander liegenden Gene die selbe Orientierung zeigen, die rote Linie die Gene mit entgegengesetzter Orientierung (siehe auch die entsprechenden Angaben unter [4.2.3]).

In dem gezeigten Diagramm sieht man auf den ersten Blick ein Häufigkeitsmaximum für Intergene Regionen der Länge "-4". Hierbei ist zu beachten, dass den errechneten Daten die beiden Werte für Beginn und Ende des jeweiligen Genes in der Genomdatei zugrunde liegen, die auch das Stop-Codon umfassen. Dies entspricht dem in der Literatur meist genannten Wert "-1", bei dem das Stop-Codon nicht mitgezählt wird. Für *rpoE1* und *rpoE2* ergibt sich folgende Konstellation:

```
E' . . . ggtaagaagtga
                    gtgagtgaaaaa . . . E''
```

tga = Stop-Codon, **gtg** = alternatives Start-Codon bei *Bacteria* und *Archaea*. Wie man sieht, entspricht die Anordnung der beiden Gene genau dieser häufigsten Anordnung mit festem Abstand. Die "-4-Position" (-1) stellt sowohl bei *Archaea* (überprüft wurde dies auch bei den übrigen *Pyrococcus*-Arten, *Halobacterium* und *Sulfolobus*) als auch bei *Bacteria* (*Escherichia* und die extrem thermophile Art *Thermotoga maritima*) ein Häufigkeitsmaximum dar. Bei *Eukarya* gibt es ein solches Muster natürlich nicht, weil die Genomstruktur grundsätzlich anders ist, und eine Organisation von Genen in Operons sowie polycistronische Transkripte nicht vorkommen. Bei *P.furiosus* zeigen immerhin **228** von **2065** Protein-Genen die "-4"-Konstellation, davon **177** unter Beibehaltung der Lese- bzw. Translationsrichtung und **54** verbunden mit einem Richtungswechsel (die 112 Nicht-Protein-Gene sind unbeteiligt). Hierbei muß bedacht werden, dass für die gekoppelte Translation nur die in der selben Richtung angeordnete Gene in Frage kommen. Bei entgegengesetzter Richtung erfolgt die Expression der Gene natürlich über zwei separate mRNAs, was eine Kopplung ausschließt. Trotzdem bestätigt die Häufung entsprechender Konstellationen in archaeellen und bakteriellen Genomen die Tatsache, dass hier die Gene im Gegensatz zu *Eukarya* viel enger gestaffelt und platzsparend angeordnet sind.

Einige der Minus-Positionen bei überlappenden Genen sind übrigens schon aus rein logischen Gründen ausgeschlossen (jedenfalls bei Protein-Genen): So kann z.B. ein Stop-Codon nicht gleichzeitig als Start-Codon fungieren, weshalb in allen gezeigten Diagrammen die -3-Position ein Minimum darstellt. Weitere erkennbare Regelmäßigkeiten im Minus-Bereich hängen neben den "kombinatorischen" Beschränkungen der Zahl möglicher Codons vielleicht auch damit zusammen, dass nicht alle Positionen gleichermaßen gut für eine Reinitiation der Translation am Ribosom geeignet sind.

4.6 Zusammenfassende grafische Darstellung von homologen Genabschnitten

Eine detaillierte Übersicht über die bei *Archaea* und *Eukarya* homologen bzw. konservierten Bereiche ist deshalb so wichtig, weil die 3D-Struktur von *S.cerevisiae* (inklusive genauer Interpretation, s.u.) neben den Daten der erfolgreichen Rekonstitution der *M.jannaschii*-RNA-Pol aus rekombinanten Untereinheiten (Werner & Weinzierl, 2002) die wichtigste Referenz für die Untersuchungen der archaeellen RNA-Polymerase darstellt (bisher ist eine entsprechende 3D-Struktur abgesehen vom rpoE'-F-Komplex (Todone *et al.*, 2001) für *Archaea* nicht verfügbar). Auch wenn es vermutlich charakteristische Abweichungen in Aufbau und Funktion des archaeellen RNA-Pol-Komplexes gegenüber *Eukarya* gibt, wird die Grundstruktur und der resultierende katalytische Mechanismus in vielen Details den nicht nur für *Eukarya* (mit Pol I – III) und *Bacteria*, sondern auch für monomere DNA-abhängige RNA-Polymerasen wie die T7-RNA-Polymerase gültigen Leitprinzipien folgen (Sonntag & Darai, 1995).

Im folgenden besonders betonte Details sind einerseits das teilweise oder komplette Fehlen von bei *S.cerevisiae* gefundenen Funktionsgruppen bei *Archaea*, und andererseits die Korrelation dieser Sequenzabschnitte mit dem in dieser Arbeit durch umfangreiche Alignments gewonnenen archaeellen Konsensus, was direkt in Form von 3D-Modellen dargestellt wird. Bei Regionen, für die sich keine (oder kaum) Sequenzähnlichkeit zwischen *Archaea* und *Eukarya* findet, kann diskutiert werden, ob es sich für die Funktion der RNA-Polymerase nicht so entscheidende (und deshalb "beliebig" abwandelbare) Bereiche handelt, oder ob es sich gerade um die Stellen handelt, an denen die archaeelle RNA-Polymerase (z.B. als Anpassung an extreme Lebensräume) eine modifizierte Form und ein abweichenden Wirkmechanismus besitzt. Es kann auch die Möglichkeit der Betrachtung der archeellen RNA-Polymerase als eine "herunterskalierte" Form der eukaryotischen Pol II bedacht werden (Eloranta *et al.*, 1998).

Rpb1 (rpoA1 + rpoA2):

Sce Rpb1:

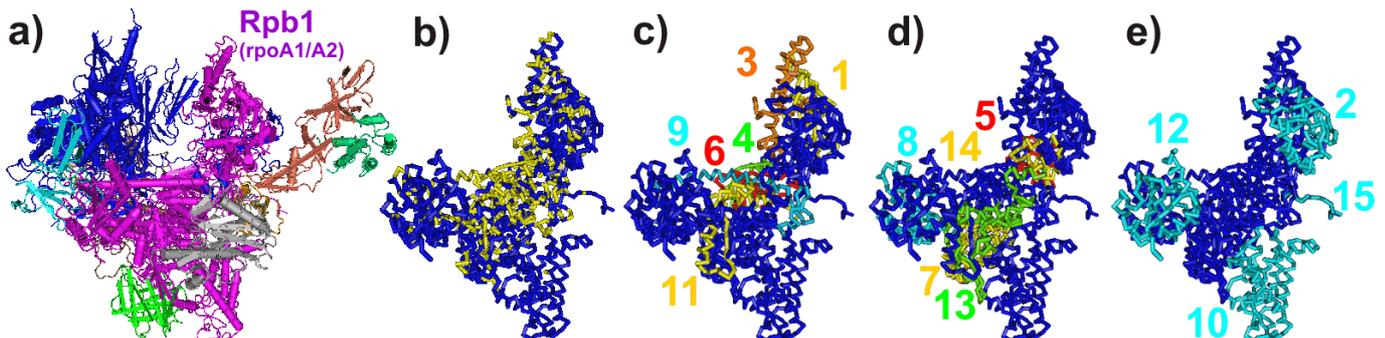
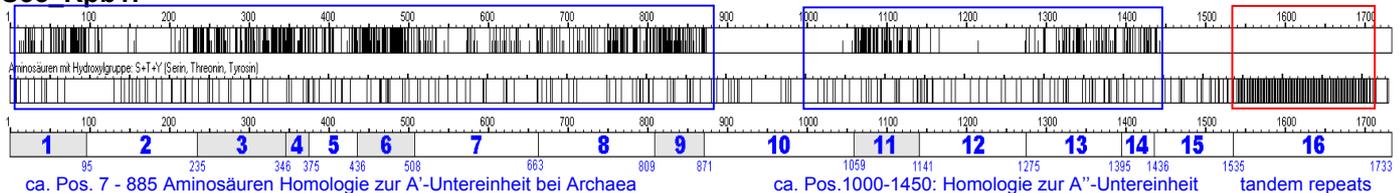


Abb.4-20: Gegenüber *Archaea* konservierte Genabschnitte in Rpb1 aus *S.cerevisiae*

(1733 Aminosäuren Gesamtlänge von Rpb1; obere Zeile Übereinstimmung mit archaeellem Konsensus; mittlere Zeile: Serin, Threonin und Tyrosin markiert, um die Regulation der Aktivität der eukaryotischen RNA-Pol wichtige Phosphorylierungsregion (CTD) hervorzuheben). Zeile mit fettgedruckten blauen Zahlen: Zuordnung von Struktur und (vermutlicher) Funktion entsprechend (Cramer *et al.*, 2001). Englische Bezeichnungen entsprechend Publikation: **1**: „Clamp core“, **2**: „Clamp head“, **3**: „Clamp core“, **4**: „Active site“, **5**: "Dock", **6**: "Active site (Metal A)", **7**: "Pore 1", **8**: "Funnel", **9**: "Cleft", **10**: "Foot", **11**: "Cleft", **12**: "Jaw", **13**: "Cleft", **14**: "Clamp core", **15**: "Linker", **16**: "CTD". Grau hinterlegt: Abschnitte mit besonders hoher Übereinstimmung zu *Archaea*. Die beiden blauen Rahmen zeigt die Dimensionen von rpoA1 und rpoA2 bei *Archaea*. Roter Rahmen: CTD mit "tandem repeats". **a**) Vollständiger RNA-Pol-Komplex zur Orientierung; Anordnung entsprechend "Vorderansicht" (Cramer *et al.*, 2000; Armache *et al.*, 2003). **b**) Regionen mit besonders großen Sequenzhomologien zu *Archaea* entsprechend Alignments (siehe [8.1.3.2] und [8.1.3.3]): gelb = ähnlich, blau = abweichend. **c**) Zuordnung besonders ähnlicher Bereiche zur 3D-Struktur, **d**) Zuordnung von Bereichen mit mittlerer Ähnlichkeit, **e**) Bereiche, die bei *Archaea* komplett fehlen bzw. sehr stark abweichen.

Eine Besonderheit am C-Terminus von Rpb1 aus *S. cerevisiae* ist eine ca. 170 Aminosäure-Positionen lange Phosphorylierungsregion, innerhalb derer ca. 25 mal eine 7 Aminosäuren lange, repetitive Sequenz („YSPTSPS“) mit leichten Variationen wiederholt wird. Diese auch „C-terminal repeat domain“ (CTD) genannte Region (Landick, 2001; Armache *et al.*, 2003) ermöglicht offensichtlich über vielfache Phosphorylierung der Hydroxylgruppen der Serin- (Tryptophan- und Tyrosin-) Reste eine Regulation der Aktivität der eukaryotischen RNA-Pol II (Cramer *et al.*, 2001). Diese auffällige Phosphorylierungsregion ist keine Besonderheit von *S. cerevisiae* allein, sondern stellt offensichtlich ein essentielles Merkmal der größten Untereinheit der Pol II dar. Beispielsweise findet sie sich auch bei *Homo sapiens* (Accession NP_000928). Wie aus der Grafik zu erkennen ist, scheint dieser Mechanismus zur Regulation der Enzymaktivität bei *Archaea* nicht etabliert zu sein – es findet sich keinerlei Hinweis auf ähnliche „tandem repeats“.

Offensichtlich fehlt nicht nur diese CTD, sondern auch ein Großteil des direkt vor dieser Region liegenden "Linkers", der ohne seine vermittelnde Rolle in bezug auf die CTD wahrscheinlich nutzlos wäre. Bei *Eukarya* dient die phosphorylierte CTD abgesehen von der erwähnten regulatorischen Funktion auch als Andockstelle für Enzyme, die für posttranskriptionale Modifikationen verantwortlich sind, d.h. „Capping“, „Splicing“ und Polyadenylierung. Dies ist deshalb sinnvoll, weil die räumliche Nähe dieser Enzyme zum RNAP-Komplex eine möglichst schnelle und effektive Ausführung der Modifikationen der Prä-mRNA ermöglicht. Somit korrespondiert das Fehlen der CTD bei *Archaea* auch mit der einfacheren Struktur der mRNA.

Der Aufbau des archaeellen Homologs zu Rpb1 aus zwei Untereinheiten (rpoA1 und rpoA2) bei allen hier untersuchten *Archaea*, unterstützt die Vorstellung, dass in vielen Fällen die Aufteilung eines großen Proteins in mehrere Teile, weder einem korrekten, spontanen "Assembly", noch einer unverminderten katalytischen Aktivität im Wege stehen muß – sofern nur nicht innerhalb der "Struktur-Module" geteilt wird (Severinov *et al.*, 1996).

Vor diesem Hintergrund muß auch die Aufteilung der B-Untereinheit bei vielen *Archaea* (u.a. den "Methanogenen") betrachtet werden (siehe [Abb.4-05]). Gleiches gilt für E' und E", wobei bei letzterer Untereinheit noch nicht nachgewiesen wurde, dass sie tatsächlich im archaeellen RNA-Pol-Komplex enthalten ist. Rpb7 aus *S. cerevisiae* ist jedenfalls nur zur archaeellen E'-Untereinheit homolog.

Interessant ist, dass sich bei der ansonsten hochkonservierten Untereinheit Rpb1 die "Lücken" in der Ähnlichkeit zum archaeellen Konsensus tatsächlich recht gut den Grenzen der postulierten Struktureinheiten zuordnen lassen: "Clamp head" (2), "Foot" (10) und "Jaw" (12) sind neben der schon erwähnten CTD-Region samt Linker auffallend gering konserviert. Innerhalb von "Clamp head" existiert ein Insert von ca. 33 Aminosäuren ohne Entsprechung bei *Archaea*. Ein noch längeres Insert findet sich an der selben Stelle bei *Halobacterium* (Ng *et al.*, 2000). Der Abschnitt "Foot" entfällt bei *Archaea* komplett. Auch innerhalb "Jaw" finden sich ca. 33 bei *S. cerevisiae* inserierte (bzw. bei *Archaea* deletierte) Aminosäure-Positionen.

"Active site (Metal A)" (6) mit dem katalytischen Zentrum gehört dagegen den Erwartungen entsprechend zu den höchstkonservierten Strukturelementen.

Rpb2 (rpoB):

Sce Rpb2:

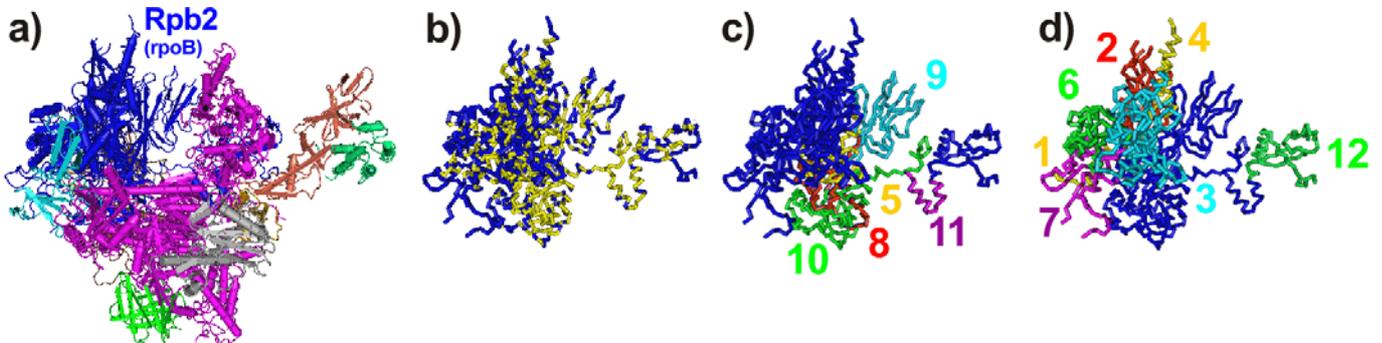
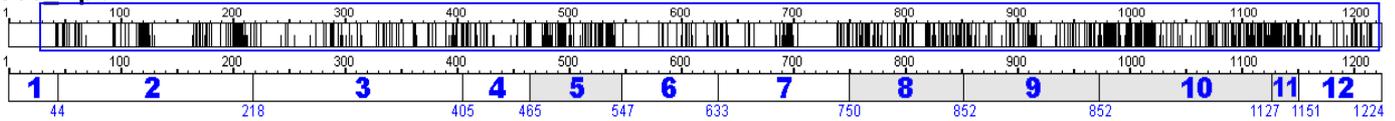


Abb.4-21: Gegenüber Archaea konservierte Genabschnitte in Rpb2 aus *S.cerevisiae*

Durchgezogene schwarze Striche: Identität mit archaeellem Konsensus. Halbe Striche: Ähnlichkeit gegenüber *Archaea*. Zeile mit fettgedruckten blauen Zahlen: Zuordnung von Struktur und (vermutlicher) Funktion entsprechend (Cramer *et al.*, 2001). Englische Bezeichnungen entsprechend Publikation: 1: „External 1“, 2: "Protrusion", 3: „Lobe“, 4: „Protrusion“, 5: „Fork“, 6 "External 2", 7 "External 1", 8 "Hybrid binding", 9 "Wall", 10"Hybrid binding", 11 "Anchor", 12 "Clamp". Der blaue Rahmen zeigt die Dimensionen von rpoB bei *Archaea*. a) Gesamtübersicht der RNAP zur Orientierung (Rpb2 ist hier blau eingefärbt); b) separate Darstellung von Rpb2 (gelb=Ähnlichkeit gegenüber *Archaea* / blau=geringe Ähnlichkeit); c) besonders konservierte Regionen; d) Regionen, in denen die Ähnlichkeit geringer ist.

Die zweitgrößte Untereinheit Rpb2 aus *S.cerevisiae* ist ungefähr gleich lang wie die B-Untereinheit bei *Archaea*, und fügt sich fast über die ganze Länge recht gut in das archaeelle Alignment. Am N-Terminus fehlt ein Großteil von "External 1". Auffällig ist eine separater Abschnitt des Proteinies, der mit "Anchor" und "Clamp" hinüber bis auf die andere Seite der großen Spalte reicht, die hauptsächlich durch Rpb1 gebildet wird. Das Zinkfingermotiv am C-Terminus (innerhalb von "Clamp") ist auch bei *Archaea* perfekt konserviert.

Rpb3 + Rpb11 (rpoD + rpoL):

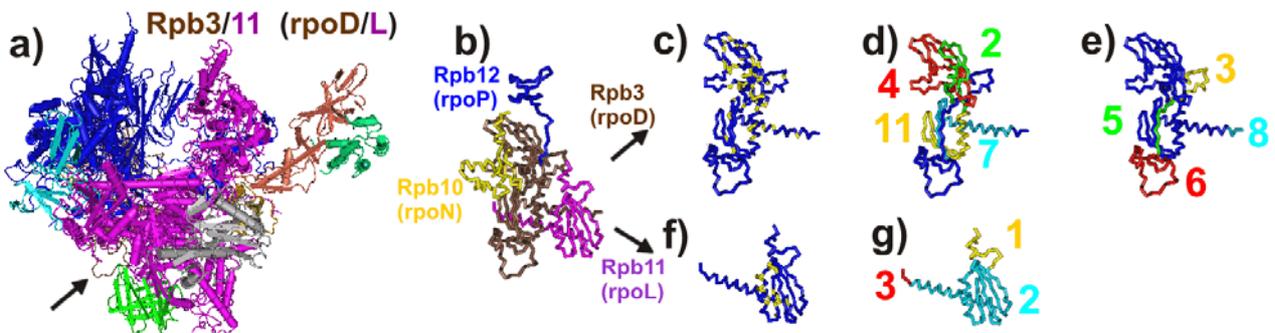
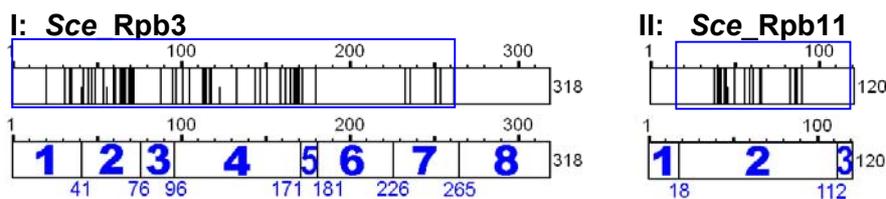


Abb.4-22: Gegenüber Archaea konservierte Genabschnitte in Rpb3 und Rpb11 aus *S.cerevisiae*

Zeile mit fettgedruckten blauen Zahlen: Zuordnung von Struktur und (vermutlicher) Funktion entsprechend (Cramer *et al.*, 2001). Englische Bezeichnungen entsprechend Publikation in I (Rpb3): 1: „Dimerization“, 2: „Domain 2“, 3: „Zinc loop“, 4: „Domain 2“, 5: "Dimerization", 6: "Loop", 7: "Dimerization", 8: "Tail". II (Rpb11): 1: "Tail", 2: "Dimerization", 3: "Tail". Die restlichen Symbole wie in der vorhergehenden Abbildung. Der Pfeil in a) weist auf einen sichtbaren Abschnitt von Rpb3 (der Rpb3/11/10/12-Komplex wird ansonsten weitgehend durch andere Untereinheiten verdeckt).

a) Gesamtübersicht der RNAP zur Orientierung; b) Darstellung des Rpb3/11/10/12-Komplexes (rpoD/L/N/P) zur

Orientierung; **c)** Rpb3 (gelb=Ähnlichkeit gegenüber Archaea / blau=geringe Ähnlichkeit); **d)/e)** Einfärbung des Rpb3-Moleküls entsprechend den genannten Struktur-/Funktionseinheiten; **f)** Rpb11 (gelb=Ähnlichkeit gegenüber Archaea / blau=geringe Ähnlichkeit); **g)** Rpb11, siehe **d)/e)**.

rpoD, das archaeelle Homolog zu Rpb3 zeigt insgesamt geringe Ähnlichkeit innerhalb *Archaea*. Der gegenüber *Archaea* ca. 49 Aminosäuren längere C-Terminus von Rpb3 betrifft das Element "Tail". In "Domain 2" (**4**) finden sich 9 zusätzliche Positionen bei *S.cerevisiae*. Bei "Loop" (**6**) fehlen ca. 21 Positionen gegenüber *Archaea*, wobei sich auch der Rest dieses Abschnittes nicht gut in das Alignment fügt (siehe [8.1.3.4]).

Rpb11 aus *S.cerevisiae* ist im Bereich von ca. 14 – 120 homolog zum archaeellen rpoL. Bemerkenswert ist, dass rpoD und rpoL, die beiden zu Rpb3 und Rpb11 homologen Proteine bei *Pyrococcus* die am stärksten vom Konsensus abweichenden RNAP-Untereinheiten darstellen. Gerade Rpb3 / rpoD nimmt aber im Rpb3/11/10/12 (rpoD/L/N/P)-Komplex eine zentrale Position ein, was mit einer allgemein hohen Mutationsrate eigentlich nicht vereinbar scheint. Vielleicht ist dies ein Bereich, wo sich artspezifische Anpassungen besonders auswirken?

Rpb7 + Rpb4 (rpoE1 + rpoF)



Abb.4-23: Gegenüber *Archaea* konservierte Genabschnitte in Rpb7 und Rpb4 aus *S.cerevisiae*

Die Abbildungen dieses Abschnitts ([4.6]) haben einen starken Bezug zu der Publikation, in der auf der Basis einer 3D-Struktur der RNAP aus *S.cerevisiae* mit 10-Untereinheiten (noch ohne Rpb7 und Rpb4) eine Gliederung in Struktur- und Funktionseinheiten vorgenommen wird (Cramer *et al.*, 2001). In der aktuellen Publikation, die sich auf die entsprechende 3D-Struktur der *S.cerevisiae*-RNAP mit allen 12 Untereinheiten (also inklusive Rpb7 und Rpb4) bezieht (Armache *et al.*, 2003) wird diese Gliederung nicht explizit fortgeführt. Für die 3D-Darstellung sei deshalb auf [Abb.4-27] verwiesen.

Blauer Rahmen: Dimension der zu Rpb7 und Rpb4 homologen RNAP-Untereinheiten bei *Archaea*. Durchgezogene schwarze Striche: Identität, halbe Striche: Ähnlichkeit gegenüber dem archaeellen Konsensus.

Die Homologie von *Sce_Rpb7* zur archaeellen Untereinheit E' scheint nicht übermäßig groß zu sein. Der ungefähr in der Mitte der Aminosäuresequenz erkennbare Bereich höherer Ähnlichkeit bezieht sich auf den Interfacebereich zwischen Rpb7 und Rpb4. Der Abschnitt, der den Kontakt zum Rest der Polymerase herstellt ist nicht so stark konserviert wie vermutet. Hiervon abgesehen hat *Sce_Rpb7* ziemlich genau die Größe seiner archaeelle Entsprechung rpoE1.

Bei Rpb4 aus Pol II findet man spärliche Identität mit dem archaeellen Konsensus und am N-Terminus 100 zusätzliche Aminosäuren ohne Entsprechung bei *Archaea*.

E'' (rpoE2) – bisher kein Homolog bei *Eukarya* identifiziert

Rpb7 besitzt nur Homologie zu E'; stellt rpoE2 also eine Besonderheit bei *Archaea* dar? Auf jeden Fall ist das Gen mit seinem Zinkfinger-Motiv genügend konserviert, um eine spezifische Funktion im Rahmen von Protein-Protein- oder Protein-DNA-Wechselwirkungen anzunehmen. *Sulfolobus acidocaldarius* stellt mit seiner einheitlichen (sowohl den E'- als auch E''-Anteil enthaltenden) E-Untereinheit eine Besonderheit gegenüber allen übrigen hier untersuchten *Archaea* dar (Lanzendörfer *et al.*, 1994).

Rpb5, Rpb6 (rpoH, rpoK):

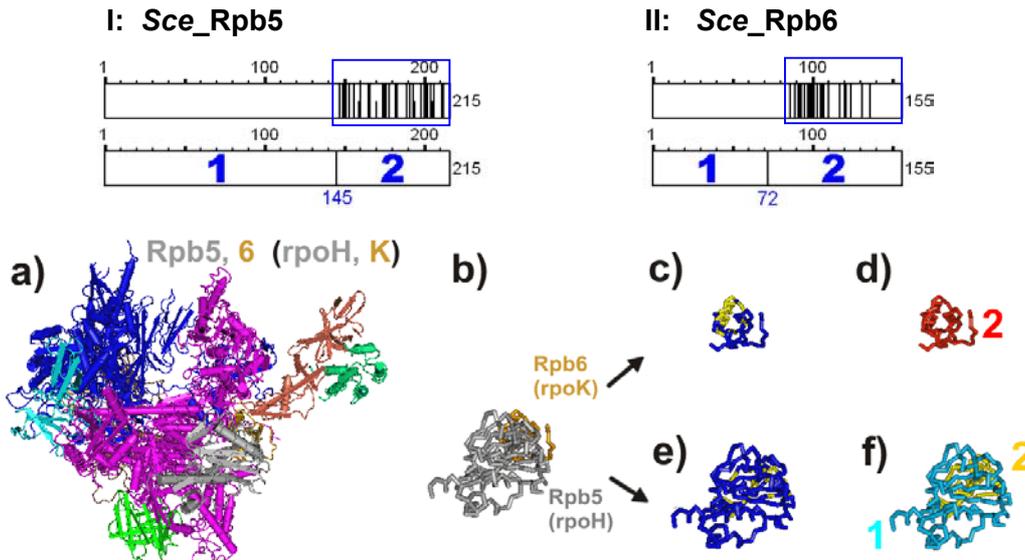


Abb.4-24: Gegenüber Archaea konservierte Genabschnitte in Rpb5 und Rpb6 aus *S.cerevisiae*

Zeile mit fettgedruckten blauen Zahlen: Zuordnung von Struktur und (vermutlicher) Funktion entsprechend (Cramer *et al.*, 2001). Englische Bezeichnungen entsprechend Publikation: **I (Rpb5):** 1: „Jaw“, 2: „Assembly“. **II (Rpb6):** 1: „Tail“ (nicht im 3D-Modell enthalten), 2: „Assembly“. Die restlichen Symbole wie in der vorhergehenden Abbildung.

a) Gesamtübersicht der RNAP zur Orientierung; **b)** Gemeinsame Darstellung von Rpb5 und Rpb6 (welches beim gewählten Blickwinkel durch Rpb5 verdeckt wird) zur Orientierung (siehe auch [Abb.4-49]); **c)** Rpb6 (gelb=Ähnlichkeit gegenüber Archaea / blau=geringe Ähnlichkeit); **d)** Einfärbung des Rpb6-Moleküls entsprechend den genannten Struktur-/Funktionseinheiten; **e)** Rpb5 (gelb=Ähnlichkeit gegenüber Archaea / blau=geringe Ähnlichkeit); **f)** Rpb5, entsprechend **d)**.

Zu rpoH homologe Region 143-215 in Rpb5 aus *S.cerevisiae*. Mit dem Konsensus übereinstimmende Positionen wurden schwarz markiert (und blau eingerahmt). Man beachte, dass die "Jaw" ("Kiefer") genannte Region bei *Archaea* komplett fehlt.

Auch Rpb6 ist ein Beispiel für eine eukaryotische RNAP-Untereinheit, bei der das archaeele Homolog (rpoK) deutlich kleiner ist. Es finden sich zusätzliche 83 Aminosäuren am N-Terminus, die kein Gegenstück bei *Archaea* haben. Ein ziemlich eng umgrenzter Bereich mit einer α -Helix im Abschnitt "Assembly" scheint auch bei *Archaea* von entscheidender Bedeutung (und deshalb hochkonserviert) zu sein.

Rpb10 + Rpb12 (rpoN + Rpb12):

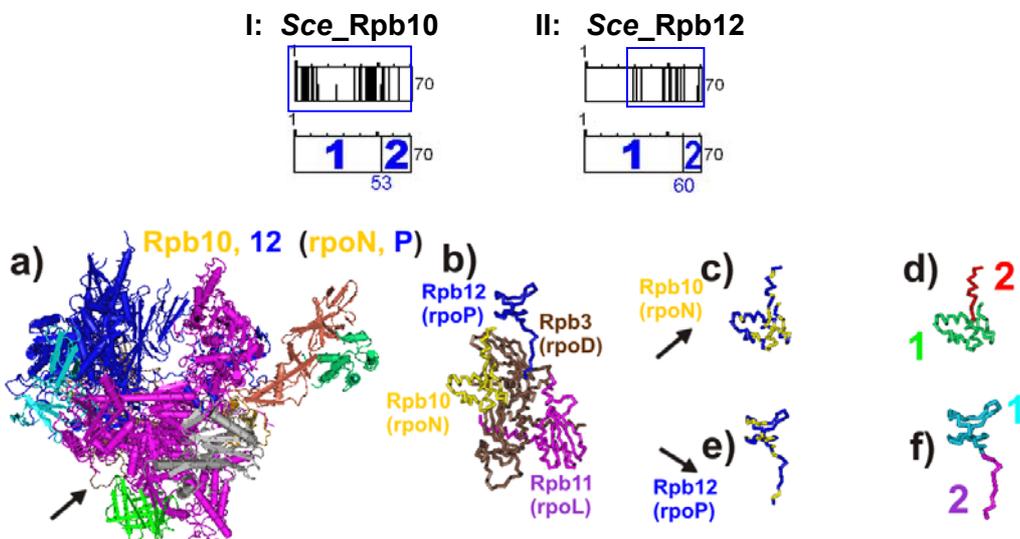


Abb.4-25: Gegenüber Archaea konservierte Genabschnitte in Rpb10 und Rpb12 aus *S.cerevisiae*

Zeile mit fettgedruckten blauen Zahlen: Zuordnung von Struktur und (vermutlicher) Funktion entsprechend (Cramer *et al.*, 2001). Englische Bezeichnungen entsprechend Publikation: **I (Rpb10):** 1: „Zink Bundle“, 2: „Tail“, **II (Rpb12):** 1: „Zink Ribbon“, 2: „Tail“. Die restlichen Symbole wie in der vorhergehenden Abbildung.).

a) Gesamtübersicht der RNAP zur Orientierung; **b)** Darstellung des Rpb3/11/10/12-Komplexes (rpoD/L/N/P) zur Orientierung; **c)** Rpb10 (gelb=Ähnlichkeit gegenüber Archaea / blau=geringe Ähnlichkeit); **d)** Einfärbung des Rpb10-Moleküls entsprechend den genannten Struktur-/Funktionseinheiten; **e)** Rpb12 (gelb=Ähnlichkeit gegenüber Archaea / blau=geringe Ähnlichkeit); **f)** Rpb12 entsprechend **d)**.

Rpb10 hat die gleiche Größe wie die N-Untereinheit der RNA-Polymerase bei *Archaea* und besitzt offensichtlich konservierte Bereiche hauptsächlich an der zum Inneren der RNAP gerichteten Seite. Rpb12 zeigt im Bereich von 26 – 70 homologe Bereiche gegenüber rpoP. Auch einige *Archaea* haben einen relativ zum Konsensus verlängerten N-Terminus. Interessant ist der auffällig basische "Schwanz" von rpoP (siehe [Abb.4-15], dem Homolog zu Rpb12 in *Pyrococcus*).

Rpb8 und Rpb9 besitzen nach gegenwärtigem Stand der Forschung keine homologen Entsprechungen bei *Archaea*. Rpb8 sitzt an der „Unterseite“ des RNAP-Komplexes und ist an der Bildung der „Pore“ bzw. des Trichters beteiligt, durch den die Nucleotide (NTPs) zum katalytischen Zentrum gelangen (siehe Schema [Abb.2-03]). Als ein Homolog zu Rpb9 wurde zunächst ein archaeelles Protein betrachtet, dem man den Namen „rpoM“ gab (z.B. bei *Thermococcus celer*: swissprot: RPOM_THECE), in der Annahme, dass es sich ebenfalls um eine RNAP-Untereinheit handelt. Es konnte jedoch gezeigt werden, dass es sich hierbei um einen zum eukaryotischen TFIIS homologen Elongationsfaktor der archaeellen Transkription handelt (Hausner *et al.*, 2000), der als TFS bezeichnet wird. Dass Rpb9 bei *Archaea* komplett fehlt ist insofern bemerkenswert, als es bei *Eukarya* eine wichtige Funktion bei der Auswahl des richtigen Transkriptionsstartpunktes spielt und offensichtlich an der Vermittlung der (RNA-) Schneideaktivität von TFIIS bei der Elongation der Transkription wesentlich beteiligt ist (Hull *et al.*, 1995; Hemming *et al.*, 2000; Cramer *et al.*, 2001).

4.6.1 3D-Struktur der RNA-Pol II mit gegenüber *Archaea* konservierten Regionen

Als wichtigste Referenz für die Konstruktion eines Modells der archaeellen RNA-Polymerase dient momentan die 3D-Struktur der 12 Untereinheiten umfassenden RNA-Pol II aus *S.cerevisiae* (Armache *et al.*, 2003). Für die Domäne der *Archaea* ist bisher keine entsprechende komplette 3D-Struktur verfügbar – nur die Kristallstruktur einiger Untereinheiten und kleine Komplexen (E'-F-Komplex bei *M.jannaschii* (Todone *et al.*, 2001)) sind bis jetzt als Datenbank-Einträge verfügbar (siehe [8.2]). Sehr wichtig ist die genaue Kenntnis dieser Strukturen für eine Plausibilitätskontrolle vor allem der Far Western Blotting-Experimente, wo in manchen Fällen die Entscheidung schwer fällt, welche Signal-Stärke noch als ein Hinweis auf Kontakt zweier RNA-Pol-Untereinheiten interpretiert werden darf, und wo der „Background“ beginnt – mit schwachen Signalen, die bei der Analyse ignoriert werden können (und müssen!). Falls starke Signale dem eukaryotischen 3D-Modell widersprechen, kann dies als mögliche Besonderheit von *Archaea* (bzw. *P.furiosus*) diskutiert werden.

Besondere Bedeutung gewinnt diese 3D-Referenz dadurch, dass es in dem benutzen Programm Cn3D 4.1 möglich ist, über eine mächtige Suchfunktion viele Sequenz-Muster gleichzeitig hervorzuheben. So konnten in jeder der 12 Untereinheiten die mit dem archaeellen Konsensus identischen und ähnlichen Regionen (siehe Alignments [8.1.3]) in der 3D-Ansicht farblich hervorgehoben werden.

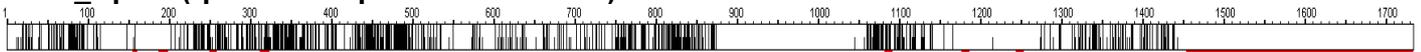
Die Umrechnung des auf *S.cerevisiae* bezogenen Konsensus in das benötigte Prosite-Suchmuster erfolgte mit dem Programm „GenbankAnalyse“ (siehe [8.4]). Beispiel: Der Anfang der Konsensus-Sequenz für Rpb1:

R-x(1)-V-x(2)-V-x(1)-FG-x(2)-SP-x(1)-EVR-x(1)-IS-x(7)-ET-x(1)-D-x(8)-GG-x(2)-D...(usw.);

(die Formulierung „-x(7)-“ steht beispielsweise für 7 beliebige Aminosäuren; „-x(1,7)“ erlaubt, dass die zwischen zwei Aminosäuren liegende Region 1 bis 7 beliebige Aminosäuren umfassen darf; mit der Formulierung „C-x(1,2)-C“ lassen sich beispielsweise schnell Zinkfinger motive vom C4-Typ finden – z.B. in Rpb2; mit "[c,h]-x(1,2)[c,h]-x(1,20)-[c,h]-x(1,2)[c,h]" werden z.B. auch Motive vom C2H2-Typ etc. erfaßt; die Angaben in eckigen Klammern stehen für "entweder oder"). Eine Einschränkung gibt es trotzdem: In den meisten der 12 diskutierten RNA-Pol II-Untereinheiten sind einige Sequenz-Abschnitte nicht aufgelöst. Die genaue Position dieser Lücken in der 3D-Struktur muß mit in eine seriöse Interpretation einbezogen werden, vor allem bei Aussagen wie „Untereinheit X kann keinen Kontakt zu Untereinheit Y haben, weil die beiden Proteine in der 3D-Struktur deutlich sichtbar voneinander getrennt angeordnet sind“.

Zunächst also eine Übersicht über diese Lücken in Kombination mit der Angabe der jeweils zum archaellen Konsensus homologen Sequenz-Abschnitte:

Sc_e_Rpb1 (rpoA1 und rpoA2 bei *Archaea*):



Sc_e_Rpb2 (rpoB bei *Archaea*):



Sc_e_Rpb3 & Sc_e_Rpb11 (rpoD und rpoL bei *Archaea*):



Sc_e_Rpb7 & Sc_e_Rpb4 (rpoE1 und rpoF bei *Archaea*):



Sc_e_Rpb5, Sc_e_Rpb6, Sc_e_Rpb10, Sc_e_Rpb12 (rpoH, rpoK, rpoN und rpoP bei *Archaea*):

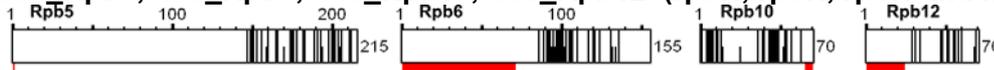


Abb.4-26: Nicht aufgelöste Abschnitte im 3D-Modell der *S.cerevisiae*-RNA-Polymerase

Rote Markierung: Bereiche, die in der verwendeten 3D-Struktur (siehe [8.1]) nicht aufgelöst wurden. Schwarze Striche: Mit dem Konsensus der archaellen multiplen Alignments identische Aminosäuren (siehe [8.1.3]); halbhohle Striche: Positionen, die entsprechend "ähnlich" sind (siehe [8.1.3.12]).

Wer schon einmal mit der Kristallisierung von Proteinen oder Protein-Komplexen zu tun hatte, weiß, wieviel Erfahrung, Know-How und Glück zu einer erfolgreichen Gewinnung eines für die Röntgenstrukturanalyse geeigneten Kristalls gehört. Selbst wenn dies gelungen ist, stellt die Interpretation der gewonnenen Strukturdaten aus einem 12-Untereinheiten-Komplex wie der RNA-Polymerase eine weitere Herausforderung dar. Aus verschiedenen Gründen ist das vorliegende 3D-Modell also nicht ganz komplett, was aber seine Bedeutung als Hauptreferenz für vergleichende Überlegungen zum Thema „RNA-Polymerase“ kaum schmälert. In der oben gezeigten Grafik wurden die nicht im 3D-Modell enthaltenen Regionen der jeweiligen RNA-Pol-Untereinheiten rot markiert (bei Rpb1 fehlt z.B. die C-terminale – für die eukaryotische Enzymregulation wichtige Phosphorylierungsregion (CTD = C-terminal repeat domain), weil dieser besonders bewegliche Proteinabschnitt im Kristall keine geordnete Struktur bildet (Landick, 2001; Armache *et al.*, 2003). Allerdings fehlt dieser Bereich bei *Archaea* ohnehin.). Zum auf der Basis von Alignments gewonnenen archaellen Konsensus homologe Sequenzabschnitte sind schwarz markiert, und dienen im folgenden zur Illustration des „archaellen Kerns“ der eukaryotischen 3D-Struktur. Diese

Formulierung ist nicht als Wertung im phylogenetischen Sinne gemeint – die Diskussion, in welcher Relation sich *Archaea*, *Bacteria* und *Eukarya* im Detail (also z.B. speziell in bezug auf die Gene des Transkriptionsapparates) zu einander befinden, ist immer wieder Gegenstand aktueller Veröffentlichungen. Der Begriff „Homologie“ wird im folgenden meist im Sinne einer (in Alignments) offensichtlichen Sequenz-Übereinstimmung verwendet; natürlich ist ein Protein auch dann noch als homolog zu betrachten, wenn nur noch ein sehr kleiner Bereich deutliche Ähnlichkeiten in der Aminosäure-Sequenz zeigt (entscheidend ist, dass mit einer verfügbaren Methode eine Abstammung von einem gemeinsamen Vorläufer nachgewiesen werden kann).

Darstellung der 3D-Struktur der *S.cerevisiae* RNA-Pol II mit allen 12 Untereinheiten:

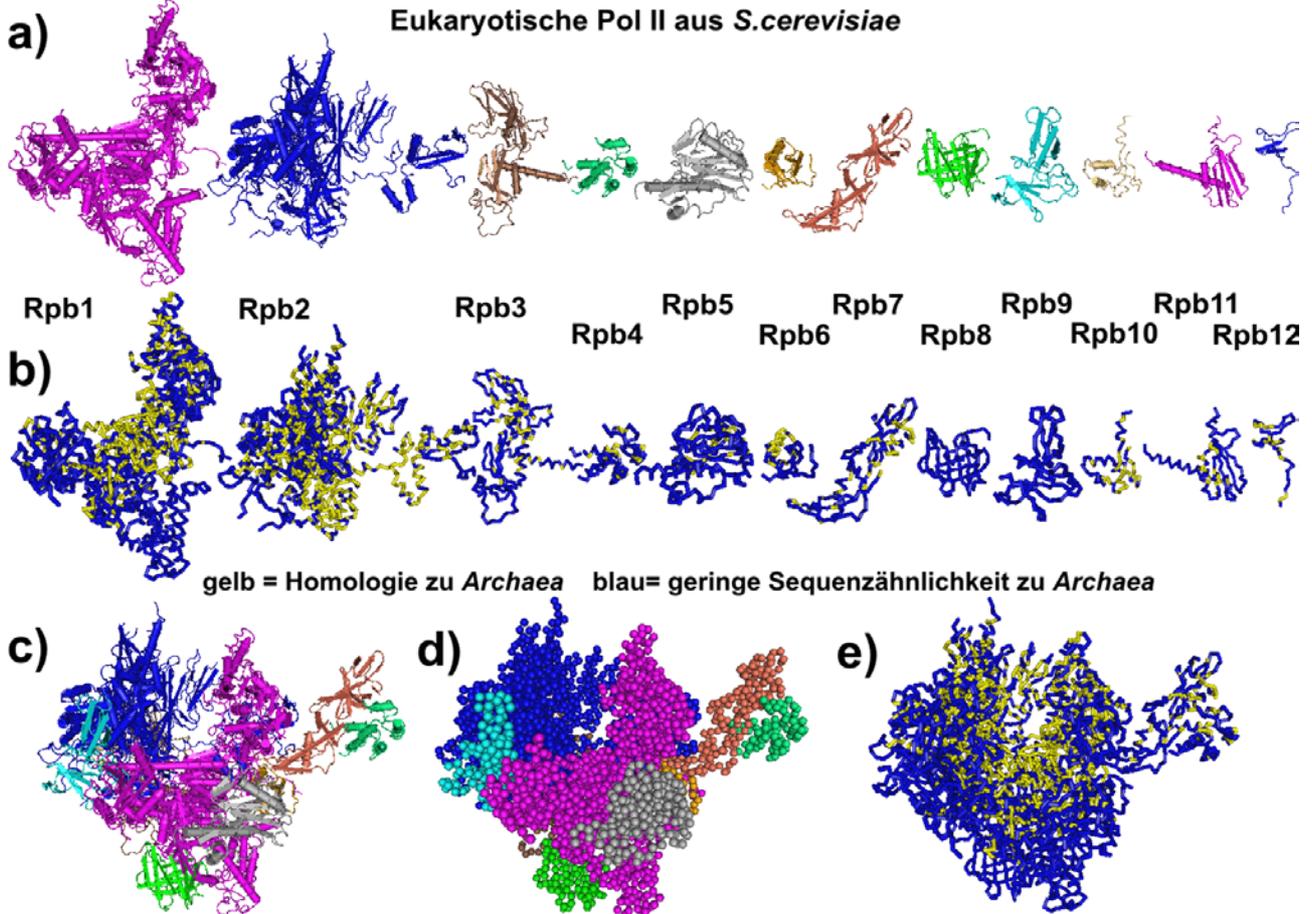


Abb.4-27: 3D-Struktur inklusive Visualisierung der gegenüber *Archaea* ähnlichen Regionen

3D-Struktur von *S.cerevisiae* (siehe [8.2]), Blick von "vorn", also bezogen auf die Richtung der Transkription und die (hier nicht eingezeichnete DNA) von stromabwärts nach stromaufwärts (alle in dieser Zusammenstellung gezeigten 3D-Moleküle mit identischem Blickwinkel). **a)** Einzelne Untereinheiten des in **c)** (und **d)** in einer etwas anderen Darstellungsweise gezeigten) Gesamtkomplex. Die Farbgebung ist einheitlich, um das Auffinden der einzelnen Untereinheiten im in der RNAP zu erleichtern. **b)** Untereinheiten entsprechend **a)**, jedoch mit gelber Farbe für gegenüber *Archaea* identischen und ähnlichen Regionen, sowie blau für Abschnitte mit geringer Sequenzähnlichkeit. Bemerkenswert ist die schon hier vielfach erkennbare lokale Konzentration der Ähnlichkeit auf eng begrenzte Bereiche. **e)** Darstellung der in **b)** demonstrierten Verteilungsmuster am Gesamtkomplex. Gut ist zu erkennen, dass sich die gegenüber *Archaea* besonders konservierten Bereiche auf die Innenseiten und den Grund der großen Spalte mit dem katalytischen Zentrum und dem Ort vielfältiger DNA-Protein- und RNA-Protein-Kontakte beschränken.

Die Darstellung der hier gezeigten 3D-Strukturen der Untereinheiten und des Gesamtkomplexes der eukaryotischen Pol II in Verbindung mit einer Visualisierung der gegenüber *Archaea* besonders konservierten Regionen auf Basis der selbst generierter Alignments stellt eine wichtige Aussage des theoretischen Teils dieser Arbeit dar, und wäre in dieser Form ohne die kürzliche Veröffentlichung der vorgestellten 3D-Struktur (Armache *et al.*, 2003) nicht möglich gewesen (zumindest war vorher im 3D-Modell nicht der wichtige Komplex aus Rpb4 und Rpb7 enthalten (Cramer, 2002), der bei *Archaea* durch rpoE' und rpoF gebildet wird). Die eigentlichen Grafiken wurden mit dem von NCBI kostenlos zur Verfügung gestellten, 3D-Programm „Cn3D 4.1“ hergestellt, wobei die

Markierung der „archaeellen Regionen“ wie oben beschrieben über Prosite-Suchmuster erfolgte (obwohl das Programm eigentlich ausgefeilte Möglichkeiten für die Behandlung von Alignments enthält, ist dies der einfachste Weg; die weitergehenden Programm-Funktionen liefern teilweise nicht besonders hilfreiche Fehlermeldungen, so dass eine längere Einarbeitungszeit und die Erarbeitung des Wissens, in welchen Konstellationen diese Funktionen überhaupt zur Verfügung stehen nötig wäre).

Eine weitere Abbildung mit Darstellungen der konservierten Bereiche des Gesamtkomplexes aus verschiedenen Blickwinkeln heraus rundet die Übersicht ab:

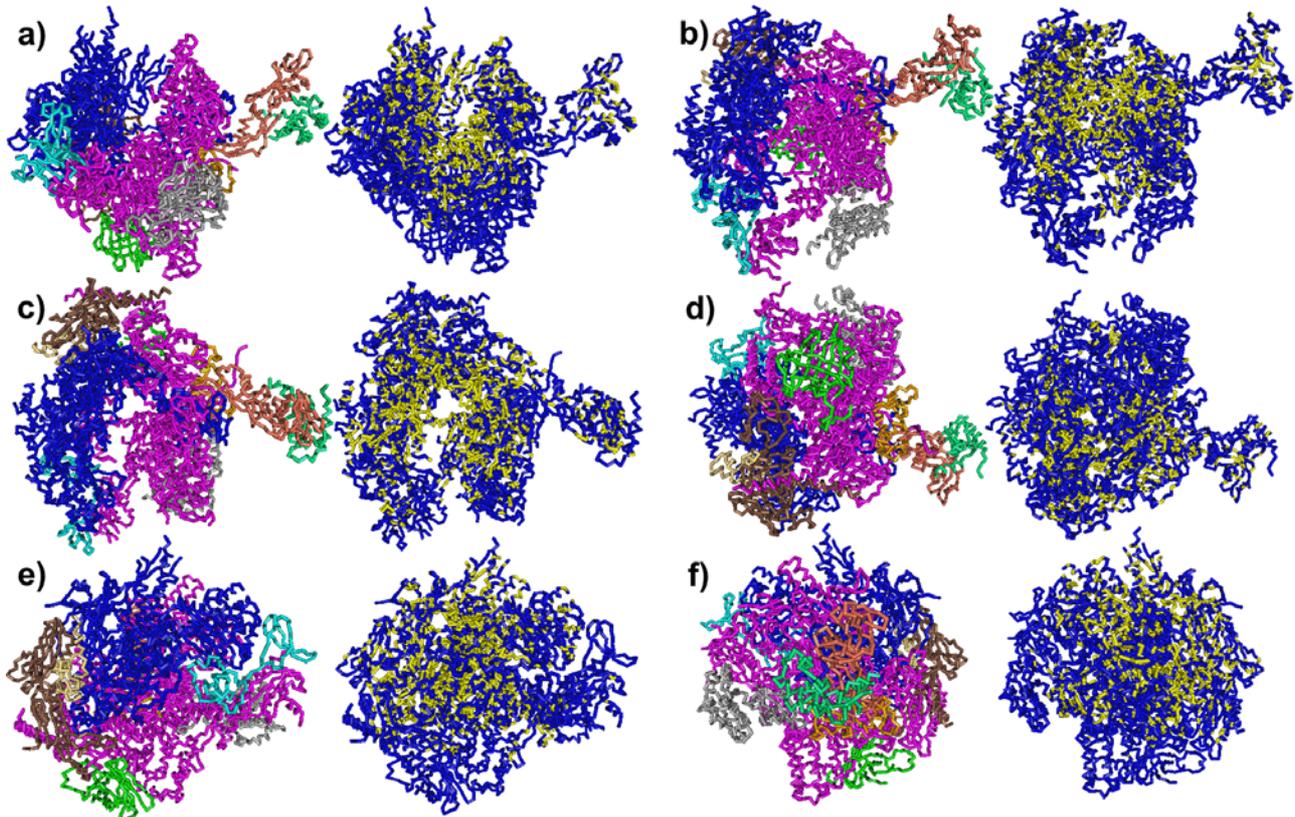


Abb.4-28: 3D-Struktur mit gegenüber *Archaea* konservierten Regionen, Gesamtkomplex

Da in der gedruckten Form das Molekül nicht wie im 3D-Programm (in diesem Fall Cn3D 4.1) interaktiv gedreht werden kann, um eine Vorstellung von der räumlichen Ausdehnung der konservierten Regionen zu gewinnen, wird der RNAP-Komplex in **a) – f)** aus 6 verschiedenen Blickwinkeln gezeigt. Zur besseren Orientierung wurde jeweils eine Molekül-Version mit farblich markierten Untereinheiten hinzugefügt (die Zuordnung der Farben ist dieselbe wie in der oben gezeigten Übersicht mit den einzelnen Untereinheiten; gelb = große Ähnlichkeit).

a) zeigt die Orientierung des RNA-Pol-Komplexes „von vorn“, entsprechend (Armache *et al.*, 2003), also von der Seite in die Spalte mit dem katalytischen Zentrum und den für die Bindung von DNA, mRNA und Ribonukleotiden wichtigen konservierten Bereichen hinein. **b)** nach Drehung um ca. 90° nach vorn, also von oben in die Spalte hinein; **c)** nach weiterer Drehung um ca. 90° nach vorn, mit seitlichem Blick in die Spalte „von hinten“, also bezogen auf die DNA von stromaufwärts; **d)** nach weiterer Drehung mit Blick von unten; **e)** ausgehend von **a)** Drehung um ca. 90° entgegen dem Uhrzeigersinn um eine vertikale Achse; **f)** entsprechend **e)**, allerdings in die entgegengesetzte Richtung, so dass der „Keil“ mit Rpb7 und Rpb4 (also den rpoE1- und rpoF-Homologen) nach vorn weist. Neben **a)** ist vor allem in **b)** und **c)** die stark konservierte Region um das katalytische Zentrum herum zu erkennen. **d) – f)** zeigen den RNAP-Komplex aus Blickwinkeln, bei denen die stark veränderte Außenhülle die bei *Eukarya* gegenüber *Archaea* ähnlichen Bereiche weitgehend verdeckt. Ein Vergleich des Ähnlichkeitsmuster zwischen **a)** und **c)** zeigt, dass sich die konservierten Regionen bei **c)** deutlich stärker „aus der großen Spalte hinaus“ in den Außenbereich hinein erstrecken als in **a)**. Diese Beobachtung paßt gut zu der Tatsache, dass **c)** die Seite zeigt, mit der die RNAP an den Präinitiationskomplex bindet, und zu diesem Zweck ein charakteristisches „Interface“ bildet, das aus funktionalen Gründen nicht zu sehr verändert werden darf.

Zwar entspricht es der Erwartung, dass das bei der Untersuchung von homologen Enzymen für die biologische Funktion besonders wichtige katalytische Zentrum stark konserviert ist. Bei einem so großen und aus soviel verschiedenen Untereinheiten bestehenden Komplex wie dem der RNA-Polymerase ist aber das erkennbare Konsensus-Muster (in der 3D-Struktur gelb markiert) erstaunlich homogen und schlüssig. Diese Darstellung ergänzt den in (Cramer *et al.*, 2001) unter

(Fig.11 und Fig.12) grafischen Vergleich homologer Bereiche einerseits zwischen Hefe und Mensch, und andererseits zwischen Hefe und *Bacteria*. Auch hier ist der zentrale Bereich um das katalytische Zentrum herum besonders konserviert.

Die in dieser Arbeit herausgearbeitete Ähnlichkeitsverteilung bezieht sich einerseits auf „*Archaea*“ versus „*Eukarya*“. Andererseits ist höchstwahrscheinlich auch der Faktor „Hyperthermophilie“ beteiligt, weil die meisten Vertreter in den zugrunde liegenden Alignments hyperthermophil sind.

In diesem Sinne könnte die starke Veränderung der äußeren Hülle des RNA-Pol-Komplexes als notwendige, überlebenswichtige Anpassung an die Bedingungen extremer (u.a. heißer) Standorte der *Archaea* diskutiert werden. Umgekehrt könnten die offensichtlichen Unterschiede bei *Eukarya* als Anpassung an die veränderten Anforderungen betrachtet werden, die sich aus der Integration in den in vieler Hinsicht komplexeren (mesophilen) Zellstoffwechsel höherer Lebewesen ergeben (z.B. in bezug auf eine Vielzahl von RNAP-assoziierte Hilfsproteine, die keine direkte Entsprechung bei *Archaea* haben).

4.7 Liefert die Aminosäuresequenz Hinweise auf Hitzestabilität bzw. Löslichkeit?

Auf den folgende Seiten wird untersucht, wie über eine Änderung der Häufigkeit bestimmter Aminosäuren ein Beitrag zu höherer Thermoadaptation erzielt wird, und welche Konsequenzen dies für die RNA-Pol-Untereinheiten hat.

Untereinheiten der RNA-Polymerase aus *P.furiosus*: Aminosäure-Frequenz

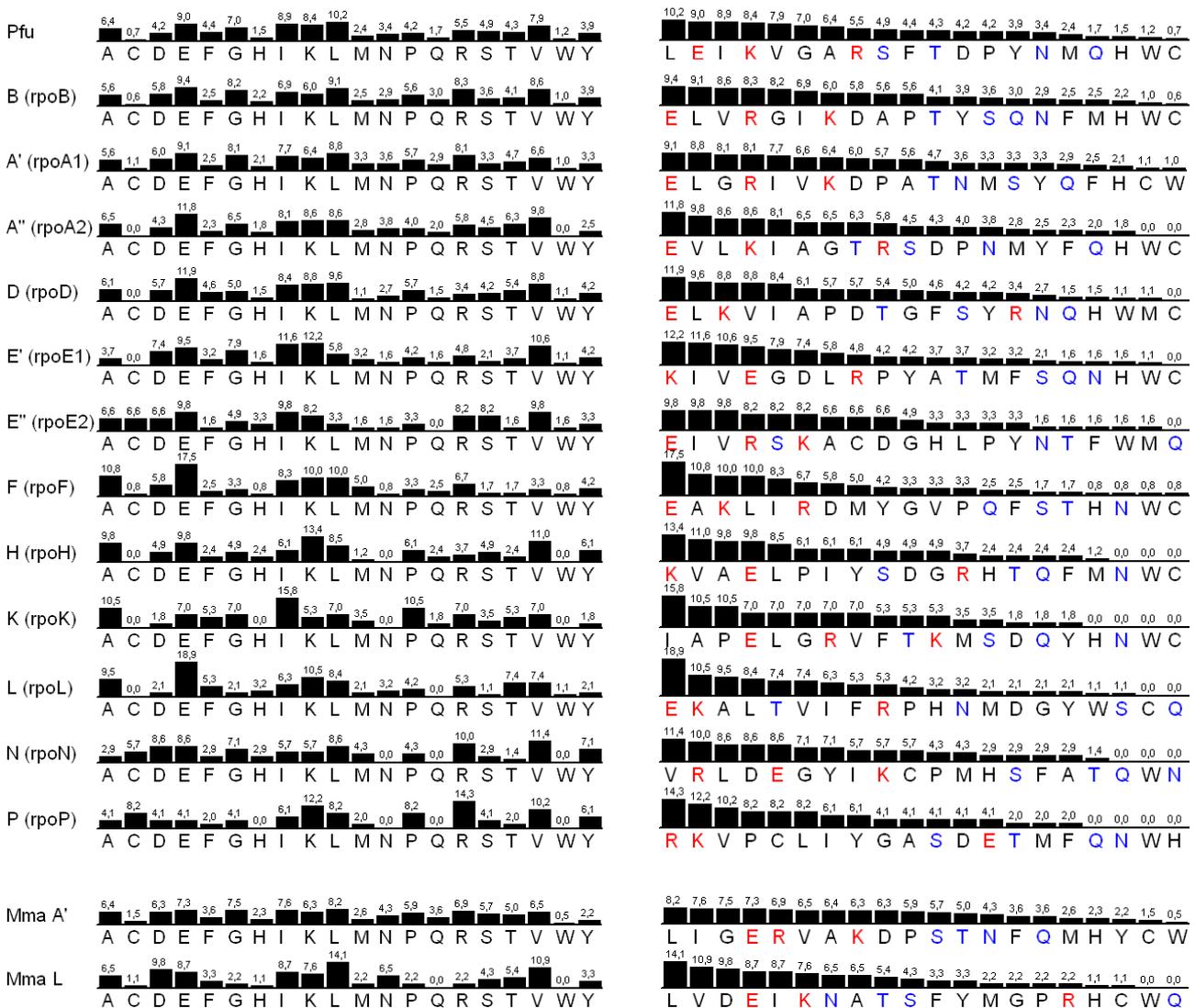


Abb.4-29: Prozentuale Aminosäurehäufigkeit in den RNA-Pol-Genen von *P.furiosus*

Die erste Zeile dieser Darstellung mit der Bezeichnung „Pfu“ stellt eine Referenz-Zeile mit den Durchschnittswerten aus 2065 Protein-Genen aus *P.furiosus* (siehe weiter unten, [Abb.4-34]). Es handelt sich also um den Erwartungswert, von dem die untersuchten Untereinheiten der RNA-Polymerase mehr oder weniger signifikant abweichen. Die Zahlen sind die Prozentwerte der Häufigkeit einzelner Aminosäuren in dem jeweiligen Protein.

In der rechten Spalte der Grafik wurden mit roter Farbe die Aminosäuren (E+K+R) gekennzeichnet, deren häufiges Auftreten auf eine Thermoadaptation des Organismus hinweisen. Ebenso soll die verringerte Frequenz der blau gekennzeichneten Aminosäuren (N+Q+S+T) einen Hinweis auf Thermostabilität liefern. Dieses aus einem Vergleich von mesophilen und hyperthermophilen Vertretern des Genus *Methanocaldococcus* (früher *Methanococcus*) entwickelte Modell (Haney *et al.*, 1999) scheint hier recht gut anwendbar: Sowohl das gesamte Verteilungsmuster (1.Zeile) als auch die Signatur der einzelnen RNA-Pol-Gene weist *Pyrococcus* als einen extrem Hyperthermophilen aus (zumindest findet sich auf Basis der genannten Kriterien kein

offensichtlicher Widerspruch). Unter [Abb.4-35] wird diese Faustformel zur Kontrolle auf eine Reihe bakterieller Genome angewandt.

Als Vergleich dient die A'-Untereinheit des bei 37°C wachsenden *M.mazei* wo die Verteilung der o.g. Aminosäure-Gruppen weniger extrem ist (Mma_rpoA1: E+K+R=20,5%, N+Q+S+T=18,6%; Pfu_rpoA1: E+K+R=23,6%, N+Q+S+T=14,5%); für die nicht so stark konservierte L-Untereinheit ergeben sich stärker divergierende Werte (Mma_rpoL: 18,5%, 16,2%, Pfu_rpoL: 34,5%, 11,8%).

Dieser Trend gilt auch für die anderen Untereinheiten der RNA-Polymerase, wie die folgende Abbildung zeigt.

Thermoadaptation: Vergleich der RNA-Pol-Untereinheiten von *P.furiosus* und *M.mazei*

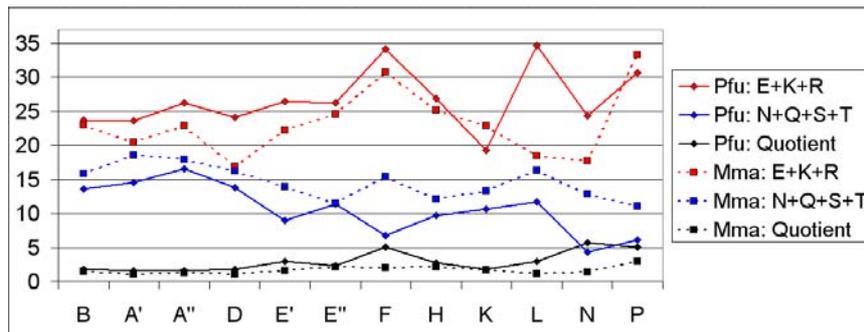


Abb.4-30: Indizien für Thermoadaptation. Vergleich von *P.furiosus* und *M.mazei*

Auswertung von Indizien für mögliche Thermoadaptation. Durchgezogene Linien: *P.furiosus*. Gestrichelte Linie: *M.mazei*. Der Quotient wurde aus den Prozentwerten von E+K+R und N+Q+S+T gebildet. *M.mazei* besitzt im Ggs. zu *P.furiosus* eine in B' und B'' geteilte B-Untereinheit – die Werte wurden für diesen Vergleich addiert.

In den meisten Fällen ist bei der hyperthermophilen Art *P.furiosus* die Summe aus E+K+R größer und die Summe aus N+Q+S+T kleiner als bei der mesophilen Art *M.mazei*. Der Quotient scheint in dieser Abbildung ziemlich ausgewogen – dies täuscht aber: Die Werte von ca. 5 bei *Pfu_rpoF* und *Pfu_rpoN* deuten auf eine stark verschobene Aminosäure-Verteilung hin (auf Genom-Ebene beim Vergleich der Durchschnittswerte der Protein-Gene unterschiedlicher Arten (siehe [Abb.4-36]) bewegen sich die Quotienten Werte in einem Bereich von ca. 0.7 bis 1,7.

Bevor darauf eingegangen wird, was für eine Bedeutung diese Daten für die praktische Arbeit mit den RNA-Pol-Untereinheiten und Transkriptionsfaktoren haben könnte, wird zunächst noch die Aminosäure-Zusammensetzung der *Pyrococcus*-Proteine etwas genauer untersucht:

Gibt es bei den RNA-Pol-Untereinheiten Auffälligkeiten, die sich von dem für *P.furiosus* üblichen Aminosäure-Verteilungsmuster abheben?

Zur weiteren Auswertung wurde eine „Normalisierung“ auf die für *P.furiosus* typische Aminosäure-Zusammensetzung vorgenommen. Zweck dieser Vorgehensweise ist es, die Besonderheiten einzelner Proteine durch Ausblenden des für *P.furiosus* typischen Aminosäure-Verteilungsmusters besser erfassbar zu machen (auch abgesehen von „Hyperthermophilie“ haben die verschiedenen Aminosäuren einen unterschiedlichen „Erwartungswert“ – Cystein oder Tryptophan sind z.B. generell ziemlich selten). Hierbei wurde die durchschnittliche, prozentuale Aminosäure-Frequenz in 2065 Proteingenen als Ausgangsbasis verwendet. Die Umrechnung erfolgte auf einheitliche 5% für jede der 20 proteinogenen Aminosäuren ($20 * 5\% = 100\%$). Ein Protein, das genau dem „Konsensus“ entspricht enthält also 5% von jeder Aminosäure; dies wurde durch die rote Linie in den Grafiken angedeutet. Signifikant sind die deutlichen Abweichungen von diesem Standardwert. Obwohl dieses Verfahren praktikabel – und meiner Meinung nach aussagekräftig – ist, sollte bedacht werden, dass die prozentuale Aminosäure-Zusammensetzung bei sehr kleinen Proteinen

häufig etwas „extremer“ (und mehr von „Zufällen“ abhängig) ist, während bei großen Proteinen mit einer langen Aminosäure-Sequenz eine gewisse statistische Nivelierung eintritt.

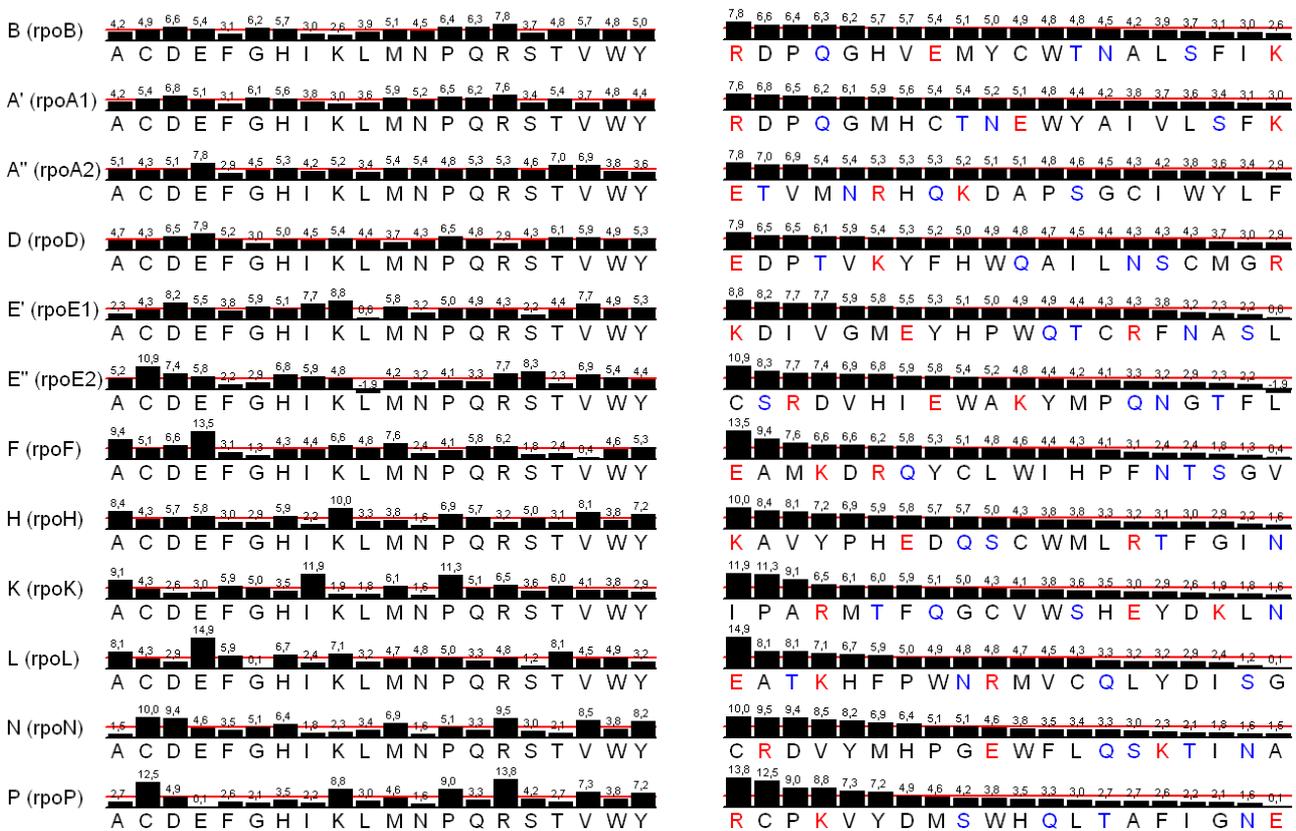


Abb.4-31: Prozentuale Aminosäurefrequenz nach "Normalisierung"

Selbst nach Abzug des für *Pyrococcus* typischen Durchschnitts-Verteilungsmusters ist in 9 von 12 Fällen die häufigste Aminosäure Glutamat, Lysin oder Arginin. Allerdings ist durch die Normalisierung auch ein Teil der auf Thermoadaptation hindeutenden Signatur herausgerechnet worden, was sich an einer stärkeren „Durchmischung“ der beiden Gruppen E+K+R (größere, stärker hydrophobe und geladene Aminosäuren) und N+Q+S+T (ungeladene, polare Aminosäuren) zeigt. Die rote Linie bei 5% markiert den Erwartungswert der jeweiligen Aminosäure in einem Standard-Protein in *Pyrococcus*.

Folgende Korrekturen wurden vorgenommen:

A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
6,4	0,7	4,2	9,0	4,4	7,0	1,5	8,9	8,4	10,2	2,4	3,4	4,2	1,7	5,5	4,9	4,3	7,9	1,2	3,9
-1,4	+4,3	+0,8	-4,0	+0,6	-2,0	+3,5	-3,9	-3,4	-5,2	+2,6	+1,6	+0,8	+3,3	-0,5	+0,1	+0,7	-2,9	+3,8	+1,1

(2. Zeile: Durchschnittliche prozentuale Aminosäure-Verteilung in 2065 *P.furiosus*-Proteingenen. 3.Zeile Normalisierung auf jeweils 5%)

Es sind recht starke Abweichungen vom *P.furiosus*-Standard-Protein festzustellen, was in dem folgenden Diagramm bezüglich der diskutierten Thermoadaptation ausgewertet wird:

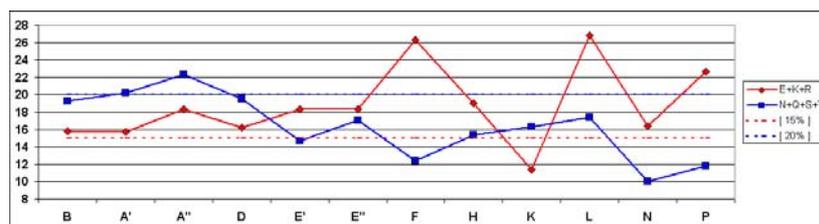


Abb.4-32: Häufigkeit von E+K+R gegenüber N+Q+S+T in den RNA-Pol-Genen nach Umrechnung

Nach dem „Herausrechnen“ der Standard-Verteilung beträgt der Erwartungswert für E+K+R 15% und für N+Q+S+T 20%, was durch die gestrichelten Linien in der selben Fabe angedeutet wird. Abgesehen von B, A', D, und H zeigen alle übrigen Untereinheiten z.T. drastische Abweichungen vom Standard. Neben der überdurchschnittlichen Thermoadaptation spielen hier aber auch RNA-Polymerase-spezifische Besonderheiten eine Rolle, wie man an dem teilweise ähnlichen Kurvenverlauf bei einem Vergleich des hyperthermophilen *P.furiosus* mit der mesophilen Art *M.mazei* erahnen kann (siehe [Abb. 4-30]).

Gibt es eine Korrelation von Thermoadaptation und Löslichkeit?

Pyrococcus ist ein hyperthermophiler Organismus. Es muß also davon ausgegangen werden, dass eine Thermoadaptation der Proteine stattgefunden hat – ansonsten würde bei 100°C eine sofortige Denaturierung der Proteine eintreten. Auf einen der Anpassungsmechanismen – nämlich die Änderung der Häufigkeit bestimmter Aminosäuren – ist bereits eingegangen worden. Zusätzlich können andere Faktoren beteiligt sein, z.B. Modifikationen der 3D-Struktur und der Position einzelner Aminosäuren zueinander (etwa, um in bestimmten Bereichen die ionischen Wechselwirkungen innerhalb des Proteins zu verstärken, und durch die hieraus resultierende Kompaktierung dem Thermostress entgegenzuwirken). Auch im Cytoplasma des Organismus gelösten Substanzen können zu einer allgemeinen Steigerung der Temperaturreistenz beitragen. Chaperone, also Hilfsproteine, die andere Proteine unter Stressbedingungen stabilisieren können, sind eine weitere Möglichkeit hierzu (Sternier & Liebl, 2001).

Weit jenseits von dem, was man normalerweise (z.B. in der Zoologie) unter „stenotherm“ oder „oligotherm“ versteht, leben hyperthermophile, extrem angepasste Mikroorganismen z.B. in kochend heißem Wasser in vulkanischen Sedimenten und „black smokers“ in der Tiefsee.

Erkauft wird diese erstaunliche „Einnischung“ durch den Verlust der Fähigkeit, in kälteren Lebensräumen zu überleben – Hyperthermophile können unter ca. 60°C nicht mehr wachsen.

Die besondere Struktur der Proteine aus Hyperthermophilen – also auch *P.furiosus* – hat verschiedene Vor- und Nachteile für die Laborpraxis:

Übliche Praxis ist es, ein in einem Organismus identifiziertes und evtl. sogar direkt aus ihm gereinigtes Protein für weitergehende Studien rekombinant in großen Mengen und evtl. versehen mit bestimmten Fusionsanteilen (z.B. His₆-Tag) in einem *E.coli*-Expressionsstamm herzustellen.

Bei vielen „hyperthermophilen“ Proteinen kann hierbei eine verblüffend einfache Reinigungsmethode angewandt werden: Die Hitzepräzipitation oder Hitzefällung. Während die meisten *E.coli*-Proteine schon längst ausgefallen sind, bleiben die thermoadaptierten Protein in Lösung. Gerade bei Proteinen oder Enzymen, die auch in der Zelle allein oder als Homomultimer vorkommen, kann in den meisten Fällen von stark erhöhter Thermostabilität ausgegangen werden (z.B. bei der Reinigung der Pyruvat-Kinase aus *A.ferroxidans* (Johnsen *et al.*, 2003)). Ein oder zwei Hitzeschritte z.B. in Kombination mit einer Nickel-Säule liefern oft schon ein hochreines Protein. Günstig ist es natürlich, wenn die trotz dieser Prozedur erhaltene korrekte Faltung des Enzyms gleich funktionell durch meßbare Enzymaktivität nachgewiesen werden kann.

Leider ist die Lage bei der hier untersuchten RNA-Polymerase aus *P.furiosus* in bezug auf die Reinigung nicht ganz so einfach:

Hitzestabilität bei Enzymkomplexen wie der RNA-Polymerase nur als Ganzes?

Ebenso wie im Rahmen der Thermoadaptation bestimmte Aminosäure-Positionen wegen ihrer Funktion bei der Protein-Protein-Interaktion in einem Enzymkomplex nicht verändert werden können (und deshalb evtl. die Statistik bei der Suche nach Thermoadaptationsmustern etwas „verfälschen“ (Haney *et al.*, 1999)), scheint sich die praktisch meßbare Thermostabilität der RNA-Polymerase bei *P.furiosus* hauptsächlich auf den komplett zusammengefügt Komplex aus den 12 Untereinheiten zu beziehen, was sich z.B. eindrucksvoll in einem (zellfreien) *In vitro* Transkriptions-Ansatz bei 70°C (oder sogar maximal 95°C) bei voll erhaltener Enzymaktivität zeigen läßt (Hethke *et al.*, 1996; Hethke *et al.*, 1999).

Bei der Reinigung der einzelnen Untereinheiten, die rekombinant in *E.coli* überexprimiert wurden zeigte sich, dass bei vielen Proteinen keine Hitzefällung angewandt werden konnte, und dass die Löslichkeit vielfach miserabel ist.

Die folgende Tabelle faßt die praktischen Erfahrungen bei der Reinigung der rekombinanten Untereinheiten der RNA-Polymerase unter den Aspekten Hitzestabilität und Löslichkeit zusammen:

<p>B (rpoB): Extreme Bildung von Einschlusskörpern bei der Expression; kaum (kein) Protein im Überstand; sehr schlechte Löslichkeit; denaturierende Reinigung aus Einschlusskörpern nötig (persönliche Mitteilung S.Naji). Diese Untereinheit wird gerade im Rahmen der Weiterführung dieses Projektes von S.Naji kloniert / gereinigt.</p>
<p>A' (rpoA1): Extreme Bildung von Einschlusskörpern bei sehr großen Proteinmengen; denaturierende Reinigung aus Einschlusskörpern nötig; schlechte Löslichkeit (s. rpoB). Hitzefällung des S100 sinnlos, weil kaum Protein.</p>
<p>A'' (rpoA2): schlechte Löslichkeit; allerdings erhebliche Mengen Protein im S100 (Überstand). An Hitzefällung ist trotzdem nicht zu denken! Aufgrund der schlechten Löslichkeit ist die Proteinkonzentration nach Ausschlußchromatographie mit der Superdex 75 sehr schwach konzentriert und läßt sich (z.B. mit Ultrafiltration) nur mit Mühe auf praktikable Konzentrationen bringen.</p>
<p>D (rpoD): Große Proteinmengen im Überstand; normale Reinigung über mehrere Säulen; Hitzefällung nicht möglich. Problemloses Handling.</p>
<p>E' (rpoE1): Etwas geringere Proteinmengen als bei rpoD. Trotzdem leidlich gute Löslichkeit der Version mit N-terminalem Fusionsanteil (die C-terminale Version liefert aufgrund stark herabgesetzter Löslichkeit sehr geringe Proteinmengen).</p>
<p>E'' (rpoE2): Hier wurde nur eine Testexpression vorgenommen, u.a. auch gemeinsam mit dem direkt stromaufwärts liegenden E'-Gen (es sollte primär gezeigt werden, dass es sich nicht um einen Sequenzierfehler mit scheinbarer Verschiebung der Leserasters handelt). Daten zur Hitzestabilität liegen nicht vor. Zum Zeitpunkt der Expression war die Bedeutung von E'', sein hochkonserviertes Zinkfinger-Motiv und die Möglichkeit der „translationalen Kopplung“ bei partiell überlappenden Genen noch nicht herausgearbeitet worden</p>
<p>F (rpoF): Im Vergleich zu den anderen RNA-Pol-Untereinheiten größte Protein-Ausbeute bzw. höchste Protein-Konzentration des gereinigten Proteins wegen seiner guten Löslichkeit. Eine Hitzefällung wurde bei der Reinigung nicht vorgenommen (und ist für ein gutes Resultat auch nicht nötig); die Hitzestabilität sollte aber noch einmal überprüft werden.</p>
<p>H (rpoH): Große Protein-Mengen; hitzestabil in Phosphat-Puffer 10min / 90°C!</p>
<p>K (rpoK): Ist nicht hitzestabil und nicht besonders gut löslich. Trotzdem gut zu reinigen, weil neben Einschlusskörpern genügend Protein im S100 gelöst ist. Das Protein fällt auf der Ni-Säule teilweise aus, und eluiert hinter dem normalen Peak in hochreiner Form.</p>
<p>L (rpoL): Ist ähnlich hitzestabil wie rpoH. Große Protein-Menge. Etwas besser löslich als rpoN. Auf der Molekularsieb-Chromatographie-Säule (Superdex 75) fiel ein Teil des Proteins aus und verursachte bei späteren Bindexperimenten eine störende Kontamination, bis erkannt war, dass durch rpoD der besser lösliche DL-Komplex gebildet wurde, und rpoL auf diese Weise doch noch von der Säule eluierte.</p>
<p>N (rpoN): Zeigt eine ähnlich hohe Hitzestabilität wie rpoH und rpoL. Die Löslichkeit ist jedoch schlecht - rpoN neigt zur Bildung von größeren Aggregaten.</p>
<p>P (rpoP): Ist hitzestabil. Bei der Version mit His₆-Tag große Protein-Menge. Es gibt keine Probleme mit der Löslichkeit. Wurde von O.v.Kampen im Rahmen seiner Diplomarbeit kloniert / gereinigt (von Kampen, 2003)</p>

Abb.4-33: Hitzestabilität und Löslichkeit der RNAP-Untereinheiten von *P.furiosus* in der Praxis

Da alle Untereinheiten eine Verschiebung der Aminosäure-Häufigkeit im Sinne der Thermoadaptation aufweisen, ist es schwer, eine Korrelation zwischen den beiden gegensätzlichen Aminosäure-Gruppen und Hitzestabilität und / oder Löslichkeit herzustellen. An erster Stelle bei der Suche nach schlechter Löslichkeit scheint zunächst die Größe der Proteine zu stehen (rpoB, rpoA1 und rpoA2). rpoF profitiert offensichtlich von der extremen Verteilung im Sinne der Thermoadaptation. Bei rpoK bewirkt der verringerte Prozentsatz an E+K+R anscheinend den Verlust der Thermostabilität, wie er ansonsten bei den ähnlich kleinen Proteinen rpoH, rpoL und rpoN festzustellen ist. Vergleicht man das Muster von rpoD und rpoE1 sowie rpoF, so kann man sagen, daß rpoE1 eher in die Richtung des extrem gut löslichen rpoF tendiert. Wie oben aufgelistet, ist rpoE1 aber etwas schlechter als rpoD löslich. Wie man sieht, dürfen keine voreiligen Schlüsse gezogen werden. Z.B. muß höchstwahrscheinlich auch die Form der RNA-Pol-Untereinheiten und die Größe der Kontaktflächen zu anderen Untereinheiten mit in die Überlegungen einbezogen werden.

Die im nächsten Abschnitt vorgestellten, etwas breiter gefächerten Daten sollten eigentlich nur sicherstellen, dass die sehr handliche Regel „mehr E+K+R und weniger N+Q+S+T = Indiz für Thermoadaptation“ (Haney *et al.*, 1999) auch über die Gattungsgrenzen (von *Methanooccus* bzw. *Methanocaldococcus*) hinaus und global, ohne langwierige Selektion bestimmter Protein-Familien anwendbar ist. Dass dies bei den hier gezeigten Daten im Großen und Ganzen tatsächlich der Fall ist, liegt vermutlich an der großen Zahl der zu Grunde liegenden Protein-Gene, die bewirkt, dass Proteine, die aus irgendeinem funktionalen Kontext heraus nicht über Aminosäure-Austausch adaptiert werden können, statistisch nicht gravierend in Erscheinung treten.

Am Anschluß daran komme ich in einem kurzen Fazit auf die Thermoadaptation bei den RNA-Pol-Untereinheiten aus *P.furiosus* zurück.

Aminosäure-Häufigkeit in den Protein-Genen verschiedener Mikroorganismen:

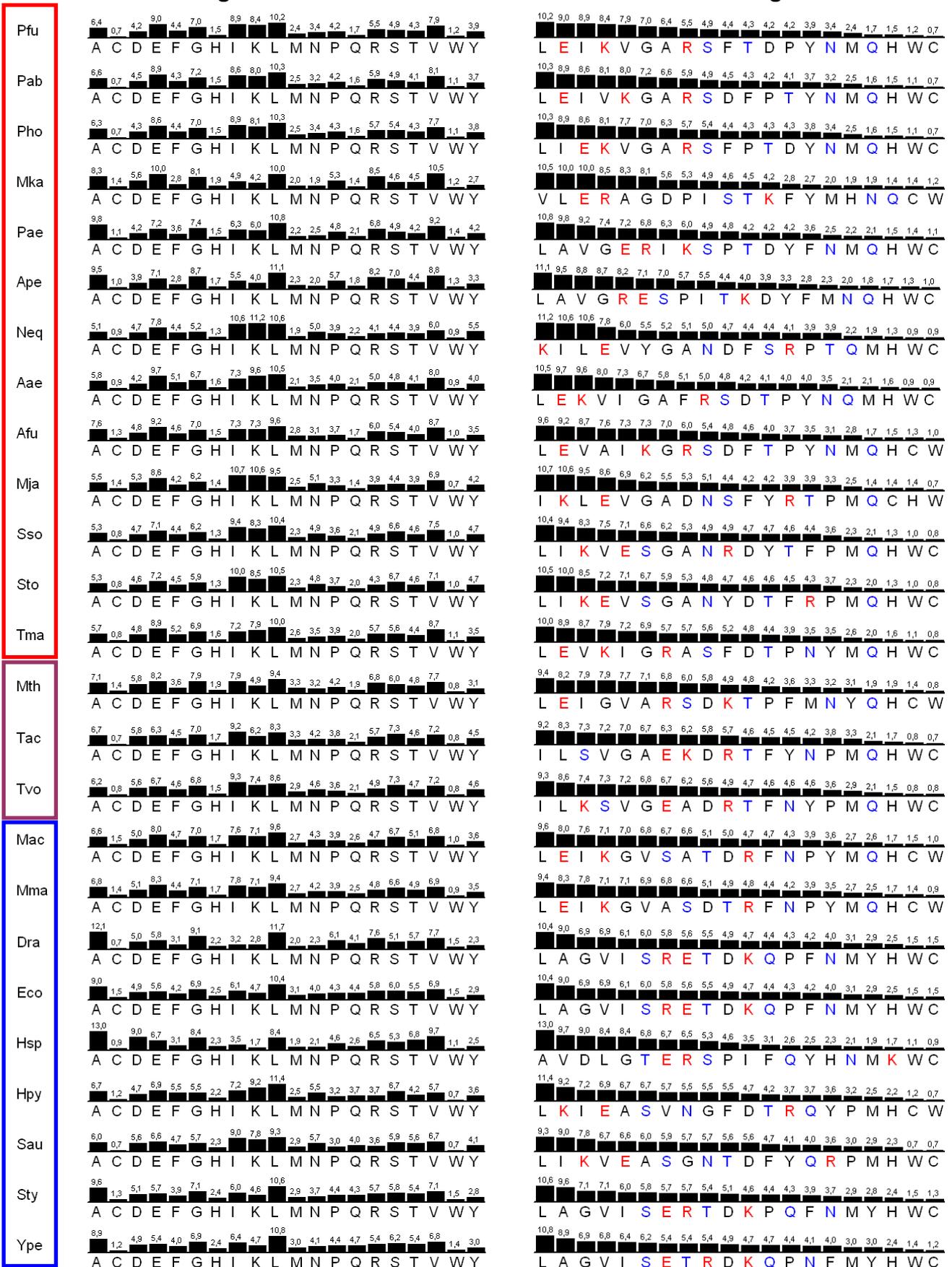


Abb.4-35: Übersicht: Aminosäurefrequenz vor dem Hintergrund potentieller Thermoadaptation
 Prozentuale Aminosäure-Frequenz in den Protein-Genen mesophiler (blau), thermophiler (violett) und extrem thermophiler bzw. hyperthermophiler Organismen (rot) (s.Anhang [8.5]). Vermehrtes Auftreten von Glutamat, Arginin, Lysin (E+R+K, größere, stärker hydrophobe und geladene Aminosäuren) sowie Verminderung der Anzahl von Serin, Threonin, Asparagin und Glutamin (N+Q+S+T, ungeladen, polar) als Indikator für Thermoadaptation (Haney *et al.*, 1999).

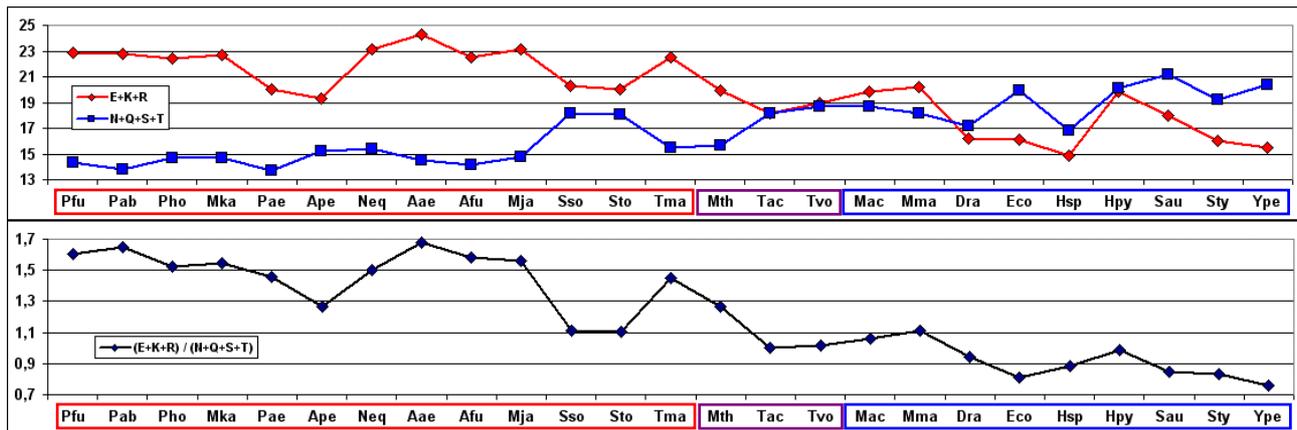


Abb.4-36: E+K+R gegenüber N+Q+S+T in hyperthermophilen, thermophilen und mesophilen Arten

Aae = *Aquifex aeolicus*, Ape = *Aeropyrum pernix*, Afu = *Archaeoglobus fulgidus*, Dra = *Deinococcus radiodurans*, Eco = *Escherichia coli*, Hsp = *Halobacterium sp. NRC-1*, Hpy = *Helicobacter pylori*, Mac = *Methanosarcina acetivorans C2A*, Mja = *Methanocaldococcus jannaschii*, Mka = *Methanopyrus kandleri AV19*, Mma = *Methanosarcina mazei Goe1*, Mth = *Methanothermobacter thermoautotrophicus*, Neq = *Nanoarchaeum equitans Kin4-M.*, Pab = *Pyrococcus abyssi*, Pae = *Pyrobaculum aerophilum*, Pfu = *Pyrococcus furiosus DSM 3638*, Pho = *Pyrococcus horikoshii*, Sau = *Staphylococcus aureus*, Sso = *Sulfolobus solfataricus*, Sto = *Sulfolobus tokodaii*, Sty = *Salmonella typhimurium LT2*, Tac = *Thermoplasma acidophilum*, Tma = *Thermotoga maritima*, Tvo = *Thermoplasma volcanium*, Ype = *Yersinia pestis KIM*.

Auswertung der Daten aus [Abb.4-35]. Oberes Diagramm: Rote Linie = Summe der Prozentwerte der Häufigkeit von E+K+R. Blaue Linie = Summe der Prozentwerte der Häufigkeit von N+Q+S+T. Unteres Diagramm: Quotient aus E+K+R und N+Q+S+T (zeigt den Trend noch etwas deutlicher). Ein regelmäßigerer Kurvenverlauf könnte vermutlich erreicht werden, wenn die Organismen genau nach ihrer optimalen Wachstumstemperatur geordnet würden – hier erfolgte nur eine grobe Zuordnung, die aber völlig ausreicht, um den vermuteten Trend erkennen zu lassen (siehe Legende zu [Abb.4-35]).

Die zusätzlichen mesophilen *Bacteria* wurden zufällig bzw. nach allgemeinem Interesse ausgewählt.

Yersinia sind psychrophil, d.h. sie vermehren sich auch bei 1 - 4° C. Deshalb wurde *Yersinia pestis* KIM als Kontroll-Organismus für das untere Ende der Temperaturskala hinzugefügt. Tatsächlich liefert der o.g. Quotient hier ein eindeutiges Ergebnis (d.h. dass es keinerlei Hinweis auf Thermoadaptation der Proteine dieses Krankheitserregers gibt). Etwas kurios ist, dass (zumindest in Bezug auf das vorgeschlagene „E+K+R / N+Q+S+T-Kriterium“) mit *Aquifex aeolicus* ein Vertreter der (von der Artenzahl nicht gerade häufigen) hyperthermophilen *Bacteria* ganz oben in der Liste steht, obwohl seine Wachstumstemperatur mit 85°C um einiges unter der von *P.furiosus* (100°C) liegt. Möglicherweise spielen hier noch andere Faktoren, wie z.B. der GC-Gehalt und entsprechende Ausstattung mit tRNAs eine Rolle.

Fazit zum Thema Thermoadaptation

Von dem diskutierten Aminosäure-Verteilungsmuster her zeigen alle Untereinheiten der RNA-Polymerase aus *P.furiosus* eine extreme Anpassung an erhöhte Temperaturen. Allerdings ist dies für die Präparation der rekombinant erstellten, einzelnen Untereinheiten wenig hilfreich (abgesehen von den kleinen Untereinheiten), weil diese Thermoadaptation offensichtlich vor allem die gesamte RNA-Polymerase mit ihren 12 Untereinheiten betrifft. Denkbar wäre es, dass vor allem die im Gesamtkomplex freiliegenden Außenflächen durch den besprochenen Aminosäure-Austausch stabilisiert werden, während im Inneren des RNA-Pol-Komplexes durch ähnliche (aber vom Muster her vermutlich nicht identische) Austauschprozesse die Protein-Protein-Bindungen intensiviert werden. Ohne die Anwesenheit ihrer Partner-Proteine sind die einzelnen RNA-Pol-Untereinheiten

zumeist nicht hitzestabil und / oder nicht gut löslich, weil normalerweise im Inneren verborgene Kontaktflächen freiliegen. Schon Zweierkomplexe sind oft schon deutlich besser löslich (z.B. bei rpoD + rpoL). Das lästige Verhalten von rpoN (Bildung von extrem schwer zu resuspendierenden Aggregaten selbst bei ganz kurzem Anzentrifugieren einer trüben (aber homogenen), recht hoch konzentrierten Lösung) läßt sich damit erklären, dass in den gebildeten Clustern die Kontaktflächen des ohnehin recht schlecht löslichen Proteins dem umgebenden Flüssigkeitsvolumen entzogen werden, und ein stabiler Zustand eintritt, der vielleicht dem von rpoN bei korrekter Integration in einen vollständigen RNA-Pol-Komplex ähnelt.

Dass eine unvollständige, bzw. aus teilweise falsch gefalteten rekombinanten Untereinheiten zusammengesetzte RNA-Polymerase nicht thermostabil ist, kann aber auch nützlich sein, und in einem finalen Präparationsschritt in Form einer Hitzefällung die intakten von den defekten Komplexen trennen, wie dies für *M.jannaschii* gezeigt wurde (Werner & Weinzierl, 2002).

Betont werden muß abschließend noch einmal, dass "Thermostabilität von Proteinen" ein faszinierendes Thema ist, welches noch nicht vollständig verstanden wird und deshalb weiter intensiv erforscht wird. Die hier untersuchte Aminosäureverteilung klammert Phänomene, die die Änderung der Proteinstruktur auf der 3D-Ebene (also der tertiären und quaternären Ebene) betreffen komplett aus. Das Problem scheint zu sein, dass die entsprechenden, bei einem bestimmten Protein verstandenen Mechanismen zur Thermoadaptation nicht unbedingt auf ein anderes Protein übertragen werden können, sondern oftmals sehr "individuell" sind. So kann es zum Beispiel sein, dass ein Protein nur in bestimmten Bereichen "starrer" wird, während andere Region entsprechend funktionaler Notwendigkeiten gut beweglich bleiben usw. (Jaenicke, 2000).

4.8 Eine zweite Kopie von TFB im Genom von *Pyrococcus*

Bei *Archaea* wurden mehrere Fälle geschildert, bei denen im Genom der untersuchten Organismen nicht eine, sondern mehrere TFB-Kopien vorkommen. Dies ist zum Beispiel für halophile *Archaea* der Fall (Reeve *et al.*, 1997), wo auch mehrere TBPs gefunden wurden. Für Organismen wie *Halobacterium* NRC-1 liegt also auf jeden Fall die Vermutung nahe, dass hier die Genexpression über mehrere TBP- und TFB-Kopien reguliert wird (Baliga *et al.*, 2000), was eine Anpassung des Organismus an den Lebensraum der halophilen *Archaea* mit häufigem und raschen Wechsel von Salinität und pH-Werte darstellen könnte. Auch für *P.horikoshii* ist die Existenz einer zweiten, von dem Standard-TFB stark abweichende Kopie schon längere Zeit bekannt (Soppa, 1999). Dadurch, dass für die Gattung *Pyrococcus* gegenwärtig sogar drei komplett sequenzierte Genome zugänglich sind (Lecompte *et al.*, 2001), ergibt sich auch für die TFB-Gene die Möglichkeit zu vergleichenden Analysen. Für die Betrachtung der RNA-Polymerase bei *Pyrococcus* ist dies deshalb relevant, weil TFB nach bisherigen Erkenntnissen die herausragende Stellung bei der Rekrutierung der RNAP an den Präinitiationskomplex einnimmt, und die Protein-Protein-Interaktionen zwischen TFB und den RNAP-Untereinheiten letztlich eine entscheidende Rolle für die Regulation der Transkription spielen.

Es zeigt sich, dass neben *P.horikoshii* auch die in dieser Arbeit näher untersuchte *Pyrococcus*-Art *P.furiosus* über ein zweites TFB-Gen verfügt, wie aus folgendem Alignment hervorgeht:

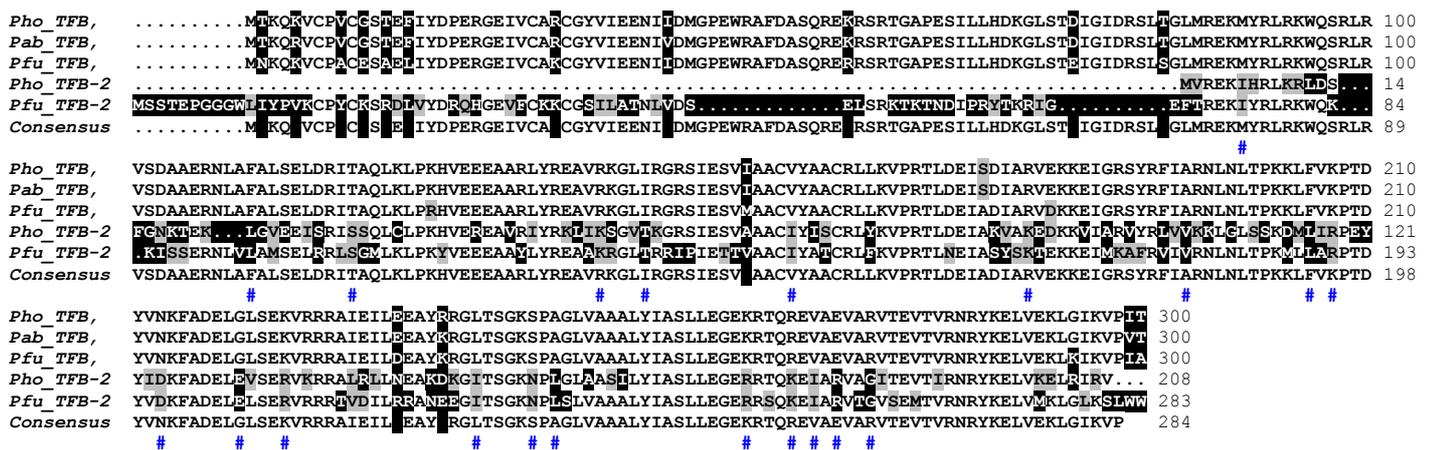


Abb.4-37: Alignment der TFB-Kopien aus *P.furiosus*, *P.horikoshii* und *P.abysii*

Das Alignment wurde mit ClustalX generiert. Der Konsensus und die Grauschattierungen wurden mit Bioedit hinzugefügt. Pho: *P.horikoshii*, Pab: *P.abysii*, Pfu: *P.furiosus*. TFB: In allen drei Arten sehr ähnliche "Standard-Kopie" von TFB; TFB-2: Zweite Kopie von TFB im Genom von *P.horikoshii* und *P.furiosus*. Namen der jeweiligen TFB-Gene im Genom (Locus): *Pab_TFB*: PAB1912; *Pho_TFB*: PH1482; *Pho_TFB-2*: PH0864; *Pfu_TFB*: PF1377; *Pfu_TFB-2*: PF0687. Blaue Raute: Positionen, an denen die beiden TFB-2-Kopien aus *P.furiosus* und *P.horikoshii* einen identischen Aminosäureaustausch gegenüber den anderen drei TFB-Genen zeigen.

TFB besteht grundsätzlich aus drei deutlich zu unterscheidenden Abschnitten: Einer N-terminalen Domäne mit einem Zinkfinger-Motiv (Zhu *et al.*, 1996) und zwei C-terminalen direkten Wiederholungen ("direct repeats"). Dies gilt sowohl für *Archaea*, als auch *Eukarya* (bei *Bacteria* gibt es kein TFB).

TFB ähnelt sich in den drei *Pyrococcus*-Arten sehr stark: Zum Beispiel beträgt die Länge einheitlich 300 Aminosäuren, und die theoretische Nettoladung in allen Fällen +13 (bei einem Umgebungs-pH von 7,6). Die Nettoladung von TFB-2 aus *P.horikoshii* beträgt dagegen +21, die von TFB-2 aus *P.furiosus* +23. Auch wenn *Pho_TFB-2* (208aa) gegenüber *Pfu_TFB-2* (283aa) stark verkürzt ist (der N-Terminus mit dem Zinkfinger-Motiv fehlt fast komplett), gibt es eine Reihe Indizien, dass es zu der Genduplikatur vor der Abspaltung beider Arten kam, und dass das TFB-2-Gen zumindest eine Zeit lang eine wichtige Rolle im Zellstoffwechsel gespielt hat, so dass sich Anzeichen für die Einwirkung positiven (konservierenden) Selektionsdruckes auf die Aminosäuresequenz feststellen läßt: An insgesamt 21 Aminosäurepositionen ist ein identischer Austausch gegenüber den drei übrigen (Standard-)TFB-Genen zu beobachten (im Alignment mit

einer blauen Raute markiert). Dies ist ein starkes Argument für das Auftreten der Genduplikatur bereits zu einem frühen Zeitpunkt in dem gemeinsamen Vorläufer. Bei 14 dieser Positionen handelt es sich zudem um einen "ähnlichen" Austausch im engeren Sinne, was ein wichtiger Hinweis auf eine aktive Rolle der untersuchten Gene im Zellstoffwechsel beziehungsweise im Rahmen der Transkription darstellt (interessanter Weise sind auch viele der zwischen *Pho*_TFB-2 und *Pfu*_TFB-2 ausgetauschten Positionen "ähnlich").

Folgende Grafik verdeutlicht die erwähnten Ladungsunterschiede zwischen den TFB-Kopien:

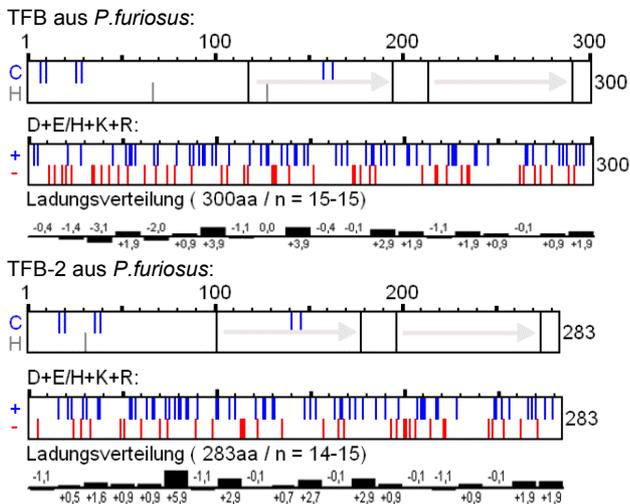


Abb.4-38: Vergleichende Darstellung der Struktur und Ladungsverteilung in TFB und TFB-2

1.Zeile: Kombinierte Darstellung von bekannten Funktionsabschnitten und Positionen von Cystein / Histidin-Resten zur Lokalisierung des Zinkfinger-Motivs (hier vom C4-Typ: CXXCX₁₅₋₁₇CXXC, wobei "X" für eine beliebige Aminosäure steht). Die beiden grau schattierten Pfeile deuten die beiden C-terminalen "direct repeats" an. Die für TFB bekannten Bereichsgrenzen wurden auf TFB-2 übertragen. 2.Zeile: Blau – basische Aminosäuren, rot – saure Aminosäuren. 3.Zeile: Darstellung der Ladungsverteilung, wobei n für die Anzahl Aminosäuren steht, die für die jeweilige Berechnung herangezogen wurden.

Obwohl die Unterschiede im Ladungsmuster auf den ersten Blick nicht sehr ausgeprägt zu sein scheinen, wird die bei TFB-2 gegenüber TFB stark erhöhte positive Ladung hauptsächlich durch die veränderte N-Terminale Domäne verursacht, wenn man die Ladung der N-terminalen und der C-terminalen Hälfte der Proteine errechnet: TFB [+3 / +10] gegenüber TFB-2 [+12 / +11].

Zusammenfassend läßt sich sagen, dass die beiden TFB-2-Gene (bzw. die resultierenden Proteine) der beiden *Pyrococcus*-Arten in Ladung und Sequenz recht ähnlich sind, und dass sie wahrscheinlich aktiv sind. Im Gegensatz zu *Pho*_TFB-2, wo die N-terminale Domäne weitgehend deletiert ist, besitzt *Pfu*_TFB-2 jedoch fast die gleiche Größe wie das Standard-TFB. Zwar täuscht der bloße Vergleich der Länge der Aminosäuresequenz etwas, weil *Pfu*_TFB-2 innerhalb der N-terminalen Domäne zwei größere Deletionen aufweist, die teilweise durch eine zusätzliche, 10 Aminosäuren lange und nicht zu den übrigen TFB-Genen homologe Region kompensiert wird. Wie aus der Grafik zu entnehmen ist, bleibt das Zinkfinger-Motiv jedoch unversehrt, so dass hier die gegenüber TFB wesentlich veränderte N-terminale Domäne weiterhin für potentielle Protein-DNA- oder Protein-Protein-Wechselwirkungen zur Verfügung steht. Da aber selbst TFB_C (also ein TFB, das nur aus den beiden C-terminalen Repeats besteht) in der Lage ist, an die DNA in der Nähe der TATA-Box zu binden (Qureshi & Jackson, 1998), scheint es fast zwingend, dass dieses Protein eine Funktion im Rahmen der Transkription hat, wie zum Beispiel die Beteiligung an der Transkription einer speziellen Teilmenge von Genen (Soppa, 1999) oder eine allgemein "modulierende" Wirkung auf Transkription durch ein Konkurrieren um gemeinsame Bindestellen mit dem Standard-TFB.

Dass TFB-2 keine so prominente (sondern eher "ergänzende") Rolle wie TFB zukommt, deutet sich in dem Fehlen des Genes in *P.abysyi* an. Da schon der gemeinsame Vorläufer von *P.furiosus* und *P.horikoshii* das zweite Gen besessen hat, und *P.abysyi* einen (späteren) gemeinsamen Vorläufer mit *P.horikoshii* besessen hat (siehe Stammbaum [4-18,a]), muß das Gen im Verlauf der Evolution zu einem relativ späten Zeitpunkt inaktiviert oder deletiert worden sein (in diesem Fall ließen sich

auch mit einer "lokalen BLAST-Suche" (siehe [8.3]) keine eindeutigen Spuren der zweiten TFB-Kopie identifizieren).

TFB-2 (PF0687) aus *P.furiosus* wurde zwar im Rahmen dieser Arbeit kloniert und einer Testexpression unterzogen. Da das resultierende Protein mit ca. 26 kDa jedoch zu stark von dem theoretischen Molekulargewicht von 33 kDa abwich, wurden praktische Versuche mit diesem Protein zugunsten anderer Experimente zurückgestellt.

Auf TFB mit seinen drei Domänen wird noch einmal unter 4.11.1.2 ("Far Western Blotting") näher eingegangen.

4.9 Klonierung der Untereinheiten der DNA-abhängigen RNA-Polymerase

Zu Beginn dieser Arbeit lag das Genom von *Pyrococcus furiosus* noch in einer recht fehlerbehafteten, unannotierten Rohfassung vor, was jedoch trotzdem einen großen Fortschritt für die Erforschung von *P. furiosus* darstellte. Dass das Genom der Schwesterart *P. horikoshii* zu diesem Zeitpunkt schon recht gut aufbereitet war und viele Informationen über eine (inzwischen stark abgeänderte) Webseite im Internet bereitgestellt wurden (<http://www.bio.nite.go.jp/ngac/e-home/ot3-e.html>), erleichterte die Identifizierung einer Reihe von RNAP-Untereinheiten im Genom, die anschließend gezielt in geeignete Expressionsvektoren kloniert werden konnten.

Auch aus der Literatur waren schon Daten über die archaeele RNAP-Gene vorhanden, z.B. für *Sulfolobus acidocaldarius* (Lanzendörfer *et al.*, 1994).

Als Expressionssystem wurde die schon in der Arbeitsgruppe etablierte Kombination von so genannten pET-Vektoren (Novagen, Inc.) in Verbindung mit *E.coli*-BL21(DE3)-Expressionsstämmen gewählt (siehe [3.1]). Die Transformation des Ligationsansatzes (bzw. den erfolgreich ligierten Vektoren inklusive der inserierten RNAP-Gene) erfolgte in der Regel in *E.coli* JM109, einem Sicherheitsstamm, aus dem die Plasmide in hoher Kopienzahl für Sequenzierung und anschließende Transformation in BL21(DE3) präpariert wurden.

Zum Einsatz kamen neben pET-14b und pET-15b auch das Plasmid pET-17b (<http://www.merckbiosciences.de/g.asp?f=NVG/pETtable.html>). Während die beiden erst genannten Vektoren ein rekombinantes Protein mit zusätzlichem Fusionsanteil liefern, enthält das bei pET-17b resultierende, überexprimierte Protein nur die eigentliche Gensequenz. Letzteres ist von Vorteil, wenn das Protein relativ leicht gereinigt werden kann, bzw. ein individuelles Reinigungsschema erarbeitet werden soll. Da das rekombinante Protein von der Aminosäuresequenz her (und somit auch mehr oder weniger in der 3D-Struktur) identisch mit dem entsprechenden Protein aus dem Original-Organismus (also in diesem Fall *P. furiosus*) ist, entfallen mögliche Interferenzen mit dem zusätzlichen Fusionsanteil.

Zu einem späteren Zeitpunkt im Verlauf des RNA-Polymerase-Projektes wurde zusätzlich das Plasmid pET-33b(+) verwendet, das ähnlich wie pET-14b und pET-15b eine N-terminale Zusatzsequenz mit "His₆-Tag" für die Reinigung nach einem vereinfachten Reinigungsschema per Nickel-Säule auf der Basis von Affinitätschromatographie bietet. Allerdings ist der Fusionsanteil bei pET-33b(+) nicht 20, sondern 25 Aminosäuren lang, da hier eine 5 Aminosäuren lange Erkennungssequenz für radioaktive Markierung mit Proteinkinase A enthalten ist.

Die PCR wurden vor allem bei den Genen für die größeren Untereinheiten mit DNA-Polymerasen durchgeführt, die über eine (im Vergleich zur Standard-*Taq*-DNA-Polymerase) erhöhte Korrekturlesefunktion aufweisen, wie z.B. *Pfu*-DNA-Polymerase. Restriktionshydrolyse, Ligation, Transformation erfolgte mit den üblichen molekularbiologischen Methoden entsprechend den Angaben des Herstellers.

4.10 Expression und Reinigung der RNA-Pol-Untereinheiten

Die Überexpression der gewünschten rekombinanten Proteine in *E.coli* BL21(DE3) erfolgte über Induktion mit dem Substrat-Analogen IPTG. Dieses verändert die Konformation eines Repressors, der daraufhin die Ablesung des Genes für die ins *E.coli*-Genom eingeschleuste T7-RNA-Polymerase aktiviert, welche anschließend die in die pET-Vektoren eingefügten Gene so stark transkribiert, dass der Zellstoffwechsel fast vollständig auf die Produktion des Fremdproteins umgestellt wird. Folgende Abbildung zeigt eine Auswahl der exprimierten und gereinigten Proteine:

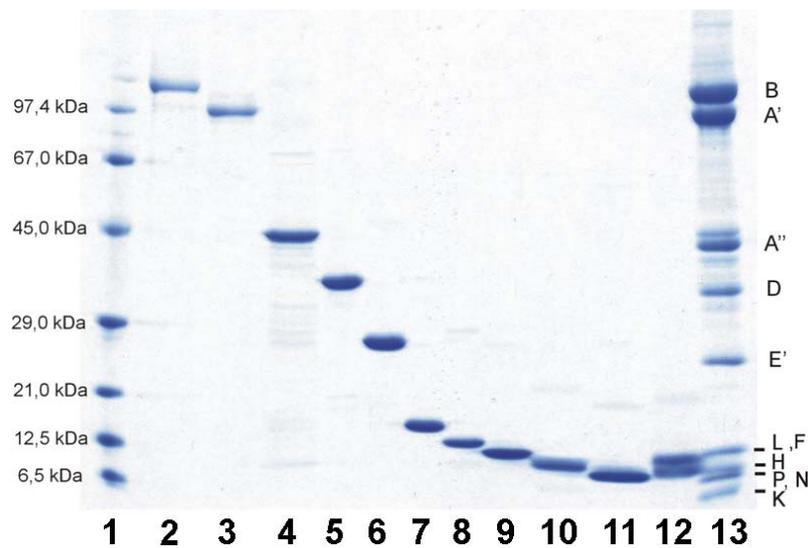


Abb.4-39: Gegenüberstellung rekombinanter RNAP-Untereinheiten und der nativen RNAP

Denaturierendes Gradientengel (8 – 15% PA, 0,1% SDS), Coomassie-Färbung.

1: Standard, 2 – 3: Untereinheiten B und A' (aus einer nativen RNAP-Präparation; Banden aus einem 6%igen SDS-Gel ausgeschnitten und eluiert; A' liegt jedoch auch rekombinant vor); 4: - 12: rekombinante Untereinheiten A'', D, E', F, L, H, N, K, P. N und P wurden TCA-gefällt, um eine bessere Darstellung zu erreichen. 13: nativ aus *P.furiosus* gereinigte RNAP. Die Untereinheiten wurden nach ihrem theoretischen Molekulargewicht in absteigender Reihenfolge sortiert.

Untereinheit P (Spur 12) wurde mir freundlicher Weise zur Verfügung gestellt (von Kampen, 2003). Wie bei den anderen rekombinanten RNAP-Untereinheiten (UE) handelt es sich um die Version inklusive Fusionsanteil, welcher ein im Vergleich zum nativen Protein (Banden in Spur 13) erhöhtes Molekulargewicht und folglich ein langsames Wandern im elektrischen Feld bei der Elektrophorese bewirkt. Während besonders bei den größeren UE eine Zuordnung von rekombinantem und nativem Protein gut möglich ist, fällt dies bei den kleineren UE schwerer, was vermutlich am relativ starken Einfluß der 25 Aminosäuren langen Zusatzsequenz liegt: Während deren Anteil beispielsweise bei der D-Untereinheit mit normalerweise 261 Aminosäuren 8,7% beträgt, liegt sie bei der kleinsten Untereinheit P mit 49 Aminosäuren schon bei ca. 33% des resultierenden Fusionsprotein. Aus diesem Grund ist die optionale (teilweise) Abtrennung des His₆-Tags per Thrombin-Schnittstelle vorgesehen. Untereinheit P zeigt darüber hinaus ein zusätzlich atypisch langsames Laufverhalten in der Elektrophorese, was durch ihre positive Ladung (ca.+9) bedingt ist, welche die negative Ladung des SDS teilweise kompensiert.

Aber auch wenn man die genannten Molekulargewichtsdifferenzen einkalkuliert, bleiben für die Zuordnung von 5 kleinen Untereinheiten (bzw. 6 mit F) nur 4 gut erkennbare Banden in der nativ gereinigten RNAP-Präparation (Spur 13) übrig. Für P konnte durch direkten Vergleich des Laufverhaltens einer rekombinanten Version ohne Fusionsanteil (pET-17b, s.oben) eine Zuordnung zur zweituntersten Bande vorgenommen werden, die vermutlich auch die Untereinheit N enthält. Dass die obere Bande im Bereich von 12,5 kDa allerdings auch Untereinheit F enthält, war ein (kürzlich aus Western Blotting-Experimenten resultierender) überraschender Befund (siehe [Abb.4-57]). Zuvor mußte davon ausgegangen werden, dass F möglicherweise während eines der Schritte bei der Reinigung der nativen RNA-Polymerase aus *P.furiosus* verloren geht, und folglich in der Darstellung fehlt.

Eine deutliche Zunahme der bei der Expression resultierenden Menge an rekombinantem Protein konnte durch den Wechsel des Expressionsstammes von einem Standard-BL21(DE3)-Stamm hin zu einer Variante mit zusätzlichen tRNA-Genen für in *E.coli* normalerweise seltene Codons erreicht werden:

BL21-CodonPlus™-RIL (Stratagene, Inc.; http://www.stratagene.com/vol12_2/p49-51.htm)

Die folgende Tabelle zeigt die Häufigkeit in *E.coli* seltener Codons in Genen, die im Rahmen dieser Arbeit oder in der Arbeitsgruppe von Prof.Dr.M Thomm kloniert wurden, bzw. allgemeine Bedeutung für die Transkription in *P.furiosus* besitzen:

	AGG (Arg)	AGA (Arg)	CGA (Arg)	CTA (Leu)	ATA (Ile)	CCC (Pro)	% gesamt
rpoB (PF1564)	21 (18,8)	69 (61,8)	0 (0)	17 (15,2)	34 (30,4)	10 (9,0)	13,5
rpoA1 (PF1563)	15 (16,5)	56 (61,7)	0 (0)	11 (12,1)	28 (30,9)	8 (8,8)	13,0
rpoA2 (PF1562)	9 (22,7)	13 (32,7)	0 (0)	5 (12,6)	11 (27,7)	2 (5,0)	10,1
rpoD (PF1647)	4 (15,3)	5 (19,2)	0 (0)	4 (15,3)	7 (26,8)	2 (7,7)	8,4
rpoE1 (PF0256)	5 (26,5)	4 (21,2)	0 (0)	4 (21,2)	11 (58,2)	3 (15,9)	14,3
rpoE2 (PF0255)	1 (16,4)	3 (49,2)	0 (0)	1 (16,4)	3 (49,2)	0 (0)	13,1
rpoF (PF1036)	2 (16,7)	5 (41,7)	0 (0)	3 (25)	6 (50)	0 (0)	13,3
rpoH (PF1565)	2 (24,4)	1 (12,2)	0 (0)	1 (12,2)	1 (12,2)	0 (0)	6,1
rpoK (PF1642)	2 (35,1)	2 (35,1)	0 (0)	0 (0)	4 (70,2)	1 (17,5)	15,8
rpoL (PF0050)	2 (21,1)	3 (31,6)	0 (0)	2 (21,1)	5 (52,6)	1 (10,5)	13,7
rpoN (PF1643)	3 (42,9)	4 (57,2)	0 (0)	2 (28,6)	1 (14,3)	2 (28,6)	17,1
rpoP (PF2009)	3 (61,2)	3 (61,2)	0 (0)	2 (40,8)	1 (20,4)	1 (20,4)	20,4
TBP (PF1295)	4 (20,9)	3 (15,7)	0 (0)	2 (10,5)	4 (20,9)	0 (0)	6,8
TFB (PF1377)	10 (33,3)	22 (73,3)	0 (0)	4 (13,3)	5 (16,7)	2 (6,7)	14,3
TFB-2 (PF0687)	6 (21,2)	23 (81,3)	1 (3,5)	3 (10,6)	5 (17,7)	2 (7,1)	14,1
LrpA (PF1601)	1 (7,1)	3 (21,3)	0 (0)	2 (14,2)	7 (49,6)	2 (14,2)	10,6
TFE (PF0491)	0 (0)	14 (71,8)	1 (5,1)	2 (10,3)	4 (20,5)	0 (0)	10,1
Phr (PF1790)	7 (34,7)	16 (79,2)	0 (0)	10 (49,5)	12 (59,4)	3 (14,9)	23,8
TFS (PF0986)	5 (45,5)	4 (36,4)	0 (0)	0 (0)	3 (27,2)	2 (18,2)	12,7
NusA (PF1560)	2 (13,8)	7 (48,3)	0 (0)	2 (13,8)	9 (62,1)	0 (0)	13,8
NusG (PF1990)	2 (13,2)	5 (32,9)	0 (0)	2 (13,2)	5 (32,9)	3 (19,7)	11,2
Tip49 (PF1847)	9 (20,4)	20 (45,4)	1 (2,3)	7 (15,9)	23 (52,2)	2 (4,5)	14,1

Abb.4-40: In *E.coli* seltene Codons in für die Transkription relevanten Proteingenen in *P.furiosus*

Fettgedruckte Zahlen: Frequenz des Codons im jeweiligen Gen. Zahl in Klammern: Codon-Frequenz umgerechnet auf 1000bp (nach der Empfehlung von Stratagene kann es ab ca. 15 seltenen Codons / 1000 Codons zu Problemen bei der Expression kommen).

Die folgende Darstellung verallgemeinert die Analyse der verwendeten Codons, indem alle in der aktuellen Genomdatei für *P.furiosus* (siehe [8.5]) Proteingene untersucht werden:

Übersicht über die in 2065 Proteingenen aus <i>P.furiosus</i> verwendeten Codons (Codon Usage):						
Anzahl analysierter Gene: 2065; Codons insgesamt: 583472						
A (Ala):	GCA: 14424	GCC: 7999	GCG: 3441	GCT: 12350		
R (Arg):	CGA: 719	CGC: 652	CGG: 432	CGT: 670	AGA: 16266	AGG: 12311
N (Asn):	AAC: 9851	AAT: 10376				
D (Asp):	GAC: 9320	GAT: 16096				
C (Cys):	TGC: 1309	TGT: 2164				
Q (Gln):	CAA: 5565	CAG: 4627				
E (Glu):	GAA: 28202	GAG: 23513				
G (Gly):	GGA: 21607	GGC: 4477	GGG: 8196	GGT: 7242		
H (His):	CAC: 4701	CAT: 4090				
I (Ile):	ATA: 23707	ATC: 7932	ATT: 19018			
L (Leu):	CTA: 10362	CTC: 9918	CTG: 4809	CTT: 14126	TTA: 11396	TTG: 8147
K (Lys):	AAA: 22179	AAG: 24833				
M (Met):	ATG: 12758					
F (Phe):	TTC: 10653	TTT: 15104				
P (Pro):	CCA: 11594	CCC: 5362	CCG: 1768	CCT: 6215		
S (Ser):	TCA: 5974	TCC: 3650	TCG: 1465	TCT: 5291	AGC: 6322	AGT: 5975
T (Thr):	ACA: 9284	ACC: 4430	ACG: 3339	ACT: 8682		
W (Trp):	TGG: 7207					
Y (Tyr):	TAC: 10721	TAT: 12496				
V (Val):	GTA: 11174	GTC: 6198	GTG: 8065	GTT: 20608		
Stop:	TAA: 692	TAG: 423	TGA: 990			

Abb.4-41: Frequenz der verschiedenen Codons (Codon Usage) in *P.furiosus*

Pro Aminosäure wurde das häufigste Codon fett gedruckt. Die "seltene Codons" in *E.coli*, für die BL21-CodonPlus™-RIL zum Zweck besserer Expression zusätzliche tRNAs enthält, wurden grau hinterlegt. (Die Daten wurden mit "GenbankAnalyse" berechnet).

Gut erkennbar ist, dass das Codon "AGA" für Arginin und "ATA" für Isoleucin in allen Proteingenen von *P.furiosus* am häufigsten vorkommt, und deshalb zusätzliche tRNAs sicherlich generell eine Verbesserung der Expression von *Pyrococcus*-Proteinen in *E.coli* bewirken.

Unter [Abb.4-33] finden sich die Schilderungen von Erfahrungen bei der Expression von rekombinanten RNAP-Untereinheiten bei *Pyrococcus furiosus* unter den Aspekten "Löslichkeit" und "Hitzestabilität".

4.11 Untersuchung von Protein-Protein-Wechselwirkungen

Die wichtigste in dieser Arbeit zur Detektion von Protein-Protein-Wechselwirkungen eingesetzte Methode ist das "Far Western Blotting". Darüber hinaus wurden Bindungsstudien mit Hilfe von Polyacrylamid-Gel-Elektrophorese unter nativen Bedingungen, sowie mit Molekularsiebchromatographie durchgeführt.

4.11.1 Bindungsstudien: Far Western Blotting-Experimente

Das Far Western Blotting kann grundsätzlich in zwei unterschiedlichen Formen durchgeführt werden, die sich in der Art unterscheiden, wie die zu untersuchenden Proteine auf die Nitrocellulose-Membran gelangen: Erstens durch Auftropfen der Proteine auf die Membran, und zweitens durch Transfer der nach Größe aufgetrennten Proteine aus einem denaturierenden PA-(Gradienten-)Gel auf eine Nitrocellulose- oder PVDF-Membran (Elektroblot). Das Dot Blotting hat vom Prinzip her den Vorteil, dass die aufgebrachten Protein-Mengen recht genau eingestellt werden können. Ein Nachteil ist, dass hier die zusätzliche Information über die Proteingröße fehlt. Außerdem bleiben eventuell vorhandene Kontaminationen der Protein-Präparation unentdeckt. Ein Nachteil beim Elektroblot der Proteine aus einem Gel auf die Membran ist, dass der Transfer je nach Größe, Form, Ladung und Isoelektrischem Punkt des Proteins, sowie der Prozentigkeit des PA-Geles unterschiedlich gut erfolgt. Erschwerend kann sich auch eine allgemein schlechte Löslichkeit der Proteine auswirken.

Überdurchschnittlich gute Löslichkeit kann sich aber auch störend bei der Inkubation der Membranen in Blockierlösung bemerkbar machen, weil, falls ein oder mehrere Proteine die Tendenz zum vorzeitigen Ablösen besitzen, keine exakten Angaben über die Proteinmengen möglich sind. Dieser prinzipielle Nachteil gilt aber auch für andere Techniken wie z.B. den „klassischen“ Western Blot mit Antikörpern: Auch hier kann es zu unrealistisch schwachen Signalen kommen, falls sich das Protein samt gebundenem Antikörper während der Blockier- und Waschschrte noch vor der eigentlichen Detektion von der Membran ablöst. Als Kontrolle diente bei den hier gezeigten Protein-Protein-Bindungsstudien eine Ponceau S-Färbung, mit der jeweils ganz am Ende gezeigt wurde, dass sich wirklich Protein auf der Membran befunden hat (ein negatives Ergebnis wäre sonst wenig aussagekräftig).

Abgesehen von der unterschiedlichen Präparation der Membran mit den Zielproteinen ist der Versuchsablauf für beide Methoden identisch: Eines der Proteine wird über die PKA-Markierungsregion im Fusionsanteil (siehe [3.6.1]) spezifisch radioaktiv markiert, und dient im folgenden als „Sonde“, die ähnlich einem Antikörper in freier Lösung zugegeben wird. Die Zielproteine sind an eine Membran gekoppelt (meist Nitrocellulose), und werden in Blockierungslösung teilrenaturiert. An die, in die Flüssigphase hineinragenden Proteinanteile und die mutmaßlichen Bindedomänen kann die Sonde durch spezifische Interaktion binden. Falls die Wechselwirkung stark genug ist, bleiben die Proteine auch während der folgenden Waschschrte komplexiert, was (nach dem Trocknen der Membran) mit Hilfe einer Autoradiographie nachgewiesen werden kann.

4.11.1.1 Far Western Dot Blotting

Wie oben erwähnt, zeigte sich bald das Problem, dass die Untereinheiten der RNA-Polymerase bei *Pyrococcus* recht unterschiedlich gut löslich sind. Nachdem der Auftrag rechnerisch ähnlicher Proteinmengen in der Ponceau S-Färbung drastische Unterschiede gezeigt hatte, wurde bei späteren Experimenten eine Angleichung der Proteinmengen entsprechend der resultierenden Färbung vorgenommen (grundsätzlich muß einkalkuliert werden, dass die gebräuchlichen Farbstoffe Proteine teilweise etwas unterschiedlich gut anfärben; die Unterschiede in der Anfärbbarkeit wurden in diesem Fall als vernachlässigbar gering eingeschätzt). Hierzu wurde in Abhängigkeit von der Proteinkonzentration der jeweiligen Protein-Präparation bis zu viermal auf einen Punkt aufgetropft. Auch wenn sich in der Praxis gezeigt hat, dass mit der Far Western-Technik selbst bei recht stark abweichenden Proteinmengen vergleichende Aussagen über die Affinität eines Proteins zu einem anderen Protein gemacht werden können, muß aus den oben genannten Gründen von einer

semi-quantitativen Methode gespochen werden. Für eine vollständig quantitative Aussage über das Bindungsverhalten zweier Proteine müßte das Ziel-Protein z.B. durch kovalente Bindung an eine feste Matrix gekoppelt werden.

Um die relevanten Signale von Artefakten unterscheiden zu können, wurde jedes Zielprotein in Form zwei senkrecht übereinander angeordneten Spots (Punkte) aufgetropft.

Aus Platzgründen können in diesem Abschnitt nicht alle Autoradiographien von Far Western Blots gezeigt werden – vor allem die Dot Blots wurden vielfach wiederholt und abgewandelt (verschiedene Proteinmengen wurden eingesetzt und z.B. die Trocknungsdauer und –temperatur der Membran variiert). Es handelt sich um eine Auswahl von Ergebnissen. Um Wiederholungen zu vermeiden, werden die Ergebnisse der beiden Varianten „Far Western Dot Blotting“ und „Far Western Blotting“ nur knapp kommentiert. Im Anschluß werden dann in einer Zusammenfassung alle Ergebnisse für jede Untereinheit beziehungsweise TFB und TBP diskutiert.

RNAP-Untereinheiten, die als Sonde verwendet wurden: D, E', F, H, K, L, N, K, P

Bei den hier gezeigten Autoradiogrammen von Far Western Dot-Blots wurde wie beschrieben (siehe [3.6.4]) die Nitrocellulose-Membran direkt mit den verschiedenen Proteinlösungen betropft und anschließend getrocknet. Für eine leichtere Identifizierung und Zuordnung auf dem später aufgelegten Film bei der Autoradiographie wurden mit Hilfe einer Schablone alternierend jeweils 2 Spots pipettiert. Die gleiche Schablone wurde nach dem Einscannen der Autoradiographien zur besseren Orientierung halbtransparent über die Fotos gelegt.

Die großen Untereinheiten B, A' und A'' wurden nur als Zielprotein, nicht jedoch als Sonde verwendet. Während Untereinheit B in dieser Arbeit nicht kloniert wurde, sondern aus einem SDS-Gel mit nativ gereinigter RNA-Polymerase ausgeschnitten und eluiert wurde (Zerkleinern der PA-Gel-Stückchen und Schütteln in 0,1% SDS über Nacht; anschließend Entfernen des SDS durch einen Fällungsschritt, und Renaturierung in einem geeigneten Puffer), erfolgte die Klonierung von rpoA1 und rpoA2 zunächst in das Plasmid pET-14b (und nicht pET-33b), so dass hier im resultierenden rekombinanten Protein keine Erkennungssequenz für die radioaktive Markierung mit PKA vorhanden ist.

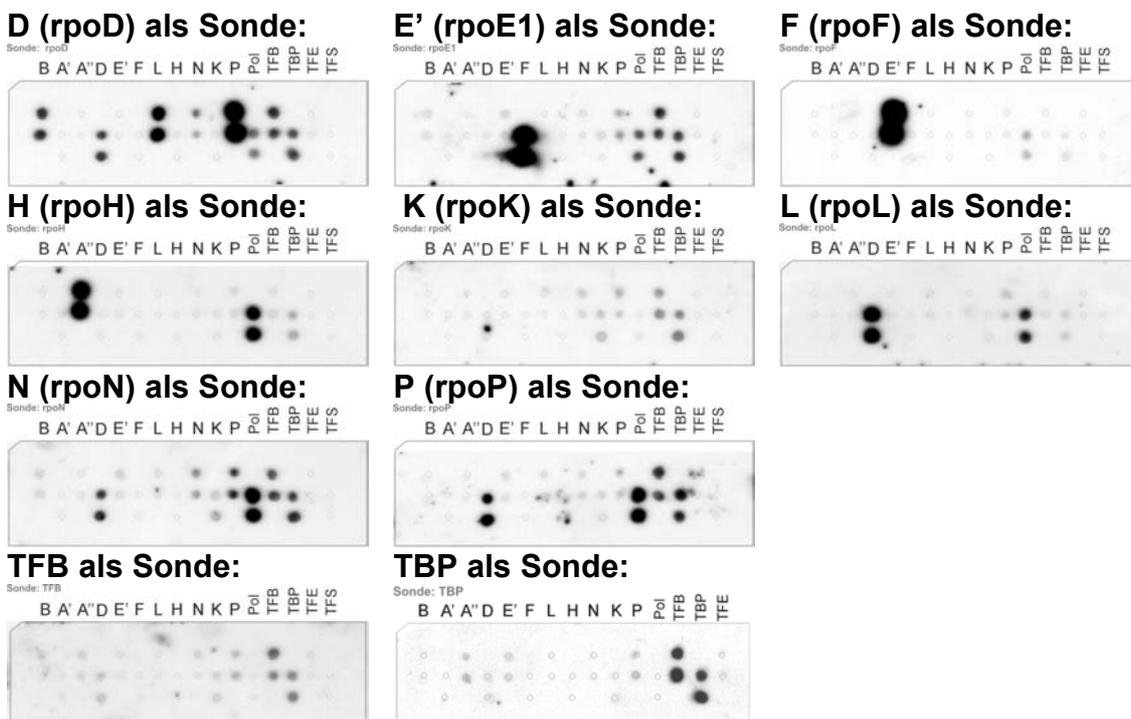


Abb.4-42: Ausgewählte Ergebnissen aus Far Western Dot-Blot-Experimenten

Zuordnung der Signale mittels der über den Spots angebrachten Bezeichnungen: **B – P**: RNAP-Untereinheiten aus *P.furiosus*. **Pol**: Nativ gereinigte (komplette) RNAP, **TFB**: archaeeller Transkriptionsfaktor B (homolog zu TFIIB bei *Eukarya*), **TBP**: archaeelles TATA-Box-Bindeprotein, **TFE** und **TFS**: zwei weitere Transkriptionsfaktoren.

Die wiederholte Durchführung der Dot Blot-Experimente zeigte, dass der gleichmäßig gute Zustand der verschiedenen Protein-Präparationen (in diesem Fall z.B. 16) für ein aussagekräftiges Ergebnis entscheidend ist, und die Anzahl von Auftau – und Gefrierzyklen möglichst gering gehalten werden sollten. Auch wenn es dem entsprechend leichte Variationen der gemessenen Bindungsstärken gab, waren die wesentlichen Ergebnisse sehr gut reproduzierbar.

Eine der stärksten und spezifischsten gemessenen Bindungen war die zwischen Untereinheit E' und F, was sich sowohl mit E' und membrangebundenem F, also auch umgekehrt zeigte. Vor allem F scheint beinahe ausschließlich an E' zu binden, wogegen die meisten übrigen markierten Proteine jeweils an mehrere Zielproteine banden. Teilweise mußte entschieden werden, welche Signalstärke noch als spezifisch, und welche als „Hintergrund“ einzustufen war. Wird H als Sonde verwendet, findet sich eine ähnlich eindeutige (und starke) Bindung an die Untereinheit A'' wie F an E', d.h. es sind sonst fast keine Signale erkennbar. Umgekehrt sind die Interaktionen der anderen markierten Proteine mit Untereinheit H als Zielprotein ausgesprochen gering (und A'' ist schließlich momentan noch nicht als Sonde verfügbar). Die letztgenannten Beobachtungen müssen jedoch mit Vorsicht gehandhabt werden: Gerade rpoH ist ein „Problem-Protein“, das sich durch seine gute Löslichkeit während der Inkubations- und Waschschriffe leicht von der Membran ablöst. Dieses (an der obligatorischen Ponceau S-Färbung erkennbare) unterschiedliche Verhalten von Proteinen ist in die Auswertung mit einzubeziehen.

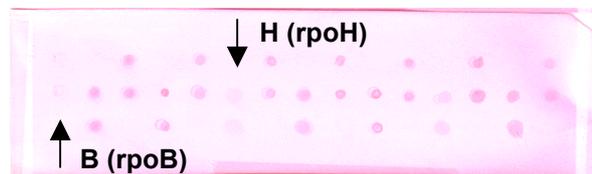


Abb.4-43: Ponceau S-Färbung einer Dot Blot-Membran: Optimierung der Proteinmengen

Gefärbte Nitrocellulose-Membran mit aufgetropften Proteinen. Wegen der unterschiedlichen Protein-Eigenschaften wurde in einigen Fällen durch mehrfaches Auftropfen die Proteinmenge erhöht, um die unterschiedliche Neigung der Proteine zum Abdiffundieren von der Membran während der Inkubationsschritte auszugleichen (es wurde jeweils gewartet, bis der vorhergehende Tropfen getrocknet war). Von links nach rechts Auftrag folgender Proteine: RNAP-Untereinheiten B, A', A'', D, E', F, L, H, N, K, P, RNAP, TFB, TBP, TFE. Neben rpoH zeigte auch rpoB eine verstärkte Neigung zum Ablösen von der Membran, was bei letzterem aber vermutlich auch an dem Auftrag in denaturierter Form nach der Elution aus einem 6%igen SDS-Gel resultierte. Durch kurzzeitig erhöhte Temperatur bei der Trocknung (42°C bzw. 50°C) konnten in beiden Fällen bessere Ergebnisse erzielt werden. Proteine, bei denen die Auftragsmenge erhöht wurde, waren besonders A', A'', F, H, TFE und in geringerem Maße B, P und RNAP. Bei den übrigen Proteinen war die Neigung zum Ablösen von der Membran gering. Insgesamt wurden Proteinmengen im Bereich von 0,5 – 3µg aufgetragen.

Andererseits hat sich unter anderem bei Untereinheit B gezeigt (die aus einem SDS-Gel eluierten Protein-Mengen beziehungsweise die bis zur finalen Ponceau S-Färbung auf der Membran verbliebenen Protein-Mengen waren in einigen Fällen recht gering), dass, falls zwei Proteine tatsächlich eine hohe Affinität zueinander besitzen, trotzdem ein starkes Signal resultiert, selbst wenn sich nur noch Spuren des Proteins auf der Membran befinden. Hier zeigt sich das Far Western Blotting ähnlich sensitiv wie ein "klassischer" Western Blot.

An dritter Stelle sind die Interaktionen des DLNP-Komplexes (gemeint ist D-L-N-P – die diskutierten Proteinkomplexe werden im folgenden meist ohne Bindestrich geschrieben) zu nennen, wobei auch die große Untereinheit B mit ins Spiel kommt, an die diese Gruppe kleinerer Proteine hauptsächlich assoziiert ist.

Untereinheit K und TBP lieferten grundsätzlich schwache Signale, was bei TBP vielleicht dadurch bewirkt wird, dass dieses Protein im nativen Zustand stabile Dimere bildet, bei denen einige Bereiche des Proteins von möglichen weiteren Interaktionen ausgeschlossen bleiben. Für K hatten die geringen Signalstärken zur Folge, dass aufgrund der Versuchsergebnisse bis zum Abschluß dieser Arbeit nicht mit Sicherheit gesagt werden konnte, an welche anderen RNA-Pol-Untereinheit rpoK bindet (von der 3D-Struktur kommen A'', A', E' und eventuell sogar B in Frage; siehe Zusammenfassung der Ergebnisse der Far Western Blot-Experimente [Abb.4-47]). Allerdings konnte bei einer TFB-Variante mit leichten N- und C-Terminalen Deletionen eine etwas stärkere Interaktion mit K beobachtet werden (siehe [4.11.1.2], weiter unten).

Mit TFB als Sonde wurden in einigen Fällen ähnlich geringe Bindungen detektiert, was in diesem Fall allerdings vor allem an der (trotz Blockierungslösung erfolgenden) großflächigen, unspezifischen Anlagerung von TFB an die Nitrocellulose-Membran lag. Möglicherweise hätte man hier mit PVDF-Membran bessere Ergebnisse erzielt. Allerdings lieferten TFB vor allem bei aus einem PA-Gel transferierten Zielproteinen trotzdem genügend deutliche Signale.

Das in dieser Versuchsreihe mit auf die Membranen aufgetragene TFE (Voss, 2003) und TFS (aus *M.jannaschii*) ergab in keinem Fall ein deutliches, verwertbares Signal.

Da die Interaktion der RNA-Polymerase mit den Transkriptionsfaktoren einer der wichtigsten Schritte bei der Bildung eines Initiationskomplexes (im Rahmen der Initiation der Transkription) darstellt, wurden weitere Untersuchungen bezüglich potentieller Wechselwirkungen von TBP und TFB mit einzelnen RNAP-Untereinheiten unternommen:

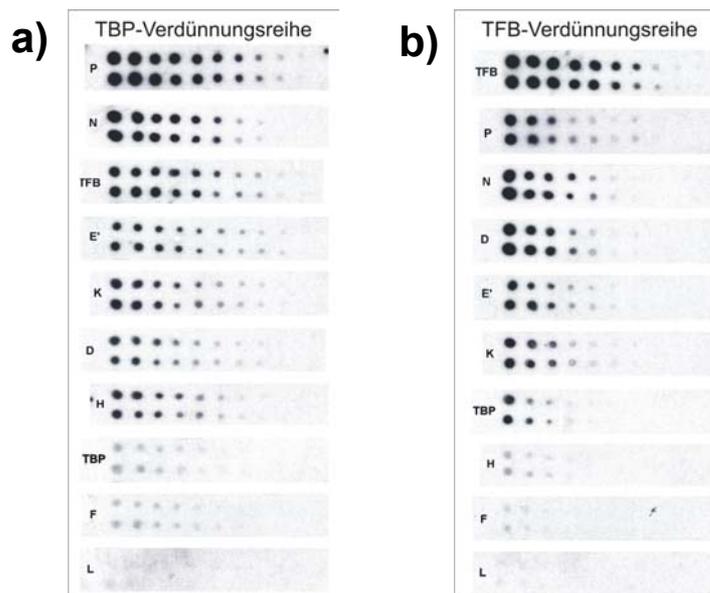


Abb.4-44: Far Western Dot Blotting-Experimente: TBP- und TFB-Verdünnungsreihe

Es wurden jeweils 10 Verdünnungsreihen für TBP und TFB auf separate Nitrocellulose-Membranen aufgetropft, so dass diese mit allen bisher verfügbaren, radioaktiv markierbaren Proteine inkubiert werden konnten. Die Fotos der resultierenden Autoradiographien wurden entsprechend der beobachteten Signalstärken absteigend sortiert. **a)** TBP-Verdünnungsreihe: 1:1 (1,2µg) bis 1:38,4 (31ng), **b)** TFB-Verdünnungsreihe: 1:1 (2µg) bis 1:38,4 (52ng). Jeweils links neben den Autoradiographien Angabe des als Sonde verwendeten Proteins (D, E', F, H, K, L, N, P: Untereinheiten der RNAP, sowie TBP und TFB). Menge eingesetzten Sondenproteins: TFB und rpoD: 1,2µg; TBP, rpoE1: 1µg, rpoF: 0,6µg; rpoL, rpoH, rpoN, rpoK: ca. 0,5µg; rpoP: 0,2µg.

Die Entscheidung, für diesen technisch umfangreichen Versuch (mit insgesamt 20 separaten, parallel durchgeführten Far Western Dot-Blotting-Experimenten) TBP und TFB an die Nitrocellulose-Membran gekoppelt (und nicht als Sonde) zu verwenden, war die zu erwartende höhere Reproduzierbarkeit der Ergebnisse (beide Proteine binden recht gut an die Membran, so dass die Verdünnungsreihen wirklich die angegebenen Proteinmengen enthalten; bei TBP entfällt vermutlich teilweise das oben geschilderte Problem der Dimerisierung; bei TFB wird das Problem einer unspezifischen Bindung an die Nitrocellulose-Membran umgangen und die Wahrscheinlichkeit von (für Bindungsstudien ungünstigen) intramolekularen Wechselwirkungen verringert (siehe Erläuterungen zu [Abb.4-46] im Text).

Die Sortierung der jeweiligen Ergebnisse nach Signalstärken ergab folgende, nach abnehmenden Protein-Protein-Wechselwirkungen sortierte Reihen:

TBP: P, N, TFB, E', K, D, H, TBP, F, L

TFB: TFB, P, N, D, E', K, TBP, H, F, L

Obwohl die entsprechend [Abb.3-01, b)] durchgeführte Kontrolle der 10 verschiedenen Markierungsreaktionen ergeben hatte, dass rpoL etwas schwächer als die meisten übrigen Proteine markiert war, ergibt sich hier eindeutig eine minimale Affinität sowohl zu TBP als auch zu TFB.

Das Ergebnis paßt gut zu den übrigen Versuchen mit L als Sonde (siehe [Abb.4-42]). Somit scheidet rpoL (zumindest auf Basis der hier unternommenen Bindungsstudien) als Kandidat für Wechselwirkungen mit TFB und TBP im Rahmen des weiter unten diskutierten "erweiterten Interface der archaeellen RNAP" aus. rpoP (als weiterer Bestandteil des DLNP-Komplexes) nimmt bei den Protein-Protein-Wechselwirkungen sowohl bei TBP als auch TFB eine herausragende Position ein (die starke Bindung von TFB an sich selbst dagegen könnte die gleiche Ursache haben, wie die – zumindest für TFIIB bei *Eukarya* nachgewiesene (siehe oben und [Abb.4-46]) Neigung zu intramolekularen Wechselwirkungen des N-Terminus mit dem C-Terminus, und ist deshalb in diesem Zusammenhang nicht so interessant).

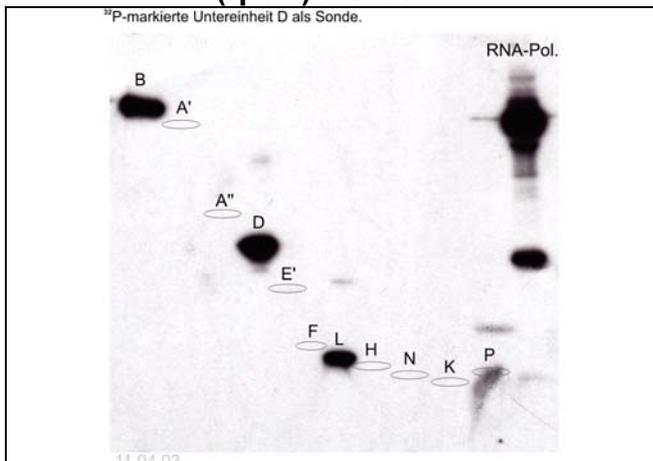
Anschließend folgt rpoN, welches auch zum DLNP-Komplex gehört, allerdings eigentlich (entsprechend der eukaryotischen 3D-Struktur der RNAP) nicht mehr in direkter "Reichweite" von TFB/TBP liegt. Zusammenfassend kann man sagen, dass Kontakte von rpoK und rpoE1 mit Komponenten des Präinitiationskomplexes aus der Literatur bekannt sind, beziehungsweise vermutet werden, was die hier detektierten Signale plausibel erscheinen lassen würde. rpoF zeigt nur schwache Bindungen, obwohl (aufgrund der hohen Beweglichkeit des "Keils" aus rpoE' und rpoF) ein Kontakt durchaus denkbar wäre; siehe [Abb.4-50]).

rpoD schließlich zeigt genügend starke Interaktionen sowohl mit TFB als auch mit TBP, dass seine potentielle Beteiligung am archaeellen "RNAP-Interface" zusammen mit anderen Komponenten des DLNP-Komplexes weiter unten diskutiert werden kann.

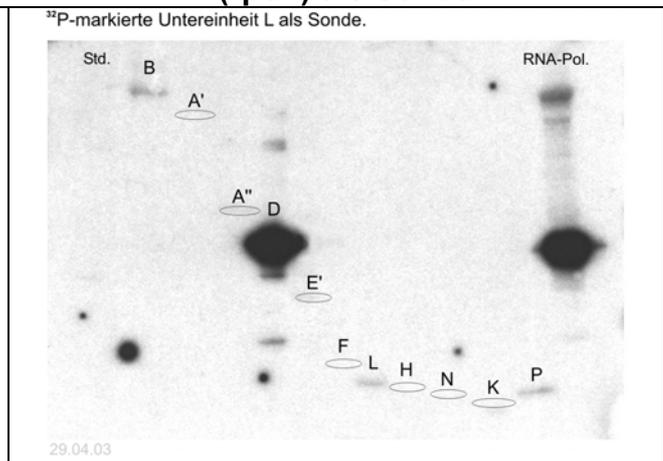
4.11.1.2 Far Western Blotting

Die Technik des Far Western Blottings ist in bezug auf Blockier-, Markierungs- und Waschschrte identisch mit dem Far Western Dot Blotting. Allerdings sind die Zielproteine nach Auftrennung in einem PA-Gradientengel mit 0,1% SDS bzw. nach dem Elektroblob schon vollständig denaturiert, wenn sie an die Membran binden. Dies ist kein Nachteil, weil während der (in der Regel über Nacht durchgeführten) Blockierung der Membran genügend Zeit für die nötige partielle Renaturierung ist. Einerseits erhält man gegenüber einem Dot Blot eine zusätzliche Information bezüglich Größe und eventuell vorhandene Nebenbanden. Die nativ gereinigte, hier in ihre Untereinheiten aufgetrennte RNA-Polymerase zeigt in den meisten Fällen markierte Banden auf der gleichen Höhe, wie die rekombinanten Untereinheiten, und dient auf diese Weise als Kontrolle. Andererseits besteht aber auch das Problem, dass je nach Größe, Form, Ladung, Löslichkeit etc. unterschiedlich viel Protein im Gel zurückbleibt, so dass die später auf der Membran nachweisbaren Proteinmengen nicht unbedingt exakt denen entsprechen, die auf das SDS-Gel aufgetragen wurden.

D (rpoD) als Sonde:



L (rpoL) als Sonde:



(Legende: siehe folgende Seite)

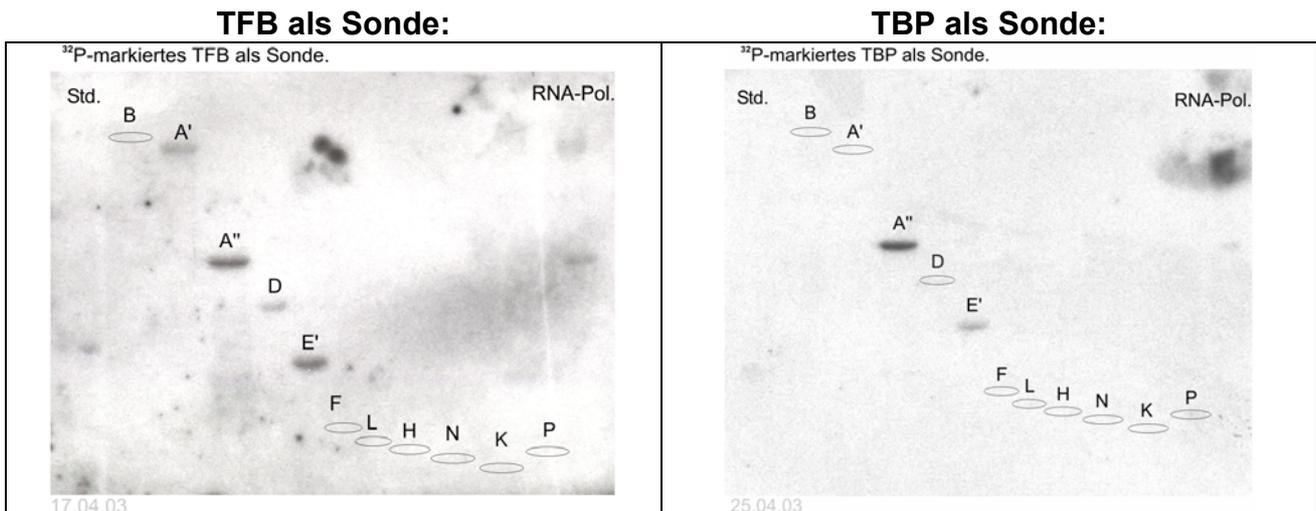


Abb.4-45: Ergebnisse aus Far Western Blotting-Experimenten mit rpoD, rpoL, TFB u. TBP als Sonde
Std.: Molekulargewichtsstandard, **RNA-Pol.:** nativ gereinigte RNAP, **B:** aus einem 6%igen SDS-Gel extrahierte Untereinheit B der nativen RNAP, **A' – P:** Untereinheiten der RNAP; denaturierendes PA-Gradientengel (PA 8-15%, 0,1% SDS).

Zu Einsatz kam bei den dargestellten Experimenten jeweils ein denaturierendes PA-Gradientengel, aus dem die aufgetrennten Proteine auf eine Nitrocellulose-Membran per Elektroblothing transferiert wurden.

Die beiden Filme der Far Western Blots mit Untereinheit D und L als Sonde bestätigen mit kräftigen Signalen die Ergebnisse der entsprechenden Dot Blotting-Experimente (Interaktionen innerhalb des DLNP-Komplexes, der hauptsächlich über die viertgrößte Untereinheit D an die größten Untereinheit B bindet). Während D (ähnlich wie auch bei einigen anderen Proteinen) trotz seiner guten Löslichkeit Komplexe mit weiteren rpoD-Molekülen bildet und deshalb an der entsprechenden Position auf dem Film eine starke Schwärzung verursacht, scheint L kaum an andere L-Untereinheiten sondern hauptsächlich an D zu binden. Dies paßt recht gut zu den Nativ-PAGE-Ergebnissen mit Untereinheit D und L, wo markiertes D eine diffuse „Wolke“ bildet, während markiertes L eine scharfe Bande liefert (siehe [Abb.4-53], [Abb.4-54]).

Das Signalmuster auf den Filmen der Far Western Blots mit TFB und TBP als Sonde schien zunächst etwas überraschend, weil eine starke Bindung an A'' und E' in den Dot Blot-Experimenten nicht so klar sichtbar war. Die weitere Analyse mit Hilfe der eukaryotischen 3D-Struktur und den im Anhang gezeigten Alignments zeigte (neben entsprechenden Literaturangaben) aber, dass eine Bindung an E' durchaus plausibel wäre. Zu dem starken Signal bei A'' ist zu sagen, dass ein entsprechender Kontakt in einem korrekt zusammengesetzten Initiationskomplex nach bisherigen Vorstellungen ziemlich unwahrscheinlich ist, weil sich A'' gemeinsam mit rpoH auf der gegenüberliegenden (bezogen auf die DNA stromabwärts gelegenen) Seite der RNAP befinden, also deutlich außerhalb der Reichweite der "Interface-Seite" der RNAP, was auch Ergebnisse aus aktuelle Crosslinking-Experimente zeigen, in denen der genaue Weg der DNA durch die RNAP aus *P.furiosus* untersucht wurde (Bartlett *et al.*, 2003). Dieses Ergebnis wird also erstmal zur Kenntnis genommen, bis weitere Daten vorliegen.

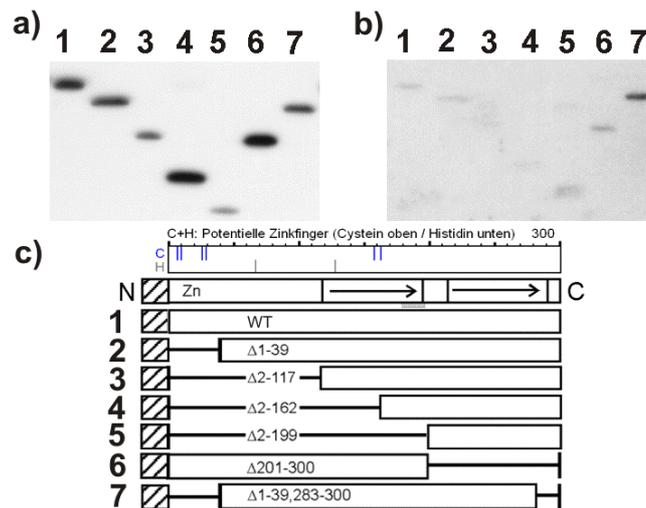


Abb.4-46: Far Western Blot mit rpoD und rpoK als Sonde gegenüber teilweise deletiertem TFB

Ziffern 1-7 in a) – c): 1: Komplettes TFB, 2: TFB Δ 2-39, 3: TFB Δ 2-117, 4: TFB Δ 2-162, 5: TFB Δ 2-199, 6: TFB Δ 201-300, 7: TFB Δ 2-39/283-300. **a)** Sonde rpoD; **b)** Sonde rpoK; **c)** Schematische Darstellung der unterschiedlich umfangreich deletierten Regionen. Schräg schraffiertes Rechteck: Fusionsanteil mit His₆-Tag; obere Zeile mit Darstellung der Cystein- und Histidin-Reste; **N, C**: N- und C-Terminus; **Zn**: Zinkfinger-Motiv; **Pfeile**: "direct repeats"; **WT**: Wildtyp von TFB

In den in [Abb.4-46] gezeigten Experimenten wurden sechs verschieden stark deletierte TFB-Varianten (Goede, 1998) auf Protein-Protein-Wechselwirkungen mit rpoD und rpoK geprüft. rpoD wurde aufgrund seiner aus anderen Far Western Blotting-Versuchen bekannten starken Wechselwirkungen mit TFB ausgewählt. Die Signale sind bei rpoD sehr kräftig (zum Vergleich: beide Filme wurden über Nacht aufgelegt, **a)** 30s, **b)** 2min Entwicklungsdauer), bei rpoK dagegen deutlich schwächer. Es scheint keine spezielle Region zu geben, die allein für die rpoD-TFB-Bindung verantwortlich ist. Eher scheint eine allgemeine Affinität existieren, die vermutlich auch durch die entgegengesetzten Proteinladungen (TFB +13, rpoD –14) mitverursacht wird.

rpoK wurde aufgrund von Literaturdaten eingesetzt, die auf einen Kontakt mit TFB hinweisen (Magill *et al.*, 2001). TFB scheint in Abwesenheit spezifischer Interaktionspartner (wie z.B. an die TATA-Box gebundenes TBP oder auch Aktivatorproteine) eine spezielle Konformation anzunehmen, bei der es zu einer intramolekularen Anlagerung des N-Terminus an den zweiten "direct repeat" am C-Terminus kommt. Auf diese Weise sind bestimmte Binderegionen z.B. für die Pol II oder daran assoziiertes TFIIF blockiert. Erst die Wechselwirkung zum Beispiel mit einem "sauren Aktivator", weiteren Komponenten des Präinitiationskomplexes oder aber durch künstliche Deletion der für die intramolekulare Anlagerung von N- und C-Terminus verantwortlichen Regionen ergibt sich eine andere Konformation von TFB und die Freilegung von Bindedomänen (Roberts & Green, 1994). Dies könnte auch die Ursache für das oben gezeigte Ergebnis sein [Abb.4-46, c): Das Entfernen des N-terminalen Zinkfinger-Motivs in Verbindung mit der Deletion eines Teils des zweiten "direkt repeats" könnte die intramolekulare Anlagerung verhindern, so dass sich eine starke Bindung an dieses (relativ moderat) deletierte TFB als an den Wildtyp ergibt. Zwar gibt es Hinweise, dass gerade der N-Terminus von TFB mit seinem Zinkfinger-Motiv entscheidend für die Bindung von rpoK ist (Magill *et al.*, 2001), jedoch könnte auch der direkt hinter dem Zinkfinger liegende Bereich ausreichend für eine erhöhte Affinität zu rpoK sein, was das im Versuch stärkste Signal bei TFB Δ 2-39/283-300 erklären würde.

4.11.1.3 Zusammenfassung der Far Western Blotting-Ergebnisse und Diskussion

Zusammenfassung der in allen Far Western Blotting-Experimenten beobachteten Bindungsstärken														
Membran Sonde	B	A'	A''	D	E'	F	H	K	L	N	P	TBP	TFB	PoI
D (rpoD)	+++			++					+++	+	+++	++	++	+
E' (rpoE1)	(+)				(+)	++++		(+)			(+)	++	++	++
F (rpoF)					++++						(+)	(+)	(+)	+
H (rpoH)			+++									+		+++
K (rpoK)								(+)		(+)	(+)	+	+	
L (rpoL)	(+)			+++					(+)		(+)			++
N (rpoN)	(+)			++				(+)		(+)	++	++	++	+++
P (rpoP)	++			++					+	+		++	++	+++
TBP			+++*		+*						+	+	+	
TFB		+*	+++*	+*	+*			+		(+)	+	++	++	

* = Signale waren vor allem bei Transfer der Zielprotein aus einem PA-Gel erkennbar, deutlich schwächer jedoch beim "Dot Blotting"

Abb.4-47: Tabellarische Übersicht der beobachteten Bindungsstärken beim Far Western Blotting

Spalten: Untereinheiten der RNAP, TBP, TFB und die native RNAP, die auf einer Nitrocellulose-Membran fixiert wurden.
Zeilen: Die mittelgroßen und kleinen Untereinheiten der RNAP, TBP und TFB, die (radioaktiv markiert) als Sonden eingesetzt wurden. Dass Sonden in einigen Fällen auch an sich selbst binden, wurde in vielen Experimenten beobachtet. Die Anzahl der Pluszeichen steht für die Stärke der gemessenen Protein-Protein-Bindungen. Da jedes Experiment leichte Abweichungen in der Verteilung der Signalstärken zeigt, handelt es sich hier um einen Durchschnittswert aus jeweils mehreren Versuchen. Ein Pluszeichen in Klammern steht für ein sehr schwaches Signal, das sich aber zumindest in einigen Versuchen von den Hintergrund-Signalen abgehoben hat. In die Bewertung wurden auch Ergebnisse aus Versuchen einbezogen, die Oliver von Kampen mit Untereinheit P als Sonde durchgeführt hat (von Kampen, 2003).

Bei der Betrachtung der Daten in der oben gezeigten Tabelle fällt zunächst auf, dass die Far Western-Methode offensichtlich selektiv ist, und dass die als Sonde eingesetzten, radioaktiv markierten Proteine jeweils ein gegenüber den anderen eingesetzten Sondenproteinen individuelles Muster an Protein-Protein-Wechselwirkungen zeigen.

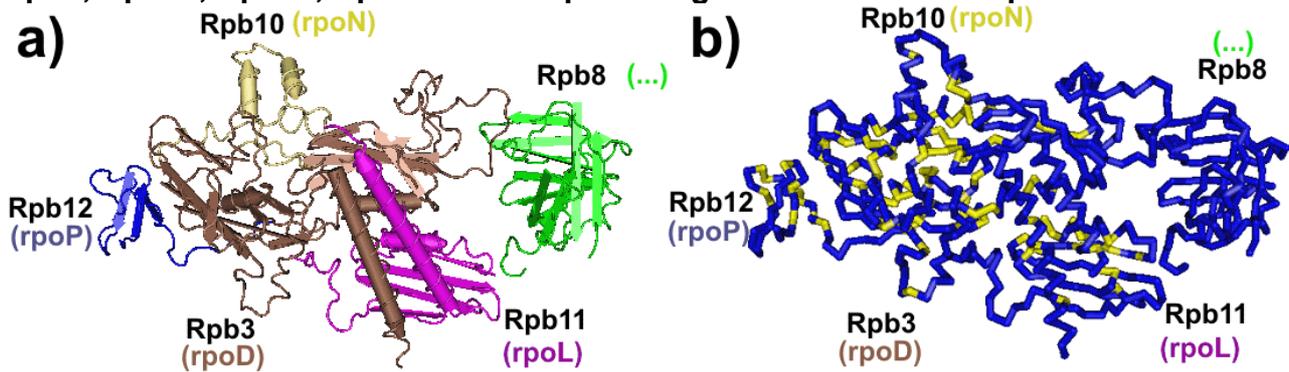
Da die großen Untereinheiten B, A' und A'' (noch) nicht in einer, per Proteinkinase A-markierbaren Form vorliegen, konnten bisher nicht alle möglichen Kombinationen von Proteinen in beiden Richtungen (das heißt jeweils sowohl als Sonde, als auch als Zielprotein, fixiert auf einer Membran) geprüft und bestätigt werden.

Weiter unten wird in einer ausführlichen Besprechung auf jede Untereinheit und die übrigen untersuchten Proteine eingegangen. Trotzdem sei an dieser Stelle schon auf das Fehlen von Signalen in der Spalte mit H (rpoH) hingewiesen, was einerseits dadurch bedingt ist, dass zwar H gegenüber A'', nicht aber A'' gegenüber H als Sonde getestet wurde. Ein zusätzliches Problem stellte die gute Löslichkeit und das (schon weiter oben erwähnte) teilweise Ablösen des Proteins von der Nitrocellulose-Membran während der Inkubationsschritte dar.

Andererseits erschwert die Fülle von Signalen bei TBP und TFB zunächst die Interpretation (siehe aber auch die Erörterungen zu [Abb.4-44]). Auch wenn sich unter Einbeziehung der eukaryotischen 3D-Struktur der RNAP viele Signale erklären lassen, scheint klar, dass die Bindung von TFB (und TBP) an RNAP-Untereinheiten über multiple Kontakte an mehrere Untereinheiten und nicht über den Kontakt zu einer einzigen Untereinheit erfolgt (was die in der Literatur beschriebenen bevorzugten Bindungen an K (rpoK) und die "Docking"-Region von Rpb1 (A' / rpoA1) nicht ausschließt).

Einer der großen Vorteile der Darstellung von 3D-Strukturen (z.B. des RNAP-Komplexes aus 12 Untereinheiten bei *S.cerevisiae*) direkt am PC mit Hilfe geeigneter Programme (z.B. Cn3D 4.1) ist die mögliche Auswahl individueller Betrachtungswinkel, Darstellungsarten und Teilstrukturen. Für die folgende Zusammenfassung und die Diskussion der jeweiligen Far Western-Versuchsdaten vor dem Hintergrund der eukaryotische 3D-Struktur erwies es sich als sehr nützlich, die beiden größten Untereinheiten Rpb1 und Rpb2 (und bei Bedarf noch weitere Untereinheiten) unsichtbar zu schalten. Auf diese Weise wurden mit dem Programm Cn3D 4.1 folgenden Detail-Grafiken der beiden Gruppen kleinerer Untereinheiten (ohne die Informationsfülle des restlichen RNA-Pol-Komplexes) erstellt, wobei Rpb9 wegen seiner isolierten Position nicht mit enthalten ist:

Gruppe 1:

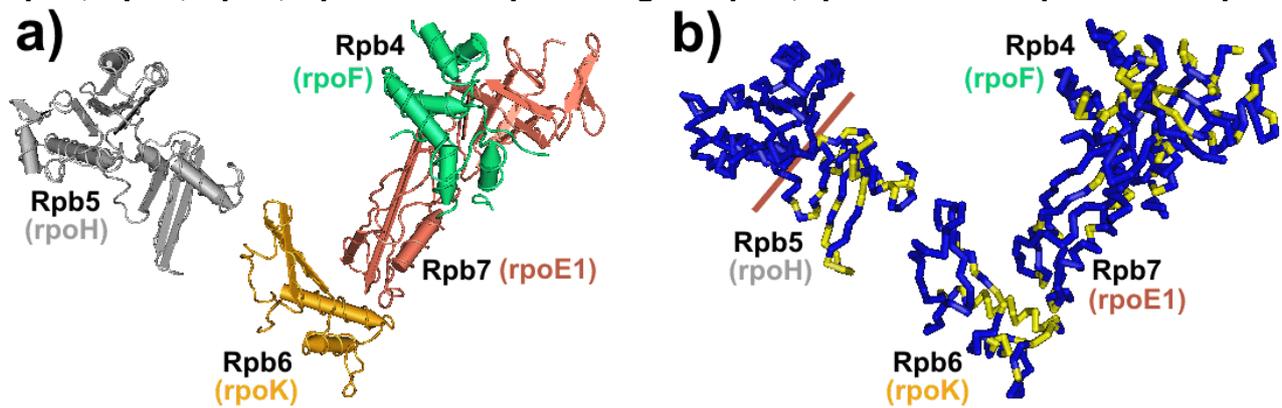
Rpb3, Rpb10, Rpb11, Rpb12 als Entsprechung für den DLNP-Komplex bei *Archaea***Abb.4-48: 3D-Struktur des Rpb3-Rpb10-Rpb11-Rpb12 Komplexes (homolog zum DLNP-Komplex)**

Darstellung von Gruppe 1 der kleineren Untereinheiten des eukaryotischen RNA-Pol-Komplexes. Blick von außen in Richtung Gesamtkomplex. Farbige Beschriftung: homologe Untereinheiten bei *Archaea*. Rpb8 besitzt kein Gegenstück im archaeellen RNAP-Komplex, befindet sich aber in räumlicher Nähe des gezeigten eukaryotischen Komplexes aus 4 Untereinheiten. a) 3D-Darstellung mit Symbolen für α -Helix- und β -Faltblatt-Strukturen. b) 3D-Darstellung mit gelb markierten Regionen mit starker Sequenz-Ähnlichkeit zu *Archaea*; blau mit geringer Ähnlichkeit.

Diese Gruppe von RNAP-Untereinheiten sitzt dem im wesentlichen durch die beiden größten Untereinheiten Rpb1 und Rpb2 (bzw. rpoB, rpoA' und rpoA'' bei *Pyrococcus*) geformten "katalytischen Kern" des Enzymkomplexes peripher auf. Der Kontakt zum Rest des Gesamtkomplexes wird hauptsächlich über Rpb2 (rpoB bei *Pyrococcus*) hergestellt. Etwa das rechte Drittel des in a) gezeigten Ausschnitts des 3D-Modells befindet sich in räumlicher Nähe zur größten Untereinheit Rpb1 (rpoA1), so dass sowohl Rpb3 (rpoD) als auch Rpb11 (rpoL) für Kontakte mit Rpb1 (rpoA1) in Frage kommen (die Bindung von Rpb8 findet ausschließlich über Rpb1 statt). Während diese Interaktion für die (bei *Eukarya* immerhin drittgrößten) Untereinheit Rpb3 nur einen kleinen Teil der Bindungskapazität ausmacht, legt das 3D-Modell für Rpb11 eine intensive Bindung auch an Rpb1 nahe. Somit bildet vor allem der eng verzahnte D-L-Anteil des DLNP-Komplexes eine Art Brücke zwischen den beiden großen Untereinheiten. Da die ähnliche Bedeutung des DLNP-Komplexes als "Assembly platform" auch für *Archaea* bzw. *M.jannaschii* (Rpb3/11/10/12-Komplex bei *S.cerevisiae*) in Rekonstitutionsexperimenten nachgewiesen wurde (Eloranta *et al.*, 1998; Werner & Weinzierl, 2002), wird vermutet, dass (neben der Sequenzhomologie der Aminosäuresequenz) auch die 3D-Struktur in vielen Details zwischen *Eukarya* und *Archaea* ähnlich ist, so dass diese Daten im folgenden zur Diskussion der Ergebnisse aus den Far Western-Experimenten herangezogen werden. So kann beispielsweise aus der 3D-Struktur abgeleitet werden, dass für Rpb10 (rpoN) und Rpb12 (rpoP) neben den Kontakten innerhalb des Viererkomplexes nur weitere Kontakte zu Rpb2 (rpoB), nicht aber zu Rpb1 (rpoA1) in Frage kommen.

Gruppe 2:

Rpb4, Rpb5, Rpb6, Rpb7 als Entsprechung für rpoH, rpoK und den rpoE'/F-Komplex

**Abb.4-49: 3D-Struktur von Rpb5, Rpb6, Rpb7-Rpb4 (homolog zu rpoH, rpoK und rpoE1-rpoF)**

Darstellung von Gruppe 2 der kleineren Untereinheiten des eukaryotischen RNAP-Komplexes. Blick von „vorn-links“ in Richtung Gesamtkomplex. Farbige Beschriftung: homologe Untereinheiten bei *Archaea*. **a)** 3D-Darstellung mit Symbolen für α -Helix- und β -Faltblatt-Strukturen. **b)** 3D-Darstellung mit gelb markierten Regionen mit starker Sequenz-Ähnlichkeit zu *Archaea*; blau mit geringer Ähnlichkeit. Rote Linie: Grenze der bei *Archaea* viel kleineren Untereinheit rpoH.

Von der hier mit "Gruppe 2" bezeichnete Ansammlung von kleinen RNAP-Untereinheiten bilden im engeren Sinn nur Rpb7 (rpoE1) und Rpb4 (rpoF) einen stabilen Subkomplex. Die Integration dieser beiden Untereinheiten in eine Kristallstruktur mit der vollständigen Anzahl aller 12 Untereinheiten für *S.cerevisiae* ist ein wichtiger Beitrag zur Erforschung der Transkription (Armache *et al.*, 2003). Dass für die Interpretation der Röntgenstruktur-Daten in bezug auf Rpb7 und Rpb4 auch ein intensiver Abgleich mit den 3D-Daten des schon früher für *M.jannaschii* veröffentlichten E'-F-Komplexes (Todone *et al.*, 2001) vorgenommen wurde, ist ein gutes Beispiel dafür, wie sich archaeele und eukaryotische Transkriptionsforschung positiv ergänzen können.

Nach bisherigen Erkenntnissen bildet der Komplex aus Rpb7 und Rpo4 eine keilförmige Einheit, die mit dem spitzen schmalen Ende über eine "Tasche" im Rest des RNAP-Komplexes ("Core") verankert ist (siehe [Abb.4-50]), wobei noch diskutiert wird, ob diese Bindung eher ein temporärer Übergangszustand oder dauerhafte Konfiguration *in vivo* darstellt (auf jeden Fall scheint die Bindung von Rpb4 an Rpb7 die Bindung von Rpb7 an die "Core-Polymerase" zu stabilisieren).

Dass neben der, an der Verankerung dieses Zweierkomplexes hauptsächlich beteiligten Untereinheit Rpb1 in geringerem Maße auch Rpb2 und Rpb6 beteiligt sind, findet seine Entsprechung in besonders konservierten Proteinregionen, wie in [Abb.4-49, b)] gut an der gegenüber *Archaea* sehr ähnlichen Kontaktfläche zwischen Rpb6 (rpoK) und Rpb7 (rpoE1) erkennbar ist. Allerdings finden über Rpb6 (rpoK) nach bisherigen Erkenntnissen entscheidende Wechselwirkungen mit den Komponenten des Präinitiationskomplexes (vor allem TFIIB (TFB)) statt, was ein weiterer triftiger Grund für die dargestellten, hochkonservierten Regionen darstellt.

Es ist offensichtlich, dass Wechselwirkungen zwischen Untereinheiten der "Gruppe 1" und "Gruppe 2" extrem unwahrscheinlich sind: Die 3D-Struktur zeigt einen großen Abstand zwischen beiden oben dargestellten Untergruppen, der durch die großen Untereinheiten Rpb1 und Rpb2 ausgefüllt wird. Vor diesem Hintergrund werden an dieser Stelle die in der obigen Tabelle [Abb.4-47] gezeigten Ergebnisse stichwortartig kommentiert, um anschließend zu einem Fazit bezüglich der gemessenen Protein-Protein-Wechselwirkungen in Form einer Grafik zu kommen:

B (rpoB):	Wurde nur membrangebunden, nicht aber als Sonde eingesetzt. Stärkste Interaktion mit rpoD und rpoP; mit rpoL und rpoN (als den übrigen Komponenten des DLNP-Komplexes) viel geringer. Die Bindung an rpoE' ist sehr schwach aber nachweisbar, und betrifft die Verankerung des rpoE'-F-Komplexes an dem Rest der RNAP.
A' (rpoA1):	Insgesamt sehr spärliche Signale. Der Kontakt mit TFB, der nicht in allen Experimenten nachgewiesen werden konnte, ist plausibel, weil rpoA1 die "Dock"-Domäne enthält, die aus der Literatur als eine für die Interaktion der RNAP mit den Komponenten des Präinitiationskomplexes (PIC) wichtige Region bekannt ist.
A'' (rpoA2):	Sehr starke Interaktion mit rpoH, was in guter Übereinstimmung mit dem 3D-Modell und der Literatur ist. Recht kräftige Signale bei TBP und TFB beim Far Western Blotting (Elektroblot aus dem denaturierenden PA-Gel): Diese Protein-Protein-Interaktionen kommen nach bisherigen Kenntnissen <i>in vivo</i> nicht vor, weil A'' gemeinsam mit rpoH "auf der anderen Seite" des RNAP-Komplexes angeordnet ist. Möglicherweise liefern weitere Experimente eine Erklärung hierfür.
D (rpoD):	Eine der Untereinheiten, für die die meisten Kontakte zu unterschiedlichen Proteinen nachgewiesen wurden. Dies sind an erster Stelle die anderen Komponenten des DLNP-Komplexes (und rpoD selbst, d.h. in Abwesenheit der übrigen Untereinheiten und insbesondere rpoL bildet rpoD irreguläre Komplexe mit sich selbst, siehe auch [Abb.4-54]). Das Signal bei TFB bei Transfer aus dem denaturierenden PA-Gel würde die hier aufgestellte Theorie unterstützen, dass bei Archaea auch rpoD und rpoP (vielleicht sogar rpoN?) eine wichtige Rolle bei Kontakten zu Komponenten des PIC besitzen.
E' (rpoE1):	Sehr starke und spezifische Bindung an rpoF, was gut zu Literaturdaten, z.B. zum E'-F-Komplex bei <i>M.jannaschii</i> (Todone <i>et al.</i> , 2001) paßt. Die Kontakte zu TBP und TFB erscheinen (zumindest für TFB) aufgrund bekannter Daten plausibel (bei <i>Eukarya</i> Rpb7 (rpoE1) neben Rpb6 (rpoK) und der "Docking-Region" von Rpb1 (rpoA1) als wichtige "Interface-Regionen").
F (rpoF):	Sehr starke und ausschließliche Bindung an rpoE1, was gut zur Tatsache paßt, dass rpoF nur vermittelt über rpoE1 Kontakt zum Rest der RNAP aufnimmt.
H (rpoH):	Neben dem starken Signal gegenüber rpoA2, wenn rpoH als Sonde eingesetzt wird keine weitere Bindungen an eine RNAP-UE. Erscheint sehr plausibel, weil rpoH (Rpb5) im 3D-Modell der eukaryotischen RNAP eindeutig nur über den zu rpoA2 homologen Abschnitt von Rpb1 an den Rest des Enzymes gebunden erscheint. Allerdings neigte rpoH zum vorzeitigen Ablösen von der Nitrocellulose-Membran – was möglicherweise dazu geführt hat, dass einige weitere schwache Bindungen nicht detektiert werden konnten. Für das Signal bei TBP siehe Erläuterungen zu rpoA2.
K (rpoK):	Insgesamt sehr schwache Reaktionen. Wechselwirkungen mit TFB und TBP plausibel (siehe rpoE1).
L (rpoL):	Sehr starke und spezifische Reaktion mit rpoD (DL-Komplex). Die Bindung an rpoA1 / rpoB scheint aufgrund der im 3D-Modell sichtbaren Kontaktfläche und die vermutete "Brückenfunktion" des DL-Komplexes zwischen rpoB und rpoA1 zu schwach. Vielleicht liefern die Experimente mit der inzwischen verfügbaren rekombinanten Untereinheit B (persönliche Mitteilung, Naji, S.) eindeutiger Ergebnisse. Ähnlich wie bei N-P (siehe unten) läßt die 3D-Struktur auch für L-P wenn überhaupt, dann nur sehr schwache Wechselwirkungen vermuten (ein schwaches Signal bei Far Western Blotting ist jedenfalls erkennbar).
N (rpoN):	Die Hauptbindung erfolgt hier an rpoD, was plausibel ist, wenn man den DLNP-Komplex bedenkt. Prinzipiell gilt für eine potentielle N-P-Bindung das bei rpoL für den möglichen L-P-Komplex gesagte. Allerdings sind bei rpoN die Signale gegenüber rpoP im Far Western Blot stärker. Für die postulierten Kontakte innerhalb des DLNP-Komplexes (bzw. Rpb3/11/10/12 bei <i>Eukarya</i>) müßte die 3D-Struktur genauer analysiert werden, als es mir möglich war (vor allem müssen dann auch die Seitenketten und nicht nur das Grundgerüst des 3D-Modells mit einbezogen werden; zweifelsfrei ist allerdings die zentrale Position von Rpb3 (rpoD), an das Rpb11 (rpoL), Rpb10 (rpoN) und Rpb12 (rpoP) angelagert erscheinen). Die recht starken Signale für TBP und TFB werden notiert, wenngleich eine tatsächliche Interaktion <i>in vivo</i> bzw. an einem korrekt geformten Initiationskomplex weniger wahrscheinlich als bei rpoP und rpoD erscheint, weil rpoN weiter abseits liegt.

P (rpoP):	rpoP zeigt starke Bindungen an rpoD und rpoB, sowie etwas schwächere Bindungen an L und N, was schon weiter oben diskutiert wurde. Gegenüber TBP und TFB gab es ebenfalls starke Signale, was sich auch in den TBP- und TFB-Verdünnungsreihen zeigte.
TBP:	Sowohl für TBP als auch TFB gilt, dass die Fülle an gemessenen Signalen in Far Western Blotting-Experimenten (vor allem, wenn TBP beziehungsweise TFB als Zielproteine membrangebunden eingesetzt werden) das resultierende Bindungsmuster etwas unspezifisch erscheinen lässt. Allerdings sind die meisten der detektierten Protein-Protein-Wechselwirkungen bei näherer Betrachtung durchaus erklärlich, auch wenn einige Fragen offen bleiben (siehe Erläuterungen zu mutmaßlichen Kontakten zu rpoA2 und rpoH). Bei TBP und TFB handelt es schließlich um "hochkommunikative" Proteine, die auf die Vermittlung von Wechselwirkungen zwischen verschiedenen Komponenten hin optimiert sind. Verglichen mit TFB zeigt sich ein ähnliches, aber in mancher Hinsicht eingeschränkteres Bindungsmuster, was ganz gut mit der geringeren Größe von TBP und mit der Reihenfolge korreliert, in der die Bindung an die Prototorelemente erfolgt (obwohl es anscheinend auch direkte TBP-RNAP-Kontakte gibt, wird doch ein Großteil der Wechselwirkung indirekt durch TFB vermittelt, welches erst nach TBP bindet, bevor die RNAP an den PIC rekrutiert werden kann).
TFB:	Bezüglich der Vielzahl von Wechselwirkungen siehe Kommentare zu TBP. Wie bei TBP haben sich einige Protein-Protein-Kontakte vor allem beim "Far Western Blotting" mit Transfer der Proteine aus einem denaturierenden PA-Gel auf die Membran und weniger beim Dot Blotting gezeigt. Besonders für TFB scheint wahrscheinlich, dass Interaktionen nicht nur mit der "Dock"-Region von rpoA1 (Rpb1), rpoK und dem flexiblen rpoE1, sondern auch mit in deren Richtung weisenden Abschnitten von Komponenten des DLNP-Komplexes möglich sind. Während rpoL außen vor bleibt, sind vor allem rpoP und proD Kandidaten für entsprechende Kontakte. Bei <i>Eukarya</i> besitzen sowohl rpoD als auch rpoP ein Zinkfinger-Motiv an exponierter (und wahrscheinlich für Interaktionen verantwortlicher Position). Von den untersuchten Archaea besitzen 9 Arten ebenfalls dieses Motiv, bei 2 Arten sind noch Teile erkennbar und bei den restlichen Arten (zu denen neben <i>P.furiosus</i> auch <i>M.jannaschii</i> gehört) ist das Motiv überhaupt nicht mehr vorhanden (siehe [8.1.3.4]). Es ist zu diskutieren, ob dies die Wahrscheinlichkeit für einen Kontakt mit TFB im Initiationskomplex verringert. rpoP dagegen besitzt bei allen untersuchten Arten ein komplettes Zinkfinger-Motiv. Die ebenfalls deutlichen Interaktionen von rpoN mit TFB im Far Western Blotting-Versuch sind auf der Basis der eukaryotischen 3D-Struktur nicht so leicht zu erklären – es müssten hierfür weitergehende geometrische Abweichungen im archaeellen RNAP-Komplex gegenüber <i>Eukarya</i> postuliert werden (siehe Anmerkungen zu rpoN weiter oben).

Abschließend noch eine grafische Darstellung der erörterten Protein-Protein-Wechselwirkungen:

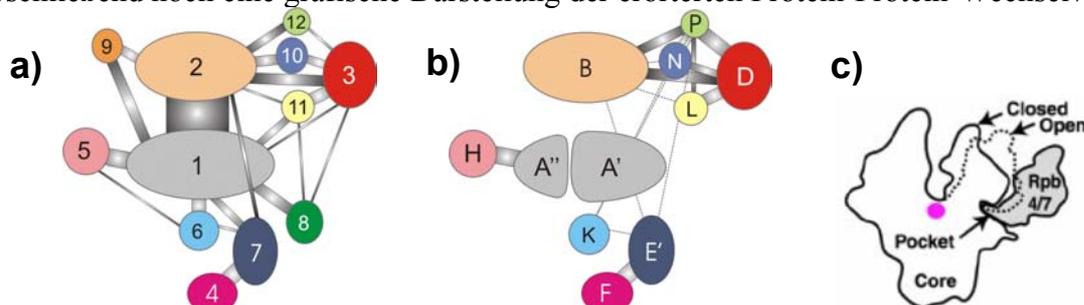


Abb.4-50: Schematische Darstellung von Bindungsstärken zwischen den RNAP-Untereinheiten

a) Abbildung modifiziert nach (Cramer *et al.*, 2001) und sinngemäß ergänzt nach (Armache *et al.*, 2003). Zahlen 1 –12: Rpb1 – Rpb12 der Pol II aus *S.cerevisiae*. Die Dicke der Balken zwischen den Untereinheiten deuten die Intensität der Protein-Protein-Wechselwirkungen zwischen den einzelnen RNAP-Untereinheiten auf der Basis des Umfangs der aus der Kristallstruktur ermittelten Kontaktflächen zwischen den Molekülen an. Der Rpb4/7-Komplex und die Dicke der hiermit verbundenen Balken wurde entsprechend den Angaben in (Armache *et al.*, 2003) sinngemäß hinzugefügt (Verankerung des Rpb4/7-Komplexes hauptsächlich über die Bindung von Rpb7 an Rpb1 mit zusätzlichen, schwächeren Kontakten zu Rpb6 und Rpb2). **b)** Entsprechende Darstellung für *P.furiosus*, in der die Dicke der Balken in etwa der Intensität der mit Far Western Blotting-Experimenten ermittelten Bindungsstärken entspricht. Wie man sieht, konnten (abgesehen von den hier nicht eingezeichneten Wechselwirkungen mit TBP und TFB) mit dieser Technik hauptsächlich die Wechselwirkungen innerhalb des DLNP-Komplexes und mit der größten Untereinheit rpoB, die Bindung von H an A'' sowie der E'F-Komplex nachgewiesen werden. **c)** Abbildung aus (Armache *et al.*, 2003): Die schematische Darstellung zeigt, wie der keilförmige Rpb4/7-Komplex über eine relativ kleinen Kontaktfläche von Rpb7 über eine "Tasche" an dem

Rest der RNAP ("Core") verankert ist. Zusätzlich wird der geöffnete und geschlossene Zustand der "Klammer" (Clamp) gezeigt, die nach erfolgreichem Start der Elongationsphase der Transkription die innerhalb der großen Spalte verlaufende DNA beziehungsweise das DNA-RNA-Hybrid einschließt und auf diese Weise den Komplex stabilisiert.

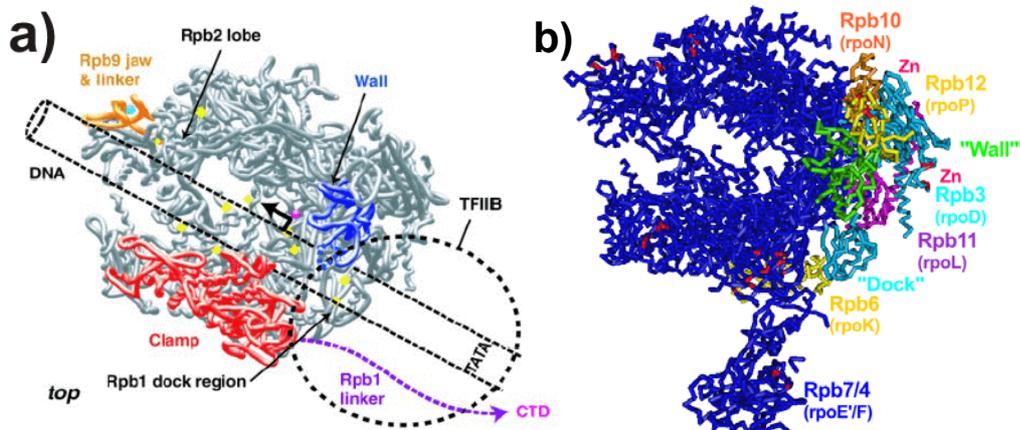


Abb.4-51: Befinden sich Rpb3/Rpb11/Rpb12 (rpoD/rpoL/rpoP) in Reichweite von TFIIIB (TFB)? [1]

a) Abbildung aus (Cramer *et al.*, 2001): Blick von oben hinein in die große Spalte mit dem katalytischen Zentrum (die violette Kugel deutet das Mg^{2+} -Ion an). Entscheidend für die Argumentation ist der durch die schwarz gestrichelten Linien angedeutete Verlauf der DNA und die ungefähre Ausdehnung von TFIIIB (TFB). Zur Orientierung dient die Angabe der der "Wall" genannten Struktur, an der vorbei die DNA und die mRNA die RNAP während der Transkription kontinuierlich verläßt. b) Mit Cn3D 4.1 auf der Basis des aktuellen, 12 Untereinheiten umfassenden 3D-Modells der eukaryotischen RNAP aus *S.cerevisiae* erstellte Grafik. Gleiche Orientierung wie in a), jedoch zusätzlich mit dem Rpb4/7-Komplex. Verschiedene Regionen wurden nachträglich farblich markiert und korrespondierend farblich beschriftet.

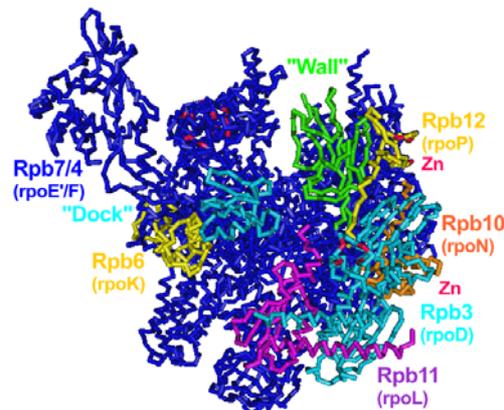


Abb.4-52: Befinden sich Rpb3/Rpb11/Rpb12 (rpoD/rpoL/rpoP) in Reichweite von TFIIIB (TFB)? [2]

Bezeichnungen und Farben wie in [Abb.5-50], b). Ansicht "back", d.h. von der Stromaufwärts-Seite her betrachtet, wo die Wechselwirkungen mit den Komponenten des Präinitiationskomplexes (also besonders TFIIIB (TFB) und TFIID (TBP) stattfinden).

Man darf nicht vergessen, dass es sich bei dem dargestellten Enzymkomplex (trotz der praktischen Beschriftung mit den Namen der archaellen Homologen) um die eukaryotische Pol II handelt. Bis die erste, aus Röntgenstruktur- oder NMR-Daten abgeleitete 3D-Struktur einer archaellen RNA-Polymerase verfügbar sein wird, darf (und muß notgedrungen) auf der Basis aller verfügbaren Informationen und Indizien über den Grad deren Ähnlichkeit gegenüber der 3D-Struktur bei *Eukarya* spekuliert werden.

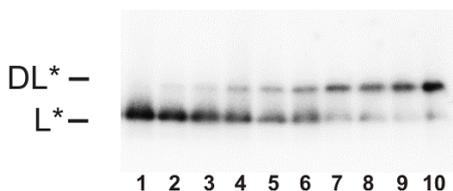
Aufgrund den Größendifferenzen der Untereinheiten zwischen *Eukarya* und *Archaea*, sowie der Abwesenheit einer Vielzahl von an die RNAP und besonders an TBP assoziierten Faktoren bei *Archaea*, werden hier leichte Abweichungen in der gesamten "3D-Geometrie" postuliert, die dazu führen, dass sich TFB und rpoD, rpoP (und vielleicht sogar rpoN) als Komponenten des DLNP-Komplexes beim Binden der RNA-Polymerase an den Präinitiationskomplex noch etwas näher kommen, als in [Abb.4-51, a) und 4-52] für *Eukarya* angedeutet (auch hier scheinen Kontakte mit TFIIIB nicht ausgeschlossen oder sogar wahrscheinlich, selbst wenn der "Dock"-Domäne, Rpo6 und Rpo7 hier eine größere Bedeutung für die Wechselwirkungen zukommen mag). Die in dieser Arbeit durchgeführten Far Western Blotting-Experimente zeigten eine starke allgemeine Affinität von rpoD zu TFB (die vermutlich unter anderem auch auf der genau entgegengesetzten Nettoladung von

+13 bei TFB und -14 bei rpoD beruht). Auch wenn bei diesen Experimenten bisher nicht die genaue Binderegion identifiziert werden konnte, begünstigen diese Ladungsdifferenzen auf jeden Fall Protein-Protein-Bindungen, wie sich auch bei der Verankerung des stark positiv geladenen "Schwanzes" von rpoP (siehe [Abb.4-25] und [Abb.4-15]) am negativ geladenen rpoD zeigt.

Auch die im folgenden noch kurz vorgestellten Ergebnisse aus Experimenten mit Native-PAGE und Molekularsieb-Chromatographie demonstrieren das Auftreten von Protein-Protein-Wechselwirkungen und die Komplexbildung zwischen bestimmten Untereinheiten der RNAP aus *P. furiosus*. Sie stellen einerseits die Bestätigung einiger Ergebnisse der Far Western Blotting-Experimenten dar. Andererseits kann man die durchgeführten Versuche als Vorversuche für das "Assembly" der *P. furiosus*-RNAP aus rekombinanten Untereinheiten betrachten, an dem bereits jetzt in der Arbeitsgruppe von Prof.Dr.M.Thomm an der Universität Regensburg intensiv gearbeitet wird.

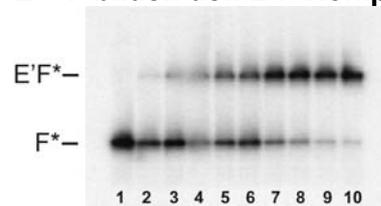
4.11.2 Bindungsstudien: Nativgel-Experimente

D und L bilden einen DL-Komplex.



a) Bildung eines Komplexes aus Untereinheit D und L der RNA-Polymerase aus *P. furiosus* bei ansteigender Konzentration von D. Autoradiographie eines 8-15% Nativgels. L wurde mittels PKA radioaktiv markiert. Bei D N-terminales His₆-Tag, jedoch ohne PKA-Markierungsregion. Inkubation erfolgte 120min bei 30°C in TEN(150mM) mit 10% Glycerin. Wegen der begrenzten Löslichkeit von L wurde zusätzlich 0,5% Tween20 hinzugefügt. Die Konzentration von L* betrug in allen Ansätzen ca. 70ng. Spur 1 ohne D, Spur 2 mit 9ng D, gleichmäßig ansteigend bis Spur 10 mit 48ng D, wo der stabile recht gut lösliche DL*-Komplex überwiegt.

E' + F bilden den E'F-Komplex



b) Bildung eines Komplexes aus Untereinheit E' und F der RNAP aus *Pyrococcus furiosus* bei ansteigender Konzentration von E'. Autoradiographie eines 8-15% Nativgels. F wurde mit Hilfe der PKA radioaktiv markiert. Bei E' handelt es sich um eine Version mit N-terminalem His₆-Tag, jedoch ohne PKA-Markierungsregion. Inkubation erfolgte 90min bei 37°C in TEN(150mM) mit 10% Glycerin. Die Konzentration von F* betrug in allen Ansätzen konstant ca. 70ng. Spur 1 ohne E', Spur 2 mit 62ng E'. In Spur 10 mit 560ng E' ist die freie, markierte Untereinheit F* fast vollständig in den E'F*-Komplex integriert worden (Spuren dazwischen mit linear ansteigender Konzentration von E').

Abb.4-53: Bildung eines DL- und E'F-Komplexes in Nativgel-Experimenten

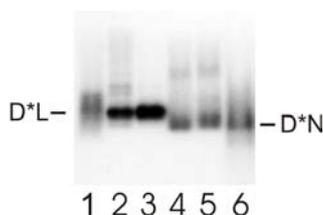


Abb.4-54: Bildung eines DL- und DN-Komplexes in Nativgel-Experimenten

Native-PAGE (8-15%). In dem gezeigten Experiment wurde rpoD statt rpoL (siehe [Abb.4-53, a]) radioaktiv markiert. Die Inkubation erfolgte 180min bei 30°C im Thermocycler. Es zeigt sich, dass im Gegensatz zu rpoL markiertes rpoD im Nativgel eine diffuse Bande bildet (1), was sehr wahrscheinlich dadurch verursacht wird, dass rpoD in Abwesenheit rpoL irreguläre Komplexe mit weiteren rpoD-Molekülen bildet. Dies kann verhindert beziehungsweise rückgängig gemacht werden durch Zugabe von rpoL (2 und 3), so dass der definierte, gut lösliche DL-Komplex gebildet wird. Auch die Zugabe von rpoN zur radioaktiv markiertem rpoD scheint einen ähnlichen Effekt zu haben, was auf die Bildung eines DN-Komplexes schließen lässt (4, 5), der die Bildung der erwähnten irregulären rpoD-Komplexe teilweise verhinert. Allerdings ist der Effekt (und vermutlich auch die Stabilität des resultierenden Komplexes) viel schwächer als beim DL-Komplex. Spur 6 zeigt zum Vergleich radioaktiv markiertes rpoN allein: Die Bande ist erkennbar diffuser als die DN-Bande und läuft auch ungefähr auf der selben Höhe (was in einem Nativgel durchaus möglich ist).

Abgesehen von den gezeigten Experimenten wurde auch versucht, Komplexe mit DLN sowie P und B zu bilden, was aber keine eindeutigen Ergebnisse lieferte. Der Einsatz eines Puffersystems mit erhöhtem pH nach (Ehlers *et al.*, 2002) brachte etwas klarere Ergebnisse, und wurde meist dem Standard-Nativgel ("SDS-Gel ohne SDS") vorgezogen. Selbst mit einem Laufpuffer pH von 9,4 ist rpoP aufgrund seines extremen pI immer noch positiv geladen, so dass die Gefahr besteht, dass das Protein im elektrischen Feld in die entgegengesetzte Richtung wandert. Erhöhte Mengen von rpoP bewirkten ein Ausfallen von Proteinkomplexen in den Geltaschen.

4.11.3 Bindungsstudien: Experimente mit Molekularsieb-Chromatographie

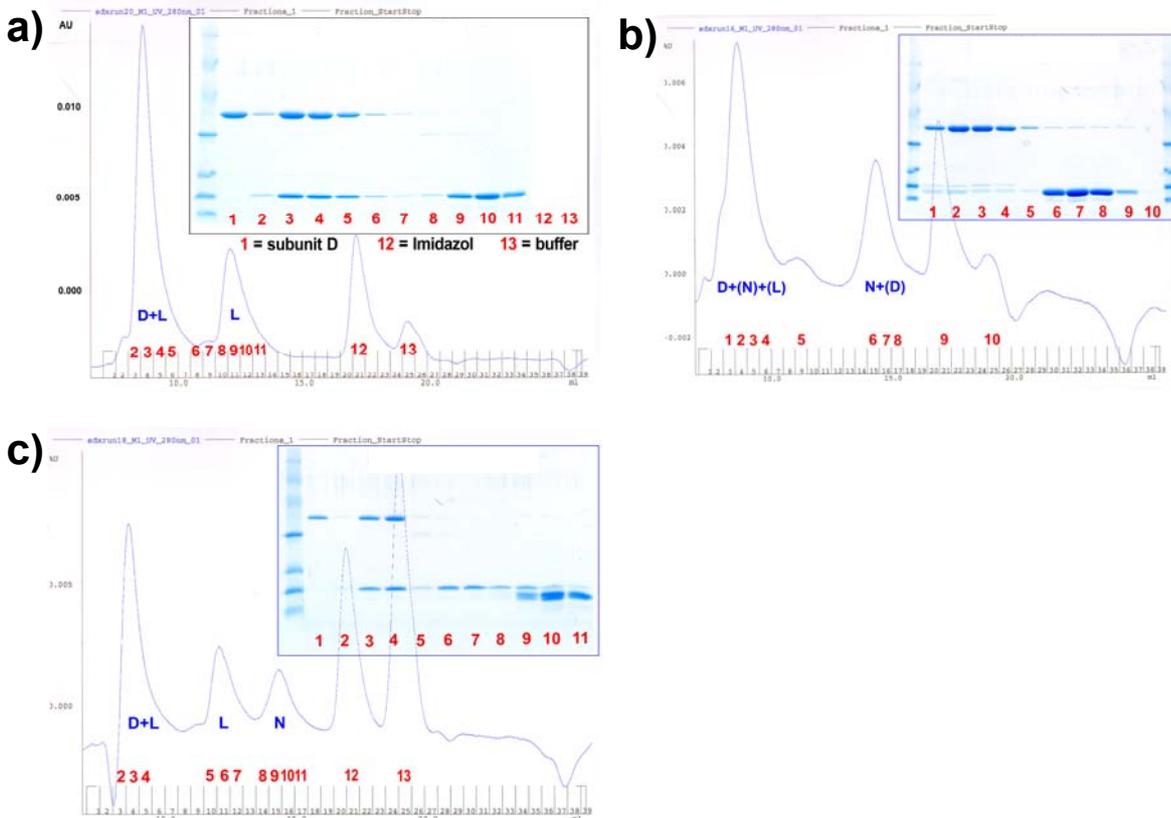


Abb.4-55: Ein DL-Komplex und mutmaßlicher DLN-Komplex bei Molekularsieb-Chromatographie

a) – c): Drei verschiedene Experimente mit der analytischen Superdex 75 (HR 10/30 von Amersham Biosciences). Das Ausschlußvolumen der Chromatographiesäule wurde mit Blue Dextran bestimmt und die Funktionsfähigkeit mit verschiedenen Testproteinen sichergestellt. Nach einer ähnlichen Inkubation wie unter [4.11.2] beschrieben, erfolgte (nach 0,2µm Filtration) der Auftrag von **a)** D+L, **b)** D+N+(L), **c)** D+L+N. **a)** demonstriert am klarsten die Ausbildung des gut löslichen DL-Komplexes: **1:** rpoD allein, **3:** DL-Komplex, **10:** L allein. (Spur 1 ist die Kontrolle mit D; natürlich eluiert der DL-Komplex zuerst von der Säule, weil er das größte Molekulargewicht besitzt). In **b)** deutet sich ein DLN-Komplex beziehungsweise zumindest ein DN-Komplex (**2**) an, nachdem rpoL teilweise auf der Säule ausgefallen war, und während des Laufes von D+N rücklöste (der Effekte der Komplexbildung "auf der Säule" hatte sich auch bei der Kontrolle mit rpoD allein in Form einer störenden Kontamination mit rpoL gezeigt: Obwohl die Säule vorher schon ausgiebig gespült worden war, und kein Protein mehr eluierte, löste sich auf der Säule ausgefallenes oder unspezifisch an die Matrix angelagertes rpoL in der Anwesenheit von rpoD, und bildete den besser löslichen DL-Komplex. **c)** zeigt ein Experiment, bei dem explizit D+L+N eingesetzt wurden, jedoch nur der DL-Komplex bildet. (**1:** D allein, **4:** DL-Komplex, **6:** L allein, **10:** N allein). Offensichtlich muß noch optimiert werden – zum Beispiel in Richtung einer Verminderung der rpoL-Konzentration und eines Überschuß an rpoN. Eine Hypothese, warum sich die DNL- und DNLP-Komplexe in der Praxis nicht so leicht bilden, ist, dass sich gerade bei den kleinen Untereinheiten der Fusionsanteil der rekombinanten Protein besonders stark bemerkbar macht (so beträgt der Fusionsanteil bei der kleinsten Untereinheit rpoP ohne Thrombin-Hydrolyse rund ein Drittel des Molekulargewichtes).

4.12 Reinigung der nativen RNA-Polymerase nach modifiziertem Reinigungsschema

Die native RNA-Polymerase aus *P. furiosus* mit ihren 11-12 Untereinheiten stellt die Ausgangsbasis für die Untersuchung von Protein-Protein-Wechselwirkungen in dieser Arbeit dar. In fast allen Far Western Blotting-Experimenten diente eine gereinigte RNAP-Präparation als Referenz und Kontrolle der Ergebnisse mit den rekombinant in *E. coli* erstellten Untereinheiten. Da die größte Untereinheit B auf *P. furiosus* erst kürzlich erfolgreich kloniert und gereinigt wurde (persönliche Mitteilung, Naji,S.), also in den in dieser Arbeit geschilderten Experimenten noch nicht zur Verfügung stand, wurde B (und teilweise auch A') durch Ausschneiden und Elution aus einem 6%igen SDS-Gel gewonnen.

Weil das etablierte Reinigungsschema für die RNAP aus *P.furiosus* recht langwierig ist (Hausner & Thomm, 1993; Hethke *et al.*, 1996; Hethke *et al.*, 1999), wurde versucht, durch ein verkürztes Schema mit weniger Schritten den Zeit- und Arbeitsaufwand zu verringern. Als Orientierung diente hierbei die Reinigungsprozedur für *Methanobacterium thermoautotrophicum* (Darcy *et al.*, 1999), die dem *Pyrococcus*-Schema in der Abfolge von vier säulenchromatographischen Reinigungsschritten (DEAE-Cellulose, Heparin-Sepharose, Mono-Q, Superdex 200) recht stark ähnelt, jedoch die beiden Fällungsschritte mit Polymin P und Ammonium-Sulfat und die anschließend nötige Dialyse wegläßt. Dieses Schema war z.B. in der Arbeitsgruppe von PD Dr.W.Hausner auch schon erfolgreich für *Methanococcus* eingesetzt worden (Schmeling, 2000). Statt Dialyse reicht in der Regel die Verdünnung der Proben zur Reduktion der Salzkonzentration vor dem Auftrag auf die nächste Säule, wozu am besten ein spezieller Verdünnungspuffer verwendet wird.

Vereinfachend wirkt sich aus, dass im Gegensatz zu *Methanobacterium* und *Methanococcus* bei *Pyrococcus furiosus* keine Einhaltung anaerober Bedingungen während der Reinigung notwendig ist (bei den genannten methanogenen *Archaea* führt Sauerstoffzutritt zu einer Verringerung oder zu einem kompletten Verlust der Enzymaktivität).

Bei der Reinigung der RNAP bei *Pyrococcus* zeigte sich jedoch, dass die Reinheit der Präparation nach dem letzten der vier säulenchromatographischen Reinigungsschritte (nämlich der Molekularsiebchromatographie mit der "Superdex 200", Amersham Biosciences) noch nicht ganz zufriedenstellend war. Es erfolgte deshalb eine erneute Konzentrierung auf der Mono-Q-Säule (Stoßelution), gefolgt von Dichtegradienten-Zentrifugation (Saccharose-Gradient) und einer weiteren abschließenden Konzentrierung auf der MonoQ (Stoßelution), bevor die Präparation portioniert und bei -70°C gelagert wurden.

Die letztgenannten drei Reinigungsschritte verlängern zwar das oben genannte Reinigungsschema. Jedoch ist der Zeitaufwand für die beiden Stoßelutionen mit der Mono-Q-Säule sehr gering, und die Dichtegradientenzentrifugation verlangt, sofern die Gradienten erst einmal gegossen und die Proben aufgetragen sind keinen großen Arbeitsaufwand (die Ultrazentrifuge läuft allerdings ca. 24 Stunden).

Für die Details bei Puffer-Lösungen und Chromatographiesäulen sei auf die oben genannte Literatur verwiesen (TMK-Puffer; DEAE-Cellulose DE52, Whatman; Heparin-Sepharose z.B. HiPrep 16/10 Heparin FF, Mono Q[®] HR 5/5 und HiLoad[™] 16/60 Superdex[™] 200, alle Amersham Biosciences; zwischen den verschiedenen Säulen jeweils TCA-Test der Fraktionen auf RNA-Polymerase-Aktivität).

Der Saccharose-Gradient wurde mit einem gewöhnlichen Zwei-Kammer-Gradientenmischer aus leichter (5% Glycerin, 10% Saccharose, 400mM KCl, TE-Puffer) und schwerer Lösung (10% Glycerin, 30% Saccharose, 400mM KCl, TE-Puffer) "von unten" (also mit der leichteren Lösung in der vorderen Kammer) mittels einer in das Zentrifugenröhrchen gestellten und gegen ungewollte Verwirbelungen des Gradienten fixierten Kanüle gegossen, die anschließend vorsichtig herausgezogen wurde. Vor Auftrag der Probe muß getestet werden, ob sie eine etwas geringere Dichte als die leichteste, oberste Schicht des Saccharose-Gradienten besitzt. Das Probenvolumen würde ansonsten gleich im Gradienten "versinken", was das angestrebte Trennprinzip zunichte machen würde! Ultrazentrifugation erfolgte ca. 24h bei 40.000Upm in der Beckman L60 Ultrazentrifuge im Ausschwingrotor SW 40 Ti bei 4°C . Abgezogen wurde der Gradient mit der aufgetrennten Probe wie beim Gießen "von unten", indem eine Kanüle vorsichtig auf den Grund des Röhrchens hinabgelassen wurde.

Als Fazit des vorgestellten, geänderten Reinigungsschemas für die RNA-Polymerase aus *P.furiosus* ergibt sich, dass der Ablauf durch den Wegfall der Fällungs- und Dialyseschritte deutlich erleichtert wird. Sofern alle Reinigungsschritte optimal verlaufen, ist die resultierende Präparation der nativen RNAP recht sauber und aktiv (siehe [Abb.4-39]). Allerdings war die Gesamtausbeute geringer als erwartet, was daran liegen könnte, dass vor allem auf der DEAE-Cellulose-Säule (und eventuell auch auf der nachfolgenden Heparin-Sepharose-Säule) Bindungskapazität durch kontaminierende Proteine verloren geht, während bei dem "alten" Reinigungsschema ein Großteil der unerwünschten Proteine schon bei den Fällungsschritten abgetrennt wird. Ein Kompromiss für zukünftige

Reinigungen könnte möglicherweise die Durchführung nur eines statt zwei Fällungsschritten, in Verbindung mit einer weitgehenden Vermeidung von Dialyseschritten in Situationen sein, wo ein Verdünnen der Proben völlig ausreicht.

Folgendes Schema faßt die durchgeführte RNAP-Reinigung zusammen:

Reinigung der DNA-abhängigen RNA-Polymerase aus *Pyrococcus furiosus* (4/2001)

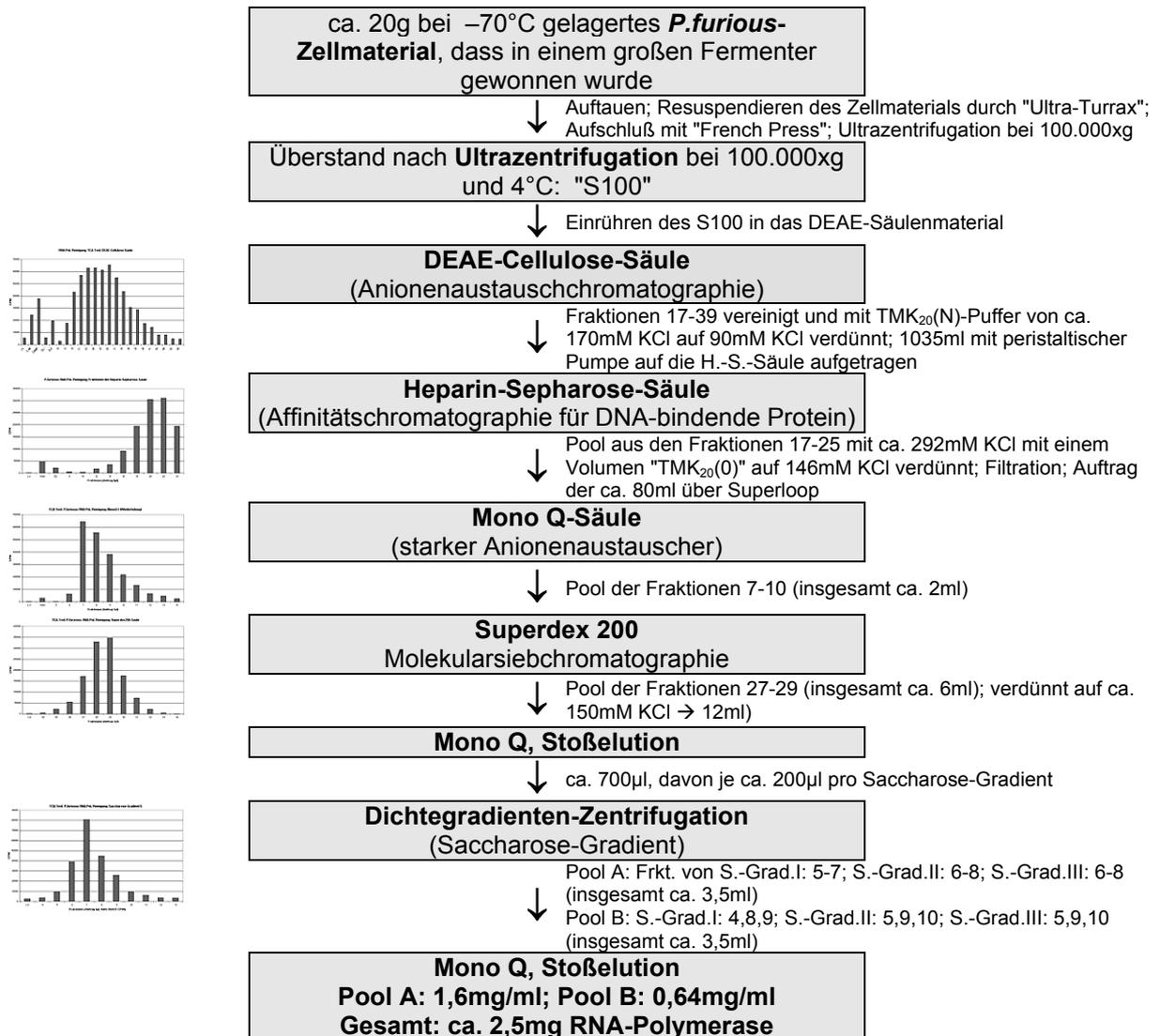


Abb.4-56: Ausführliches Schema der RNAP-Reinigung

Bezeichnungen der Chromatographie-Säulen wie oben beschrieben. Links neben einigen Reinigungsschritten wird die mittels TCA-Test ermittelte enzymatische Aktivität der RNAP in den verschiedenen Fraktionen angedeutet.

Anmerkung: Während der ersten Reinigungsschritte, und besonders bei und direkt nach dem Zellaufschluß durch "Ultra-Turrax" sollte (neben zügigem Arbeiten) besonders auf gleichbleibend niedrige Temperaturen geachtet werden: *P.furiosus* besitzt sehr aktive Proteasen, die auch bei "normalen Temperaturen" eine Gefahr für die übrigen *P.furiosus*-Proteine darstellen (Chang *et al.*, 2001). In der genannten Literatur wird zum Beispiel auch darauf hingewiesen, wie mit dem Proteaseinhibitor AAF-CMK mit Protease I zumindest eine der prominentesten der ca. 11 identifizierten Endoproteinasen spezifisch gehemmt werden kann.

4.12.1 Nachweis von Untereinheit F im Western Blotting Experiment

Für die große Untereinheit A" und die mittelgroßen Untereinheiten D, E' und F wurde die Herstellung von Antikörpern in Auftrag gegeben. Die Immunisierung wurde von der Firma Eurogentec vorgenommen. Bei D, E' und F erfolgte die Immunisierung jeweils eines Kaninchens in

mehreren, zeitlich voneinander getrennten Schritten mit jeweils ca. 100µg Protein in Lösung. Bei A" wurden, da die Präparation noch Nebenbanden zeigte entsprechende Proteinmengen in Form ausgeschnittenen PA-Gel-Stückchen verwendet.

Da die Zuordnung von A", D und E' zu den Banden der nativ gereinigten RNA-Polymerase eindeutig ist, wurde die Verwendung der Antikörper bis zum Auftreten geeigneter Fragestellungen zurückgestellt. In Bezug auf F herrschte hingegen Unsicherheit darüber, ob diese Untereinheit überhaupt permanenter Bestandteil der RNA-Polymerase bzw. der nativen RNAP-Präparation war. Die durchgeführten Far Western Blotting-Experimente zeigten zwar auch ein Signal bei der RNAP, jedoch fand sich in dem aufgrund des berechneten theoretischen Molekulargewicht erwarteten Bereich auf den SDS-Gelen keine nennenswerte Proteinbande. Und ein ähnlich irreguläres Laufverhalten wie bei der stark positiv geladenen Untereinheit P war bei einer Ladung von -8 nicht zu erwarten. Einzig der isoelektrische Punkt ist mit 4,66 der niedrigste der untersuchten RNAP-Untereinheiten. Allerdings besitzen auch D und L einen ähnlich niedrigen pI.

So war das folgende Ergebnis überraschend:

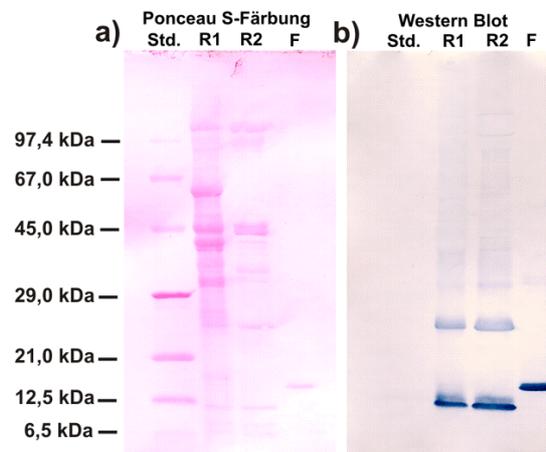


Abb.4-57: Western Blot mit Antikörpern gegen Untereinheit F

Std: Molekulargewichtsstandard. **R1:** RNAP-Präparation (von Kampen, 2003); **R2:** RNAP-Präparation, die in dieser Arbeit vorgenommen wurde (siehe [4.11]). **F:** Untereinheit F (rpoF) mit zusätzlichem Fusionsanteil (N-terminaler His₆-Tag, Molekulargewichtsdifferenz ca. 2 kDa). Denaturierendes PA-Gradientengel (8 – 15% PA, 0,1% SDS); Aufgetragene Proteinmenge bei R1 und R2 ca. 13µg, bei rpoF ca. 200ng. **a)** Ponceau S-Färbung als Vergleich und zur Identifizierung der Banden des Western Blots.

Die Durchführung der Western Blotting-Experimente erfolgte nach den üblichen Protokollen. Durch mehrfache Wiederholung des Versuches wurden die optimalen Mengen für den primären und sekundären Antikörper ermittelt. Der sekundäre (Anti-Rabbit-) Antikörper war mit Peroxidase gekoppelt (POD-Konjugat) und bewirkte in Verbindung mit "BM blue POD Substrat, präzipitierend" (Roche) eine intensiv blaue Farbreaktion, über die die Bindung des sekundären an den primären Antikörper und somit letztlich die Bindung des primären Antikörpers an die Proteine auf der Nitrocellulose-Membran nachgewiesen wurde.

Das Ergebnis dieses Experimentes ist, dass Untereinheit F in der RNAP-Präparation enthalten ist, jedoch überraschend weit unten im Gel läuft. Die beobachteten Unterschiede im Laufverhalten gegenüber dem rekombinanten Protein scheinen auch bei Einbeziehung der Molekulargewichtsdifferenz aufgrund des 20 Aminosäuren langen His₆-Tags recht groß. Vielleicht spielen hier posttranskriptionale Modifikation oder ein alternativer Translationsstart eine Rolle. Im Gegensatz zu Untereinheit E", bei der die Anwesenheit im RNAP-Komplex weiterhin unsicher ist (siehe [4.5]), ist die Situation für rpoF somit geklärt.

Ein interessanter Ansatz zur Erklärung der (deutlich schwächeren) Signale bzw. der Kreuzreaktion mit der E'-Bande wäre der, dass im Kaninchen nicht nur Antikörper gegen rpoF, sondern (nach einiger Zeit) auch Antikörper gegen die Antikörper gebildet worden sind, und so weiter. Diese Idee des "Idiotischen Netzwerkes" ist nicht mehr ganz neu, und wurde teilweise kontrovers diskutiert. Gut läßt sich mit dieser Theorie erklären, wie die Mutter einen Teil ihrer erworbenen Immunität mit

der Muttermilch als Schutz auf das Kind übertragen kann. Außerdem scheint es einen Zusammenhang mit Autoimmun-Prozessen zu geben (Shoenfeld, 2004).

Mit den in dieser Arbeit durchgeführten Experimenten wurde die starke und spezifische Bindung zwischen rpoE' und rpoF nachgewiesen. Möglicherweise ist die Kontaktfläche zwischen den beiden RNAP-Untereinheiten so charakteristisch komplementär gestaltet, dass es zu der geschilderten Antikörperbildung gekommen ist.

4.12.2 Untersuchungen mit MALDI-TOF-MS

MALDI-TOF-MS ("Matrix assisted laser desorption ionisation time-of-flight mass spectrometry") ist heute eine Standardmethode zur Ermittlung des Molekulargewichts von Proteinen. Auch für einzelne Untereinheiten der nativ gereinigten RNA-Polymerase aus *Pyrococcus furiosus* wurden Messungen vorgenommen, die freundlicher Weise von Dr.K.Strupat (Institut für Medizinische Physik und Biophysik, Münster) unter der Leitung von Prof.Dr.F.Hillenkamp durchgeführt wurden. Zur Technik: (Schleuder *et al.*, 1999).

Hauptintention der Anwendung dieser Technik war das Bemühen, eine eindeutige Zuordnung und Identifizierung der "kleinen" Untereinheiten (also F, H, K, L, N, K und eventuell E") zu erlangen.

Bei den großen und mittleren RNAP-Untereinheiten (B, A', A", D und E') war die Zuordnung dagegen schon viel früher klar.

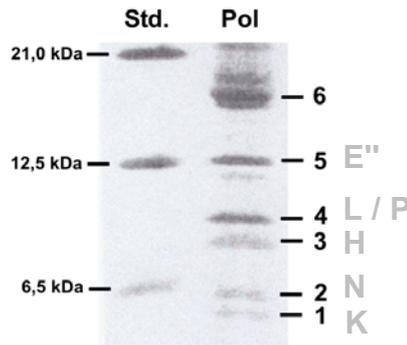


Abb.4-58: Ausschnitt einer PVDF-Membran mit den "geblotteten" kleinen RNAP-Untereinheiten

Es wurden gezielt die Banden 1 – 5 mit MALDI-TOF-MS analysiert. Beginnend mit der am weitesten im denaturierenden PA-Gradientengel gelaufenen Bande, ergaben sich folgende Messwerte: **1:** 6225 Da; **2:** 8040 Da; **3:** 8980 Da; **4:** 11230 Da; **5:** 7150 Da bei einem Massenfehler von +/- 15 Da. Da bei der Elektrophorese ein Tricin-Puffer-System verwendet wurde, weicht das Bandenmuster etwas von dem in [Abb.4-39] gezeigten ab. **6:** wurde vorläufig rpoF zugeordnet (siehe hierzu auch das Western Blotting-Experiment mit Antikörpern gegenüber rpoF im vorangehenden Abschnitt). **Std.:** Banden des Molekulargewichtsstandards; **Pol:** kleine Untereinheiten der nativ aus *P.furiosus* gereinigten RNAP. Graue Buchstaben: Hypothetische Zuordnung von Untereinheiten zu den Banden 1 – 5 aufgrund von MALDI-TOF-MS-Daten.

Es wurden zwei unterschiedliche Ansätze verfolgt: Einerseits die Messung der Präparation der nativen RNAP mit ihrem Gemisch aus 11-12 Untereinheiten, was eine Vielzahl von Signalen lieferte, die nicht alle einwandfrei zugeordnet werden konnten. Andererseits die gezielte Messung der auf der Membran voneinander separierten Proteine.

Da zum jetzigen Zeitpunkt alle RNAP-Gene im Genom von *P.furiosus* identifiziert sind, und sich die theoretischen Molekulargewichte leicht berechnen lassen, kann eine direkte Zuordnung zu den gezeigten Banden beziehungsweise zu den gemessenen Molekulargewichten durchgeführt werden:

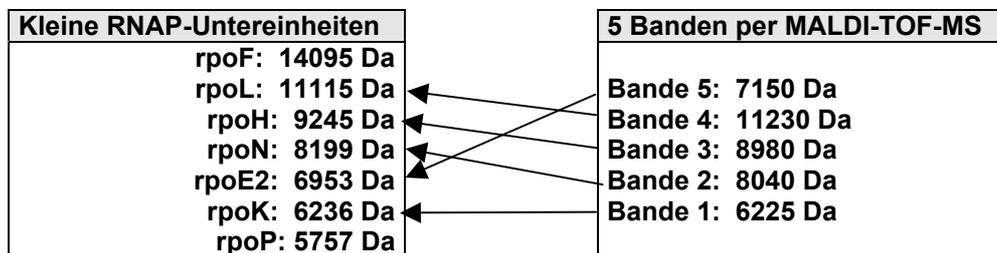


Abb.4-59: Zuordnung von RNAP-Untereinheiten auf der Basis von MALDI-TOF-MS Daten

Wie aus den Daten zu erkennen ist, war rpoF den Erwartungen entsprechend nicht unter den gemessenen Banden. Auch das bekanntermaßen aufgrund der hohen positiven Ladung in SDS-Gelen irregulär langsam laufende rpoP wurde offensichtlich nicht detektiert (eigentlich soll sich mit dem Tricin-Puffersystem gerade bei den kleinen Untereinheiten eine besonders gute Auftrennung ergeben). Da das Vorhandensein von rpoP in der RNAP-Präparation an anderer Stelle eindeutig nachgewiesen wurde, kommt eigentlich nur die Interpretation in Betracht, dass das Protein gemeinsam mit einer anderen Untereinheit in einer der analysierten Banden enthalten ist. Wenn

man entsprechend dieser Vermutung annimmt, dass eine zwei Proteine enthaltende Bande auch optisch stärker hervortritt, kommen hierfür besonders Bande 4 und 5 in Frage. Tatsächlich findet sich im "Schreiberprotokoll" der entsprechenden Messung für Bande 4 neben dem Signal bei ca. 11230 Da ziemlich genau bei 5700 Da ein starkes Signal, welches als der H^{2+} protonierte Zustand interpretiert wurde (im Gegensatz zu H^+ für das 11230 Da-Signal). Falls tatsächlich auch rpoP (mit einem MG von 5757Da, siehe oben) in der Bande enthalten wäre, könnte das Signal bei ca. 5700 Da also zusätzlich den einfach protonierten Zustand von rpoP enthalten.

Für Bande 5 kommt eigentlich nur rpoE2 in Frage, was besonders interessant ist, weil das Vorhandensein von rpoE2 in der RNAP-Präparation von *Pyrococcus furiosus* noch nicht mit Sicherheit nachgewiesen wurde.

Da allerdings die Abweichungen der theoretischen und praktische gemessenen Molekulargewichten größer als erwartet sind (rpoK/Bande1: -11 Da; rpoN/Bande2: -159 Da; rpoH/Bande3: -265 Da; rpoL/Bande4: +115 Da; rpoE2/Bande5: +197 Da), werden die Daten zunächst vorsichtig gehandhabt (nur der Messwert für rpoK/Bande1 liegt innerhalb des angegebenen Fehlers von +/- 15 Da). Als Erklärung für die beobachteten Abweichen kommen posttranslationale Modifikationen in Betracht.

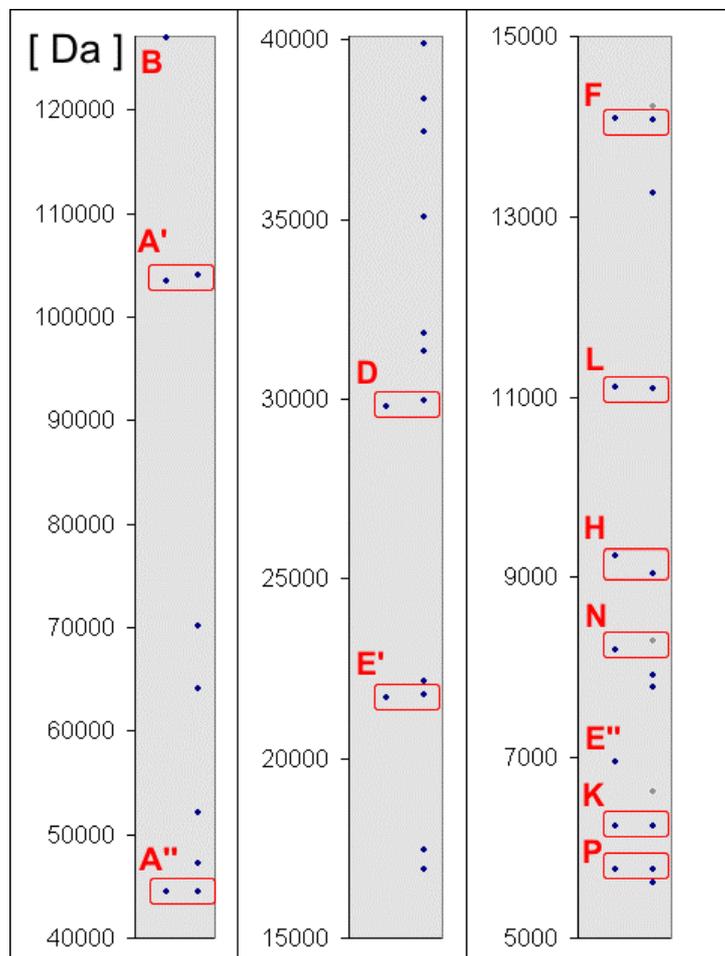


Abb.4-60: Auswertung der MALDI-TOF-MS-Daten von einem Gemisch aus RNAP-Untereinheiten

Korrelation von theoretischem Molekulargewicht der (bekannten) RNAP-Untereinheiten und den Signalen, die mittels MALDI-TOF-MS aus einer Präparation der nativen RNAP (samt allen darin enthaltenen Untereinheiten) gemessen wurden. Das Molekulargewichtsspektrum (Einheit Dalton) wurde in 3 fortlaufende, jedoch für jeden Auflösungsbereich passend skalierte Abschnitte aufgeteilt. Den 12 Untereinheiten stehen 29 Messwerte gegenüber, von den 3 als "winzig" eingestuft wurden. Diese Punkte wurden grau eingefärbt. Eines dieser schwachen Signale paßt am besten zu rpoN. Ansonsten läßt sich allen Untereinheiten bis auf E" (rpoE2) ein Molekül ähnlicher Größe zuordnen. Dass sich hier kein Signal für E" findet, steht im Widerspruch zu den oben gezeigten MALDI-TOF-MS-Messungen und verlangt weitere Untersuchungen.

5 Zusammenfassung

Pyrococcus furiosus gehört zur Domäne der *Archaea* (*Euryarchaeota*), früher auch "Archaeobacteria", und ist ein hyperthermophiler Mikroorganismus, dessen DNA-abhängige RNA-Polymerase sich aus 11-12 Untereinheiten zusammensetzt (B, A', A'', D, E', E'', F, H, K, L, N, P). Die Gene dieser Untereinheiten wurden identifiziert und bis auf die größte Untereinheit B alle kloniert, überexprimiert und gereinigt, wobei vor allem die großen Untereinheiten schlecht löslich sind. Die Untereinheit A' wurde unter denaturierenden Bedingungen aus Einschlusskörpern gereinigt. Bei den kleineren Untereinheiten (D, E', F, H, K, L, N, P) sowie den Transkriptionsfaktoren TBP und TFB erfolgte die Klonierung zusätzlich in den Vektor pET-33b(+) (Novagen, Inc.), der über die im Fusionsanteil des resultierenden Proteins enthaltene Erkennungssequenz Arg-Arg-Ala-Ser-Val eine spezifische, radioaktive Markierung mit Proteinkinase A ermöglicht.

Anschließend konnten diese rekombinanten Proteine als Sonden bei "Far Western Blotting"-Experimenten eingesetzt werden, bei denen im Gegensatz zum Western Blot statt eines Antikörpers direkt das markierte Protein zum Nachweis von spezifischen Protein-Protein-Wechselwirkungen verwendet wird.

Spezifische Interaktionen der Untereinheit D mit L, N, P und der großen Untereinheit B konnten gezeigt werden. Besonders starke Interaktionen zeigten sich für E'+F, A''+H, D+L, D+P, D+B. Die Wechselwirkungen der Untereinheit K mit TFB war schwächer als entsprechend der Literatur zu erwarten gewesen wäre. Ein Versuch mit teildeletierten TFB-Varianten lieferte einen Hinweis auf eine stärkere Bindung von Untereinheit K an eine N- und C-terminal leicht deletierte TFB-Form.

Bei Polyacrylamid-Gelelektrophorese unter nativen Bedingungen konnte die Bildung eines D-L-, D-N- und E'-F-Komplexes nachgewiesen werden.

Schließlich konnte mittels Molekularsiebchromatographie ein gut löslicher D+L-Komplex und ein mutmaßlicher D-L-N-Komplex gezeigt werden. Die letztgenannten Experimente können als Vorversuche für das "Assembly" der RNA-Polymerase aus rekombinanten Untereinheiten verstanden werden. In der Arbeitsgruppe von Prof.Dr.M.Thomm werden gegenwärtig an der Universität Regensburg im Rahmen der Weiterführung des von mir bearbeiteten Themas bereits weitere Experimente in dieser Richtung unternommen.

Neben den praktischen Experimenten, erfolgte eine intensive Analyse der Gene der RNA-Polymerase-Untereinheiten mit bioinformatischen Methoden, wobei auch ein selbstgeschriebenes Programm zur Extraktion von Daten aus Genom-Dateien, zur Protein-Analyse und zum Erstellen von Diagrammen zum Einsatz kam. Schwerpunkt war eine Bestandsaufnahme aller RNA-Pol-Gene aus den gegenwärtig 17 komplett sequenzierten *Archaea*. Resultierende Alignments dienen als Grundlage zur Visualisierung von gegenüber *Archaea* konservierten Regionen in dem seit kurzem verfügbaren 3D-Modell der eukaryotischen, aus 12 Untereinheiten bestehenden RNA-Polymerase von *Saccharomyces cerevisiae*. Es wurde eine detaillierte Gegenüberstellung der bei *Archaea* bzw. *Pyrococcus furiosus* gegenüber *S.cerevisiae* ähnlichen und abweichenden Regionen innerhalb der einzelnen RNA-Pol-Gene erarbeitet.

Eine Untersuchung der Verteilungshäufigkeit bestimmter Aminosäuren in RNA-Polymerase-Genen sowie in allen Proteingenen der untersuchten Organismen lieferte interessante Details zum Thema "Thermoadaptation".

6 Summary

Pyrococcus furiosus belongs to the domain *Archaea* (*Euryarchaeota*), formerly known as *Archaeobacteria* and is a hyperthermophilic microorganism with an optimal growth temperature of 100°C. It contains a DNA dependent RNA polymerase (RNAP) out of 11-12 different subunits (B, A', A'', D, E', E'', F, H, K, L, N, P). All RNAP genes except for the largest subunit B were cloned in *E.coli*, overexpressed and purified. Above all, the large subunits showed limited solubility. Subunit A' was purified from inclusion bodies under denaturing conditions.

Medium sized and small subunits (D, E', F, H, K, L, N, P) and the general transcription factors TFB and TBP were subjected to additional cloning into the vector pET-33b(+) (Novagen Inc.). The recombinant proteins can be specifically labeled with radioactivity via Protein kinase A catalytic subunit from heart muscle (HMK) because of the included 5-amino acid HMK recognition site ArgArgAlaSerVal.

Subsequently all ³²P-labeled proteins were used as probes in Far western blotting analysis to detect direct protein–protein contacts without additional need of antibodies.

Specific interaction of subunit D with the small subunits L, N, P and the largest subunit B were shown. Particularly strong interactions became visible between subunits E'-F, A''-H, D-L, D-P and D-B. Binding of subunit K to TFB was less intense as expected from literature. A slightly both N- and C-terminal truncated form of TFB showed increased affinity to subunit K compared with the wild type TFB.

Finally, a readily soluble D-L complex and a putative D-L-N complex were detected in size exclusion chromatography experiments. The last-mentioned approach can act as a preliminary experiment for the reconstitution of the *P.furiosus* RNA polymerase from recombinant subunits.

Already now further experiments with additional subunits are carried out in the workgroup of Prof.Dr.M. Thomm at the University of Regensburg, Germany.

Apart from practical experiments a detailed analysis of all RNAP subunit genes by bioinformatic means was done. A computer program was written to facilitate the extraction and the analysis of the genes from the available microbial genomes, to calculate certain informations about proteins and to generate space-saving diagrams. The main focus was a stock-taking of all RNAP subunit genes in the currently accessible 17 completed archaeal genomes.

An archaeal consensus sequence was calculated for each subunit via extensive alignments and visualized in the 12 subunit 3D model of the eukaryotic RNAP of *S.cerevisiae*. A detailed comparison of sequence similarities of the RNAP subunits in *Archaea* / *P.furiosus* and *S.cerevisiae* was worked out and discussed.

The analysis of amino acid frequencies within the RNAP genes and all protein genes in the genomes of different microorganisms provides data about amino acid replacement and its importance to thermoadaptation of proteins.

7 Literaturverzeichnis

- Aggarwal,G. & Ramaswamy,R. (2002). Ab initio gene identification: prokaryote genome annotation with GeneScan and GLIMMER. *J.Biosci.*, **27**: 7-14.
- Armache,K.J., Kettenberger,H. & Cramer,P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc.Natl.Acad.Sci.U.S.A.*, **100**: 6964-6968.
- Baliga,N.S., Goo,Y.A., Ng,W.V., Hood,L., Daniels,C.J. & DasSarma,S. (2000). Is gene expression in Halobacterium NRC-1 regulated by multiple TBP and TFB transcription factors? *Mol.Microbiol.*, **36**: 1184-1185.
- Bartlett,M.S., Thomm,M. & Geiduschek,E.P. (2000). The orientation of DNA in an archaeal transcription initiation complex. *Nat.Struct.Biol.*, **7**: 782-785.
- Bartlett,M.S., Thomm,M. & Geiduschek,E.P. (2003). Topography of the euryarchaeal transcription initiation complex. *J.Biol.Chem.*.
- Bell,S.D., Brinkman,A.B., Van der,O.J. & Jackson,S.P. (2001). The archaeal TFIIIEalpha homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep.*, **2**: 133-138.
- Blöchl,E., Burggraf,S., Fiala,G., Lauerer,G., Huber,G., Huber,R., Rachel,R., Segerer,A., Stetter,K.O. & Völkl,P. (1995). Isolation, taxonomy and phylogeny of hyperthermophilic microorganisms. *World Journal of Microbiology & Biotechnology*, **11**: 9-16.
- Breathnach,R. & Chambon,P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annu.Rev.Biochem.*, **50**: 349-383.
- Brinkmann,U., Mattes,R.E. & Buckel,P. (1989). High-level expression of recombinant genes in Escherichia coli is dependent on the availability of the dnaY gene product. *Gene*, **85**: 109-114.
- Bucher,P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J.Mol.Biol.*, **212**: 563-578.
- 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., Dougherty,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. & Venter,J.C. (1996). Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science*, **273**: 1058-1073.
- Burgess,R.R., Arthur,T.M. & Pietz,B.C. (2000). Mapping protein-protein interaction domains using ordered fragment ladder far-western analysis of hexahistidine-tagged fusion proteins. *Methods Enzymol.*, **328**: 141-157.
- Chang,L.S., Hicks,P.M. & Kelly,R.M. (2001). Protease I from Pyrococcus furiosus. *Methods Enzymol.*, **330**: 403-413.
- Chou,P.Y. & Fasman,G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv.Enzymol.Relat Areas Mol.Biol.*, **47**: 45-148.
- Cohen,G.N., Barbe,V., Flament,D., Galperin,M., Heilig,R., Lecompte,O., Poch,O., Prieur,D., Querellou,J., Ripp,R., Thierry,J.C., Van der,O.J., Weissenbach,J., Zivanovic,Y. & Forterre,P.

- (2003). An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol. Microbiol.*, **47**: 1495-1512.
- Cramer, P. (2002). Multisubunit RNA polymerases. *Curr. Opin. Struct. Biol.*, **12**: 89-97.
- Cramer, P., Bushnell, D.A., Fu, J., Gnat, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R. & Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*, **288**: 640-649.
- Cramer, P., Bushnell, D.A. & Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, **292**: 1863-1876.
- Dahlke, I. & Thomm, M. (2002). A *Pyrococcus* homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment. *Nucleic Acids Res.*, **30**: 701-710.
- Darcy, T.J., Hausner, W., Awery, D.E., Edwards, A.M., Thomm, M. & Reeve, J.N. (1999). *Methanobacterium thermoautotrophicum* RNA polymerase and transcription in vitro. *J. Bacteriol.*, **181**: 4424-4429.
- Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wiezer, A., Baumer, S. & Jacobi, C. (2002). The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. **4**: 453-461.
- Di Giulio, M. (2003). The ancestor of the Bacteria domain was a hyperthermophile. *J. Theor. Biol.*, **224**: 277-283.
- Dieci, G., Hermann-Le Denmat, S., Lukhtanov, E., Thuriaux, P., Werner, M. & Sentenac, A. (1995). A universally conserved region of the largest subunit participates in the active site of RNA polymerase III. *EMBO J.*, **14**: 3766-3776.
- Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. & Weissmann, C. (1983). Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. *Cell*, **32**: 695-706.
- Ehlers, C., Grabbe, R., Veit, K. & Schmitz, R.A. (2002). Characterization of GlnK1 from *Methanosarcina mazei* strain Go1: complementation of an *Escherichia coli* glnK mutant strain by GlnK1. *J. Bacteriol.*, **184**: 1028-1040.
- Eloranta, J.J., Kato, A., Teng, M.S. & Weinzierl, R.O. (1998). In vitro assembly of an archaeal D-L-N RNA polymerase subunit complex reveals a eukaryote-like structural arrangement. *Nucleic Acids Res.*, **26**: 5562-5567.
- Erauso, G., Reysenbach, A.L., Godfroy, A., Meunier, J.R., Crump, B., Partensky, F., Baross, J.A., Marteinsson, V., Barbier, G., Pace, N.R. & et al. (1993). *Pyrococcus abyssi* sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch. Microbiol.*, **160**: 338-349.
- Eskin, E., Keich, U., Gelfand, M.S. & Pevzner, P.A. (2003). Genome-wide analysis of bacterial promoter regions. *Pac. Symp. Biocomput.*, 29-40.
- Fiala, G. & Stetter, K.O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Arch. Microbiol.*, **145**: 56-61.

- Fitz-Gibbon,S.T., Ladner,H., Kim,U.J., Stetter,K.O., Simon,M.I. & Miller,J.H. (2002). Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Proc.Natl.Acad.Sci.U.S.A.*, **99**: 984-989.
- Flores,O., Lu,H. & Reinberg,D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIIH. *J.Biol.Chem.*, **267**: 2786-2793.
- Galagan,J.E., Nusbaum,C., Roy,A., Endrizzi,M.G., Macdonald,P., FitzHugh,W., Calvo,S., Engels,R., Smirnov,S., Atnoor,D., Brown,A., Allen,N., Naylor,J., Stange-Thomann,N., DeArellano,K., Johnson,R., Linton,L., McEwan,P., McKernan,K., Talamas,J., Tirrell,A., Ye,W., Zimmer,A., Barber,R.D., Cann,I., Graham,D.E., Grahame,D.A., Guss,A.M., Hedderich,R., Ingram-Smith,C., Kuettner,H.C., Krzycki,J.A., Leigh,J.A., Li,W., Liu,J., Mukhopadhyay,B., Reeve,J.N., Smith,K., Springer,T.A., Umayam,L.A., White,O., White,R.H., Conway,d.M., Ferry,J.G., Jarrell,K.F., Jing,H., Macario,A.J., Paulsen,I., Pritchett,M., Sowers,K.R., Swanson,R.V., Zinder,S.H., Lander,E., Metcalf,W.W. & Birren,B. (2002). The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.*, **12**: 532-542.
- Garnier,J., Osguthorpe,D.J. & Robson,B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J.Mol.Biol.*, **120**: 97-120.
- Ghosh,P., Ishihama,A. & Chatterji,D. (2001). Escherichia coli RNA polymerase subunit omega and its N-terminal domain bind full-length beta' to facilitate incorporation into the alpha2beta subassembly. *Eur.J.Biochem.*, **268**: 4621-4627.
- Gnatt,A.L., Cramer,P., Fu,J., Bushnell,D.A. & Kornberg,R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science*, **292**: 1876-1882.
- Goede,B. (1998). Untersuchungen zur Funktion des archaeellen Transkriptionsfaktors TFB aus *Pyrococcus furiosus*. *Diplomarbeit, Universität Kiel*.
- Gonzalez,J.M., Masuchi,Y., Robb,F.T., Ammerman,J.W., Maeder,D.L., Yanagibayashi,M., Tamaoka,J. & Kato,C. (1998). *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. *Extremophiles.*, **2**: 123-130.
- Haney,P.J., Badger,J.H., Buldak,G.L., Reich,C.I., Woese,C.R. & Olsen,G.J. (1999). Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc.Natl.Acad.Sci.U.S.A.*, **96**: 3578-3583.
- Hanzelka,B.L., Darcy,T.J. & Reeve,J.N. (2001). TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIEalpha. *J.Bacteriol.*, **183**: 1813-1818.
- Hausner,W., Lange,U. & Musfeldt,M. (2000). Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J.Biol.Chem.*, **275**: 12393-12399.
- Hausner,W. & Thomm,M. (1993). Purification and characterization of a general transcription factor, aTFB, from the archaeon *Methanococcus thermolithotrophicus*. *J.Biol.Chem.*, **268**: 24047-24052.
- Hausner,W. & Thomm,M. (2001). Events during initiation of archaeal transcription: open complex formation and DNA-protein interactions. *J.Bacteriol.*, **183**: 3025-3031.

- Hausner,W., Wettach,J., Hethke,C. & Thomm,M. (1996). Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. *J.Biol.Chem.*, **271**: 30144-30148.
- Helmann,J.D. & Chamberlin,M.J. (1988). Structure and function of bacterial sigma factors. *Annu.Rev.Biochem.*, **57**: 839-872.
- Hemming,S.A., Jansma,D.B., Macgregor,P.F., Goryachev,A., Friesen,J.D. & Edwards,A.M. (2000). RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo. *J.Biol.Chem.*, **275**: 35506-35511.
- Hethke,C., Bergerat,A., Hausner,W., Forterre,P. & Thomm,M. (1999). Cell-free transcription at 95 degrees: thermostability of transcriptional components and DNA topology requirements of *Pyrococcus* transcription. *Genetics*, **152**: 1325-1333.
- Hethke,C., Geerling,A.C., Hausner,W., de Vos,W.M. & Thomm,M. (1996). A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. *Nucleic Acids Res.*, **24**: 2369-2376.
- Higgins,D.G., Thompson,J.D. & Gibson,T.J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.*, **266**: 383-402.
- Hull,M.W., McKune,K. & Woychik,N.A. (1995). RNA polymerase II subunit RPB9 is required for accurate start site selection. *Genes Dev.*, **9**: 481-490.
- Islas,S., Velasco,A.M., Becerra,A., Delaye,L. & Lazcano,A. (2003). Hyperthermophily and the origin and earliest evolution of life. *Int.Microbiol.*, **6**: 87-94.
- Jaenicke,R. (2000). Do ultrastable proteins from hyperthermophiles have high or low conformational rigidity? *Proc.Natl.Acad.Sci.U.S.A*, **97**: 2962-2964.
- Johnsen,U., Hansen,T. & Schonheit,P. (2003). Comparative analysis of pyruvate kinases from the hyperthermophilic archaea *Archaeoglobus fulgidus*, *Aeropyrum pernix*, and *Pyrobaculum aerophilum* and the hyperthermophilic bacterium *Thermotoga maritima*: unusual regulatory properties in hyperthermophilic archaea. *J.Biol.Chem.*, **278**: 25417-25427.
- Kawarabayasi,Y., Hino,Y., Horikawa,H., Jin-no,K., Takahashi,M., Sekine,M., Baba,S., Ankai,A., Kosugi,H., Hosoyama,A., Fukui,S., Nagai,Y., Nishijima,K., Otsuka,R., Nakazawa,H., Takamiya,M., Kato,Y., Yoshizawa,T., Tanaka,T., Kudoh,Y., Yamazaki,J., Kushida,N., Oguchi,A., Aoki,K., Masuda,S., Yanagii,M., Nishimura,M., Yamagishi,A., Oshima,T. & Kikuchi,H. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. *DNA Res.*, **8**: 123-140.
- Kawarabayasi,Y., Hino,Y., Horikawa,H., Yamazaki,S., Haikawa,Y., Jin-no,K., Takahashi,M., Sekine,M., Baba,S., Ankai,A., Kosugi,H., Hosoyama,A., Fukui,S., Nagai,Y., Nishijima,K., Nakazawa,H., Takamiya,M., Masuda,S., Funahashi,T., Tanaka,T., Kudoh,Y., Yamazaki,J., Kushida,N., Oguchi,A., Kikuchi,H. & . (1999). Complete genome sequence of an aerobic hyperthermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.*, **6**: 83-52.
- Kawarabayasi,Y., Sawada,M., Horikawa,H., Haikawa,Y., Hino,Y., Yamamoto,S., Sekine,M., Baba,S., Kosugi,H., Hosoyama,A., Nagai,Y., Sakai,M., Ogura,K., Otsuka,R., Nakazawa,H., Takamiya,M., Ohfuku,Y., Funahashi,T., Tanaka,T., Kudoh,Y., Yamazaki,J., Kushida,N., Oguchi,A., Aoki,K. & Kikuchi,H. (1998). Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.*, **5**: 55-76.

- Kawashima,T., Amano,N., Koike,H., Makino,S., Higuchi,S., Kawashima-Ohya,Y., Watanabe,K., Yamazaki,M., Kanehori,K., Kawamoto,T., Nunoshiba,T., Yamamoto,Y., Aramaki,H., Makino,K. & Suzuki,M. (2000). Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proc.Natl.Acad.Sci.U.S.A.*, **97**: 14257-14262.
- Klug,A. (2001). Structural biology. A marvellous machine for making messages. *Science*, **292**: 1844-1846.
- Kreidberg,J.A. & Kelly,T.J. (1986). Genetic analysis of the human thymidine kinase gene promoter. *Mol.Cell Biol.*, **6**: 2903-2909.
- Kyngas,J. & Valjakka,J. (1998). Unreliability of the Chou-Fasman parameters in predicting protein secondary structure. *Protein Eng*, **11**: 345-348.
- Kyrpides,N.C. & Ouzounis,C.A. (1999). Transcription in archaea. *Proc.Natl.Acad.Sci.U.S.A.*, **96**: 8545-8550.
- Landick,R. (2001). RNA polymerase clamps down. *Cell*, **105**: 567-570.
- Lange,A. (2001). Untersuchungen zur Funktion der Untereinheiten der DNA-abhängigen RNA-Polymerase von *Pyrococcus furiosus*. *Diplomarbeit, Universität Kiel*.
- Lange,U. (2003). Charakterisierung des archaeellen Elongationsfaktors TFS und Studien zur Elongation der Transkription in *Methanococcus*. *Doktorarbeit, Universität Kiel*.
- Langer,D., Hain,J., Thuriaux,P. & Zillig,W. (1995). Transcription in archaea: similarity to that in eucarya. *Proc.Natl.Acad.Sci.U.S.A.*, **92**: 5768-5772.
- Lanzendörfer,M., Langer,D., Hain,J., Klenk,H.-P., Holz,I., Arnold-Ammer,I. & Zillig,W. (1994). Structure and Function of the DNA-Dependent RNA Polymerase of *Sulfolobus*. *System.Appl.Microbiol.*, **16**: 656-664.
- Lecompte,O., Ripp,R., Puzos-Barbe,V., Duprat,S., Heilig,R., Dietrich,J., Thierry,J.C. & Poch,O. (2001). Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea. *Genome Res.*, **11**: 981-993.
- Littlefield,O., Korkhin,Y. & Sigler,P.B. (1999). The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc.Natl.Acad.Sci.U.S.A.*, **96**: 13668-13673.
- Lukashin,A.V. & Borodovsky,M. (1998). GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.*, **26**: 1107-1115.
- Maeder,D.L., Weiss,R.B., Dunn,D.M., Cherry,J.L., Gonzalez,J.M., DiRuggiero,J. & Robb,F.T. (1999). Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics*, **152**: 1299-1305.
- Magill,C.P., Jackson,S.P. & Bell,S.D. (2001). Identification of a conserved archaeal RNA polymerase subunit contacted by the basal transcription factor TFB. *J.Biol.Chem.*, **276**: 46693-46696.
- Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M., Shukla,H.D., Lasky,S.R., Baliga,N.S., Thorsson,V., Sbrogna,J., Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A., Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W., Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H., Isenbarger,T.A., Peck,R.F., Pohlschroder,M., Spudich,J.L., Jung,K.W., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P., Omer,A.D., Ebhardt,H.,

- Lowe, T.M., Liang, P., Riley, M., Hood, L. & DasSarma, S. (2000). Genome sequence of Halobacterium species NRC-1. *Proc. Natl. Acad. Sci. U.S.A.*, **97**: 12176-12181.
- Nikolov, D.B. & Burley, S.K. (1997). RNA polymerase II transcription initiation: a structural view. *Proc. Natl. Acad. Sci. U.S.A.*, **94**: 15-22.
- Orphanides, G., Lagrange, T. & Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev.*, **10**: 2657-2683.
- Ouhammouch, M., Dewhurst, R.E., Hausner, W., Thomm, M. & Geiduschek, E.P. (2003). Activation of archaeal transcription by recruitment of the TATA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.*, **100**: 5097-5102.
- Page, R.D. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.*, **12**: 357-358.
- Pribnow, D. (1975). Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc. Natl. Acad. Sci. U.S.A.*, **72**: 784-788.
- Qureshi, S.A., Bell, S.D. & Jackson, S.P. (1997). Factor requirements for transcription in the Archaeon *Sulfolobus shibatae*. *EMBO J.*, **16**: 2927-2936.
- Qureshi, S.A. & Jackson, S.P. (1998). Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol. Cell*, **1**: 389-400.
- Reeve, J.N., Sandman, K. & Daniels, C.J. (1997). Archaeal histones, nucleosomes, and transcription initiation. *Cell*, **89**: 999-1002.
- Robb, F.T., Maeder, D.L., Brown, J.R., DiRuggiero, J., Stump, M.D., Yeh, R.K., Weiss, R.B. & Dunn, D.M. (2001). Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. *Methods Enzymol.*, **330**: 134-157.
- Roberts, S.G. & Green, M.R. (1994). Activator-induced conformational change in general transcription factor TFIIB. *Nature*, **371**: 717-720.
- Rother, M., Resch, A., Wilting, R. & Bock, A. (2001). Selenoprotein synthesis in archaea. *Biofactors*, **14**: 75-83.
- Ruepp, A., Graml, W., Santos-Martinez, M.L., Koretke, K.K., Volker, C., Mewes, H.W., Frishman, D., Stocker, S., Lupas, A.N. & Baumeister, W. (2000). The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature*, **407**: 508-513.
- Schaller, H., Gray, C. & Herrmann, K. (1975). Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. *Proc. Natl. Acad. Sci. U.S.A.*, **72**: 737-741.
- Schleuder, D., Hillenkamp, F. & Strupat, K. (1999). IR-MALDI-mass analysis of electroblotted proteins directly from the membrane: comparison of different membranes, application to on-membrane digestion, and protein identification by database searching. *Anal. Chem.*, **71**: 3238-3247.
- Schmeling, G. (2000). Expression und Reinigung des Transkriptionsfaktors TFIIS von *Saccharomyces cerevisiae*, sowie der RNA-Polymerase von *Methanococcus thermolithotrophicus* und *Methanococcus jannaschii*. *GMP II Praktikum am Ifam, Universität Kiel*.
- Severinov, K., Mustaev, A., Kukarin, A., Muzzin, O., Bass, I., Darst, S.A. & Goldfarb, A. (1996). Structural modules of the large subunits of RNA polymerase. Introducing archaeobacterial and

- chloroplast split sites in the beta and beta' subunits of Escherichia coli RNA polymerase. *J.Biol.Chem.*, **271**: 27969-27974.
- Shoenfeld, Y. (2004). The idiotypic network in autoimmunity: antibodies that bind antibodies that bind antibodies. *Nat.Med.*, **10**: 17-18.
- Slesarev, A.I., Mezhevaya, K.V., Makarova, K.S., Polushin, N.N., Shcherbinina, O.V., Shakhova, V.V., Belova, G.I., Aravind, L., Natale, D.A., Rogozin, I.B., Tatusov, R.L., Wolf, Y.I., Stetter, K.O., Malykh, A.G., Koonin, E.V. & Kozyavkin, S.A. (2002). The complete genome of hyperthermophile Methanopyrus kandleri AV19 and monophyly of archaeal methanogens. *Proc.Natl.Acad.Sci.U.S.A.*, **99**: 4644-4649.
- 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.
- Sonntag, K.C. & Darai, G. (1995). Evolution of viral DNA-dependent RNA polymerases. *Virus Genes*, **11**: 271-284.
- Soppa, J. (1999). Transcription initiation in Archaea: facts, factors and future aspects. *Mol.Microbiol.*, **31**: 1295-1305.
- Spitalny, P. & Thomm, M. (2003). Analysis of the open region and of DNA-protein contacts of archaeal RNA polymerase transcription complexes during transition from initiation to elongation. *J.Biol.Chem.*, **278**: 30497-30505.
- Sterner, R. & Liebl, W. (2001). Thermophilic adaptation of proteins. *Crit Rev.Biochem.Mol.Biol.*, **36**: 39-106.
- Thomm, M. (1996). Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol.Rev.*, **18**: 159-171.
- Thomm, M. (2000). Die Transkriptionsmaschinerie der Archaea. *Biospektrum*, **5. 0 0, 6.Jahrgang**: 179-185.
- Todone, F., Brick, P., Werner, F., Weinzierl, R.O. & Onesti, S. (2001). Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol.Cell*, **8**: 1137-1143.
- Tsai, F.T. & Sigler, P.B. (2000). Structural basis of preinitiation complex assembly on human pol II promoters. *EMBO J.*, **19**: 25-36.
- Vierke, G., Engelmann, A., Hebbeln, C. & Thomm, M. (2003). A novel archaeal transcriptional regulator of heat shock response. *J.Biol.Chem.*, **278**: 18-26.
- von Kampen, O. (2003). Analyse von Protein-Protein-Wechselwirkungen von Komponenten des Transkriptionsapparates von *Pyrococcus furiosus*. *Diplomarbeit, Universität Kiel*.
- Voss, E. (2003). Expression und Charakterisierung von regulatorischen Proteinen aus *Pyrococcus*. *Diplomarbeit, Universität Kiel*.
- Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G.G., Simon, M., Soll, D., Stetter, K.O., Short, J.M. & Noordewier, M. (2003). The genome of

- Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. *Proc.Natl.Acad.Sci.U.S.A*, **100**: 12984-12988.
- Werner,F., Eloranta,J.J. & Weinzierl,R.O. (2000). Archaeal RNA polymerase subunits F and P are bona fide homologs of eukaryotic RPB4 and RPB12. *Nucleic Acids Res.*, **28**: 4299-4305.
- Werner,F. & Weinzierl,R.O. (2002). A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol.Cell*, **10**: 635-646.
- Wilting,R., Schorling,S., Persson,B.C. & Bock,A. (1997). Selenoprotein synthesis in archaea: identification of an mRNA element of Methanococcus jannaschii probably directing selenocysteine insertion. *J.Mol.Biol.*, **266**: 637-641.
- Woese,C.R. & Fox,G.E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdom. *Proc.Natl.Acad.Sci.*, **74**: 5088-5090.
- Woese,C.R., Kandler,O. & Wheelis,M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc.Natl.Acad.Sci.U.S.A*, **87**: 4576-4579.
- Zawel,L. & Reinberg,D. (1993). Initiation of transcription by RNA polymerase II: a multi-step process. *Prog.Nucleic Acid Res.Mol.Biol.*, **44**: 67-108.
- Zaychikov,E., Martin,E., Denissova,L., Kozlov,M., Markovtsov,V., Kashlev,M., Heumann,H., Nikiforov,V., Goldfarb,A. & Mustaev,A. (1996). Mapping of catalytic residues in the RNA polymerase active center. *Science*, **273**: 107-109.
- Zhu,W., Zeng,Q., Colangelo,C.M., Lewis,M., Summers,M.F. & Scott,R.A. (1996). The N-terminal domain of TFIIB from Pyrococcus furiosus forms a zinc ribbon. *Nat.Struct.Biol.*, **3**: 122-124.
- Zillig,W., Stetter,K.O., Wunderl,S., Schulz,W., Priess,H. & Scholz,I. (1980). The Sulfolobus-'Caldariella' group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Arch.Microbiol.*, **125**: 259-269.

8 Anhang

8.1 Multiple Alignments

8.1.1 Vorbemerkung und Zusatzinformationen zu den Multiplen Alignments

Die hier gezeigten Alignments wurden zum Zweck größerer Übersichtlichkeit vom Ergebnisteil in den Anhang verlagert. Vor allem die „großen“ Alignments mit 17-18 *Archaea* und *S.cerevisiae* ziehen sich über viele Seiten hin. Um bestimmte Aspekte und Resultate aus diesen Sequenzvergleichen intuitiv erfassen und betonen zu können, werden im Ergebnisteil stattdessen (mit Borland Delphi 5) selbst programmierte Grafiken gezeigt. Die Details dieser Grafiken wären ohne die kompletten Alignments nicht nachvollziehbar.

Obwohl die drei *Pyrococcus*-Arten natürlich auch in den kompletten Alignments enthalten sind, wird hier zusätzlich der „kleine“ Vergleich innerhalb der Gattung *Pyrococcus* gezeigt, weil sich damit einige Details diskutieren lassen, die im großen Alignment nur schwer erkennbar wären.

Durchgängig wurden konservierte Regionen hell, ähnliche Bereiche grau, und stark abweichende Aminosäure-Positionen schwarz hinterlegt mit weißer Schrift dargestellt. Die Berechnung der Alignments erfolgte mit „Malign“ in Verbindung mit Malign2Msf (beides „Biounit“ des dkfz-Heidelberg; „HUSAR“ [3.8]). Für die Erzeugung der Konsensus-Sequenz und grafische Darstellung wurde das Programm „Bioedit 5.0.9“ (siehe auch [3.8]) verwendet. Nach Export als RTF-Datei wurden letzte manuelle Änderungen in Word vorgenommen (z.B. wurden durch überstehende, nicht homologe Enden verursachte „massive schwarze Flächen“ sowie schwarze Lücken entfernt).

8.1.2 Alignments der drei *Pyrococcus*-Arten untereinander

8.1.2.1 Untereinheit B (rpoB)

P.furiosus PF1564, *P.abyssi* rpoB / PAB0423, *P.horikoshii* PH1546

```

Pfu_rpoB MRGPTVVDVTPDDLWIVMESYWQEKGLVRQHLDSYNAFIERGLQEVVNEFGGKIPDIPDFEVKFGKIRVGEPEFQEPHGQ 80
Pab_rpoB MRGPTVVEVTPDDLWIVMESYWQEKGLVRQHLDSYNAFIDRGLQEVVDEFGGKIPDIPDFEVKFGKIRVGEPEFQEPHGQ 80
Pho_rpoB MRGPTVVEVTPDDLWIVMESYWEKGLVRQHLDSYNAFIERGLQEVVDEFGGKIPDIPDFEVKFGKIRVGEPEFQEPHGQ 80

Pfu_rpoB RKPLYPMDARIRNLTYSAPLYLEMIPIVIGKVEQDPVEVRIGELPIMLKSVCRLYGLSDEELKRLGEDPKDPGGYFIING 160
Pab_rpoB RKPLYPMDARIRNLTYSAPLYLEMIPIVIRGLEQEPVEVRIGELPIMLKSVCRLYGLSDEELKRLGEDPKDPGGYFIING 160
Pho_rpoB RKPLYPMDARIRNLTYSAPLYLEMIPIVIRGLEQDPVEVRIGELPIMLKSVCRLYGLSDEELKRLGEDPKDPGGYFIING 160

Pfu_rpoB SERVIVSIEDIAPNRTLVEKDERQEKYIAKVSYRHGYRALVVERKKGILYVCIIPNVKPKVKFVYVMRALGLERDKDI 240
Pab_rpoB SERVIVSIEDIAPNRTLVEKDERQEKYIAKVSYRHGYRALVVERKKGILYVCIIPNVKPKVKFVYVMRALGLERDKDI 240
Pho_rpoB SERVIVSIEDIAPNRTLVEKDERQEKYIAKVSYRHGYRALVVERKKGILYVCIIPNVPRPKVKFVYVMRALGLERDKDI 240

Pfu_rpoB VEAVGNDPEVQQIMFDNLEDASDITQOEALYIGKLVAPGQAREYRLKRAEYVIDNLLPHMGVSPEDRIKAYYLGM 320
Pab_rpoB VDAVGNDPEVQQILFDNLEDASDITQOEALYIGKLVAPGQAREYRLKRAEYVIDNLLPHMGVSPEDRIKAYYLGM 320
Pho_rpoB VDAVGNDPEVQQILFDNLEDASDITQOEALYIGKLVAPGQAREYRLKRAEYVIDNLLPHMGVTPEDRIKAYYLGM 320

Pfu_rpoB ALKVIELSLGKRDEDDKDYANKRRLRAGDLLKDLFRVAFSGLVKDIQYQMTKTYQRGKDYVFCNIHRFIRNSIRPDVL 400
Pab_rpoB ALKVIELSLGKRDEDDKDYANKRRLRAGDLLKDLFRVAFSGLVKDIQYQMTKTYQRGKDYVFCNIHRFIRNSIRPDVL 400
Pho_rpoB ALKVIELSLGKRDEDDKDYANKRRLRAGDLLKDLFRVAFSGLVKDIQYQMTKTYQRGKDYVFCNIHRFIRNSIRPDVL 400

Pfu_rpoB TERIEHALATGAWPGGRTGVSQLLDRNTNYVSTLSHLRRVTSPLDREQPHFEARDLHGTHWGRICPTETPEGPNCGLVKNL 480
Pab_rpoB TERIEHALATGAWPGGRTGVSQLLDRNTNYVSTLSHLRRVTSPLDREQPHFEARDLHGTHWGRICPTETPEGPNCGLVKNL 480
Pho_rpoB TERIEHALATGAWPGGRTGVSQLLDRNTNYVSTLSHLRRVTSPLDREQPHFEARDLHGTHWGRICPTETPEGPNCGLVKNL 480

Pfu_rpoB ALMAQITTAVPEEEIREYLMSLGIVPIEVRRPSPPLWRVYLVNGVLIIGTIEDGKALVRIQDRRSKGISDINVAYYEDV 560
Pab_rpoB ALMAQITTAVPEEEVKEYLMSLGIVPIEVRRPSPPLWRVYLVNGVLIIGTIEDGKALVERIKKDRRSKGISDINVAYYEDV 560
Pho_rpoB ALMAQITTAVPEEEVREYLMSLGIVPIEVRRPSPPLWRVYLVNGVLIIGTVEDGKALVERIKKDRRSKGISDINVAYYEDV 560

Pfu_rpoB KEVYVNSDDGRVRRPLIIVENGVPKLTKEHVQAVKEGRKWSDLIKMGIIEYLDAAEEENAYVAMWPWEVTEHHTHELM 640
Pab_rpoB REVYVNSDDGRVRRPLIIVENGVPKLTKEHVQAIKEGRKWSDLVRMGVIEYLDAAEEENAYVAMWPWEVTEHHTHELM 640
Pho_rpoB KEVYVNSDDGRVRRPLIIVENGVPKLTKEHVQAIKEGRKWSDLVRMGVIEYLDAAEEENAYVAMWPWEVTEHHTHELM 640

Pfu_rpoB PAAIILGLPASLIPYPEHNAAPRNTYGAGMAKQSLGLGWANFRIRVDTRGHLLHYPOIPLVNSRIMKAVGYEDRPAGQNFV 720
Pab_rpoB PAAIILGLPASLIPYPEHNAAPRNTYGAGMAKQSLGLGWANFRIRVDTRGHLLHYPOIPLVNSRIMKAVGYEDRPAGQNFV 720
Pho_rpoB PAAIILGLPASLIPYPEHNAAPRNTYGAGMAKQSLGLGWANFRIRVDTRGHLLHYPOIPLVNSRIMKAVGYEDRPAGQNFV 720

Pfu_rpoB VAVLAYGYNMEDAIIINKASIERGLARSTFFRTYEAEEKYLGQTDKFEIPDPTVRYRGEKYYRHLDEDGLIFPESK 800
Pab_rpoB VAVLAYGYNMEDAIIINKASIERGLARSTFFRTYEAEEKYLGQTDKFEIPDPTVRYRGEKYYRHLDEDGLIFPESK 800
Pho_rpoB VAVLAYGYNMEDAIIINKASIERGLARSTFFRTYEAEEKYLGQTDKFEIPDPTVRYRGEKYYRHLDEDGLIFPESK 800

```

<i>Pfu_rpoB</i>	VEGKDVLVGRSTSPRRFLEERGLGAVLSLQERRETSVAVRPSEKGVVDKVIITETGDGTKLVKVTVRDLRIPEIGDKFASRH	880
<i>Pab_rpoB</i>	VEGKDVLVIGRSTSPRRFLEERGLGAVLSLQERRETSVAVRPSEKGVVDKVIITETGDGTKLVKVTVRDLRIPEIGDKFASRH	880
<i>Pho_rpoB</i>	VEGKDVLVIGRSTSPRRFLEERGLGAVLSLQERRETSVAVRPSEKGVVDKVIITETGDGTKLVKVTVRDLRIPEIGDKFASRH	880
<i>Pfu_rpoB</i>	GQKGVIGLIVPQEDMPWTESGIVPDLIVNPHGIPSRMTVGQLLEAIGGKVAALKGRRVDGTAFIGEPEEKLRKELEELGF	960
<i>Pab_rpoB</i>	GQKGVIGLIVPQEDMPWTESGIVPDLIVNPHGIPSRMTVGQLLEAIGGKVAALKGRRVDGTAFIGEPEEKLRKELEELGF	960
<i>Pho_rpoB</i>	GQKGVIGLIVPQEDMPWTESGIVPDLIVNPHGIPSRMTVGQLLEAIGGKVAALKGRRVDGTAFIGEPEEKLRKELEELGF	960
<i>Pfu_rpoB</i>	KHTGREIMYDGITGKRLEADIFIGVYYQLRHHMVADKIHARSRGPVQVLTQPTTEGRAREGGLRFGEMERDVLVGHGAA	1040
<i>Pab_rpoB</i>	KHTGREIMYDGITGKRLEADIFIGVYYQLRHHMVADKIHARSRGPVQVLTQPTTEGRAREGGLRFGEMERDVLVGHGAA	1040
<i>Pho_rpoB</i>	KHTGREIMYDGITGKRLEADIFIGVYYQLRHHMVADKIHARSRGPVQVLTQPTTEGRAREGGLRFGEMERDVLVGHGAA	1040
<i>Pfu_rpoB</i>	MLLIERLLEESDKTEVWVCENCGHIALEDKRRERKVCYPCVCGEEERISKVEMSYAFKLLDELKAMVIRPKLNLSEV	1117
<i>Pab_rpoB</i>	MLLIERLLEESDKTEVWVCENCGHIALEDKRRERKVCYPCVCGEEERISKVEMSYAFKLLDELKAMVIRPKLNLSEV	1117
<i>Pho_rpoB</i>	MLLIERLLEESDKTEVWVCENCGHIALEDKRRERKVCYPCVCGEEERISKVEMSYAFKLLDELKAMVIRPKLNLSEV	1117

(Positionen mit Aminosäureaustausch: 66, hiervon 51 ähnlich und 15 stark abweichend).

8.1.2.2 Untereinheit A' (rpoA1)

P.furiosus PF1563, *P.abysyi* rpoA1 / PAB0424, *P.horikoshii* PH1545

Bei *P.furiosus* wurde der Startpunkt drei Aminosäure-Positionen früher als in PF1563 annotiert angesetzt. Die Daten stimmen auf diese Weise u.a. besser mit der Position der ribosomalen Bindestelle überein.

<i>Pfu_rpoA1</i>	MOSMKVIGSIEFGILSPQIEIRKMSAVEVTVPDTYDDDGYPTEGGVMDKRMGVIDPGLRCETCGGKAGECPGHFGHIELA	77
<i>Pab_rpoA1</i>	MOSVKKVIGSIEFGILSPQIEIRKMSAVEVTVPDTYDDDGYPTEGGVMDKRMGVIDPGLRCETCGGKAGECPGHFGHIELA	80
<i>Pho_rpoA1</i>	MHSVKKVIGSIEFGILSPQIEIRKMSAVEVTVPDTYDDDGYPTEGGVMDKRMGVIDPGLRCETCGGKAGECPGHFGHIELA	80
<i>Pfu_rpoA1</i>	RPVIHVGFAKTIYRILESTCRECGRIKLTDEEIEEYMKKLELAKNRRSEVNIKEIHKKAKERMVCPHCGAPQYPIKFE	157
<i>Pab_rpoA1</i>	RPVIHVGFAKTIHRIILESTCRECGRIKLTDEEIEEYMKKLELAKNRRSEVNIKEIHKKAKERMVCPHCGAPQYPIKFE	160
<i>Pho_rpoA1</i>	RPVIHVGFAKTIHRIILESTCRECGRIKLTDEEIEEYMKKLELAKNRRSEVNIKEIHKKAKERMVCPHCGAPQYPIKFE	160
<i>Pfu_rpoA1</i>	KPTIYWEIRKDEQOGNEYRHRMMPTEIRDRLEKIPDKLPLLGLDPEKSRPEWVMVLTPLVPPVPTARPSITLETGIRAEDD	237
<i>Pab_rpoA1</i>	KPTIYWEIRKDEQOGNEYKHRMMPTEIRDRLEKIPDKLPLLGLDPEKSRPEWVMVLTPLVPPVPTARPSITLETGIRAEDD	240
<i>Pho_rpoA1</i>	KPTIYWEIRKDEQOGNEYKHRMMPTEIRDRLEKIPDKLPLLGLDPEKSRPEWVMVLTPLVPPVPTARPSITLETGIRAEDD	240
<i>Pfu_rpoA1</i>	LTHKLVDIIRINNRLKQNIIEAGAPQLIIEDLWDLQYHVTTYINNEISGVPAPAKHKSGRPLKTLAQRKLGKEGRFRGNLS	317
<i>Pab_rpoA1</i>	LTHKLVDIIRINNRLKQNIIEAGAPQLIIEDLWDLQYHVTTYINNEISGVPAPAKHKSGRPLKTLAQRKLGKEGRFRGNLS	320
<i>Pho_rpoA1</i>	LTHKLVDIIRINNRLKQNIIEAGAPQLIIEDLWDLQYHVTTYINNEISGVPAPAKHKSGRPLKTLAQRKLGKEGRFRGNLS	320
<i>Pfu_rpoA1</i>	GKRVNFSARTVISPDPMISINEVGVPVEVAMELTVPEKVTEFNIEIRLKKMVLNPGDKYPGANYVIDPEGRRIRIMDSNKE	397
<i>Pab_rpoA1</i>	GKRVNFSARTVISPDPMISINEVGVPQIAMELTVPEKVTEFNIEIRLKKMVLNPGDKYPGANYVIDPEGRRIRIMESNRE	400
<i>Pho_rpoA1</i>	GKRVNFSARTVISPDPMISINEVGVPQIAMELTVPEKVTEFNIEIRLKKMVLNPGDKYPGANYVIDPEGRRIRIMESNKE	400
<i>Pfu_rpoA1</i>	NLANOLDIGWTVRHLMDGDIVLFRNQPSLHRMSIMAHVRVMPYRTRFRLNLAVCPPYNADFDGDEMNLHVPQTEEAQAE	477
<i>Pab_rpoA1</i>	NLARLMDIGWTVRHLMDGDIVLFRNQPSLHRMSIMAHVRVMPYRTRFRLNLAVCPPYNADFDGDEMNLHVPQTEEAQAE	480
<i>Pho_rpoA1</i>	NLAKLMDIGWTVRHLMDGDIVLFRNQPSLHRMSIMAHVRVMPYRTRFRLNLAVCPPYNADFDGDEMNLHVPQTEEAQAE	480
<i>Pfu_rpoA1</i>	ARILMEVQNHIIISPRYGGPIIGGIQDHISGGYLLTREGAYFTRDEVEQMLMFAGVDIKELPEPKDYENGKPLWSGKTIF	557
<i>Pab_rpoA1</i>	ARILMEVQNHIIISPRYGGPIIGGIQDHISGGYLLTREGAYFTRDEVEQMLMFAGVDIKELPEPKDYENGKPLWSGKTIF	559
<i>Pho_rpoA1</i>	AKILMEVQNHIIISPRYGGPIIGGIQDHISGGYLLTREGAYFTRDEVEQMLMFAGVDIKELPEPKDYENGKPLWSGKTIF	559
<i>Pfu_rpoA1</i>	SLLLPEDLTVWYRNKLCDEPERCEALEKLEIEKLPDPEEVRKLAYDGFVYIQNGKLLSGAIDKKAYGREDGIILDLIVR	637
<i>Pab_rpoA1</i>	SLLLPDDLTVWYRNKLCDEPERCEALEKLEIEKLPDPEEVRKLAYDGFVYIQNGKLLSGAIDKKAYGREDGIILDLIVR	639
<i>Pho_rpoA1</i>	SLLLPDDLTVWYRNKLCDEPERCEALEKLEIEKLPDPEEVRKLAYDGFVYIQNGKLLSGAIDKKAYGREDGIILDLIVR	639
<i>Pfu_rpoA1</i>	EYGVERRARQFLDQVTKLTIWVITHKGFTTGIDDEDLPEEARDRIEIREAEERVKRLIEAYKRGELEPLPGKSLEDTLE	717
<i>Pab_rpoA1</i>	EYGVERRARQFLDQVTKLTIWVITHKGFTTGIDDEDLPEEARDRIEIREAEERVKRLIEAYKRGELEPLPGKLTLEDTLE	719
<i>Pho_rpoA1</i>	EYGVERRARQFLDQVTKLTIWVITHKGFTTGIDDEDLPEEARDRIEIREAEERVKRLIEAYKRGELEPLPGKSLEDTLE	719
<i>Pfu_rpoA1</i>	SKIMAVLAEARDNAGVAEKYLGMDNHTVIMAKTGARGKILNITQMAALLGQOSIRGKRLYRGFRGRLVSHFKPGDLGAR	797
<i>Pab_rpoA1</i>	SKIMAVLAEARDNAGSIAEKYLGMDNHTVIMAKTGARGKILNITQMAALLGQOSIRGKRLYRGFRGRLVSHFKPGDLGAR	799
<i>Pho_rpoA1</i>	SKIMAVLAEARDNAGSVAEKYLGMDNHTVIMAKTGARGKILNITQMAALLGQOSIRGKRLYRGFRGRLVSHFKPGDLGAR	799
<i>Pfu_rpoA1</i>	AKGFVNSYKSGLTPQEYFFHAMGGREGLVDTAVRTAQSGYMQORRLINALQDLKVEYDGTVRSPGIVQFRYGEDGVDP	877
<i>Pab_rpoA1</i>	AKGFVNSYKSGLSPQEYFFHAMGGREGLVDTAVRTAQSGYMQORRLINALQDLKVDYDGTVRDPTGIVQFRYGEDGVDP	879
<i>Pho_rpoA1</i>	AKGFVNSYKSGLSPQEYFFHAMGGREGLVDTAVRTAQSGYMQORRLINALQDLKVDYDGTVRDPTGIVQFRYGEDGVDP	879
<i>Pfu_rpoA1</i>	MKSWGKKTVDVDRIVRITLLKMRGNKAKV	910 (korrigiert)
<i>Pab_rpoA1</i>	MKSWGKKTVDVDRIVRITLLKMRANNGK	907
<i>Pho_rpoA1</i>	MKSWGKKTVDVDRIVRITLLKMRSNCK	907

(Positionen mit Aminosäureaustausch: 72, hiervon 43 ähnlich und 29 stark abweichend).

8.1.2.3 Untereinheit A'' (rpoA2)

P.furiosus PF1562, *P.abysyi* rpoA2 / PAB0425, *P.horikoshii* PH1544

<i>Pfu_rpoA2</i>	MVSSSTIKSLIEKKGKDLPEVSKOELYEKLYKYNEKYKLTFAEVEIIDDVVKEYERALVEPGEAVGTVAASIGEPSTQ	80
<i>Pab_rpoA2</i>	MVSSSTIKSLIEKKGKDLPEVSKOELYEKLYKYNEKYKLTFAEVEIIDEVVKEYERALVEPGEAVGTVAASIGEPSTQ	80
<i>Pho_rpoA2</i>	MVSSSTIKSLIEKKGKDLPEVSKOELYEKLYKYNEKYKLTFAEVEIIDEVKEYERALVEPGEAVGTVAASIGEPSTQ	80

<i>Pfu_rpoA2</i>	MTLNTFHYAGVAEINVTLGLPRIIEIVDARKNPSTPMMTVYLDEEHYDRKAEVARRIEGTTLENLARTTLDLINME	160
<i>Pab_rpoA2</i>	MTLNTFHYAGVAEINVTLGLPRIIEIVDARKNPSTPMMTVYLDEEHYDRKAEVARRIEGTTLENLARTTLDLINME	160
<i>Pho_rpoA2</i>	MTLNTFHYAGVAEINVTLGLPRIIEIVDARKNPSTPMMTVYLDEEHYDRKAEVARRIEGTTLENLARTTLDLINME	160
<i>Pfu_rpoA2</i>	FIVEIDPERLERSGLTMEKVVKKLQSSFKSAEFEVGYTTLIVRPKADKISDLRRFAEKIKKHLRLKGLSGVGKTIIVRKEG	240
<i>Pab_rpoA2</i>	FIVEIDPERLEKSGLTMEKVLKKLQSSFKSAEFEVGYTTLIVRPKADKISDLRRFAEKIKKHLRLKGLSGVGKTIIVRKEG	240
<i>Pho_rpoA2</i>	FIVEIDPERLEKSGLTMEKVLKKLQSSFKSAEFEVGYTTLIVRPKADKISDLRRFAEKIKKHLRLKGLSGVGKTIIVRKEG	240
<i>Pfu_rpoA2</i>	DEYVIYTEGSNFKQVLKVPVDPTRTRTNNIHEIAEVLGIEAARNAI IDEIVSTMEOEQGLEVDIRHIMLVADMTLDGIV	320
<i>Pab_rpoA2</i>	DEYVIYTEGSNFKQVLKVPVDPTRTRTNNIHEIAEVLGIEAARNAI IEEIMNTMREOQGLEVDIRHIMLVADMTLDGIV	320
<i>Pho_rpoA2</i>	DEYVIYTEGSNFKQVLKVPVDPTRTRTNNIHEIAEVLGIEAARNAI IEEIMNTMREOQGLEVDIRHIMLVADMTLDGIV	320
<i>Pfu_rpoA2</i>	RPIGRHGVSVEKSSVLARAFAFEITVQHLFEAAEKGEVDNLGVIENVLIGQVPVGTGMVKLTMKPLRPQKEKEEV	397
<i>Pab_rpoA2</i>	RPIGRHGVSVEKASVLARAFAFEITVQHLFEAAERGEVDNLGVIENVLIGQVPVGTGMVKLTMKPLRPQKEKEEV	397
<i>Pho_rpoA2</i>	RPIGRHGVSVEKASVLARAFAFEITVQHLFEAAERGEVDNLGVIENVLIGQVPVGTGMVKLTMKPLRPQKEKEEV	397

(Positionen mit Aminosäureaustausch: 41, hiervon 28 ähnlich und 13 stark abweichend).

8.1.2.4 Untereinheit D (rpoD)

P.furiosus PF1647, *P.abysii* rpoD / PAB2410, *P.horikoshii* PH1637

<i>Pfu_rpoD</i>	MAGIEVQILEKEDSIRKIVLKGVHVSFANALRRITILGEVPTFAVDEVEFYENDSALFDEIIAHLRAMIPLKTPVDRFELD	80
<i>Pab_rpoD</i>	MVEIKILKEDSIRKIVLKGVHVSFANALRRITILGEVPTFAVDEVEFYENDSALFDEIIAHLRAMIPLKTPVDRFELD	78
<i>Pho_rpoD</i>	MEMVKIKLKKEDSIRKIVLKGVHVSFANALRRITILGEVPTFAVDEVEFYENDSALFDEIIAHLRAMIPLKTPVDRFELD	80
<i>Pfu_rpoD</i>	ALELDDYTVTLSLEAEGPGIVYSGDLKSDDPVVKPVPNPIPIVKLAKGQRLVFNAYAKLGRGKDHAKWQPGFYKYKPT	160
<i>Pab_rpoD</i>	ALELDDYTVTLSLEAEGPGIVYSGDLKSDDPVVKPVPNPIPIVKLAKGQRLVFNAYAKLGRGKDHAKWQPGFYKYKPT	158
<i>Pho_rpoD</i>	ALELDDYTVTLSLEAEGPGIVYSGDLKSDDPVVKPVPNPIPIVKLAKGQRLVFNAYAKLGRGKDHAKWQPGFYKYKPT	160
<i>Pfu_rpoD</i>	VHISKIPEWELKLLAKKRGLPVKEEVEEVLVITIKPFYIPKDFEYEGKEIWEIIVPNTYIFIVETNGELPVEEIVKI	240
<i>Pab_rpoD</i>	IHISKIEGWEKLLAKKRGLPVKEEVEEVLVITIKPFYIPKDFEYEGKEIWEIIVPNTYIFIVETNGELPVEEIVKI	238
<i>Pho_rpoD</i>	VHISKIPEWELKLLAKKRGLPVKEEVEEVLVITIKPFYIPKDFEYEGKEIWEIIVPNTYIFIVETNGELPVEEIVKI	240
<i>Pfu_rpoD</i>	ALKILMRKADRFISELQKLS	261
<i>Pab_rpoD</i>	ALKILMRKADRFINELQLAG	259
<i>Pho_rpoD</i>	ALRILMRKADRFINELQLAG	261

(Positionen mit Aminosäureaustausch: 49, hiervon 23 ähnlich und 26 stark abweichend).

8.1.2.5 Untereinheit E' (rpoE1)

P.furiosus PF0256, *P.abysii* rpoE1 / PAB1105, *P.horikoshii* PH1908

<i>Pfu_rpoE1</i>	MYKIVTVKDVVRIPPMFTMDPKEAAKILRETYEGHYDKDEGVILSILEVKDIDKGGIIPGDGATYHEVDFVLVWEPK	80
<i>Pab_rpoE1</i>	MYKIVTVKDVVRIPPMFTMDPKEAAKILRETYEGHYDKDEGVILSILEVKDIDKGGIIPGDGATYHEVDFVLVWEPK	80
<i>Pho_rpoE1</i>	MYKIVTVKDVVRIPPMFTMDPKEAAKILRETYEGHYDKDEGVILSILEVKDIDKGGIIPGDGATYHEVDFVLVWEPK	80
<i>Pfu_rpoE1</i>	IHEVVEGYVADVMPFGAFIRIGPIDGLVHISQLMDDYVYVDERNKQFVGKEKKYLLKIGDAVRARIINISAKSKVIRENR	160
<i>Pab_rpoE1</i>	QOEVEGTVYDADVMPFGAFVIRIGPMDGLVHISQLMDDYVYVDERNKQFVGKEKKYLLKIGDAVRARIINISAKSKVIRENR	160
<i>Pho_rpoE1</i>	QDEVVEGTVYDADVMPFGAFIRIGPMDGLVHISQLMDDYVYVDERNKQFVGKEKKYLLKIGDAVRARIINISAKSKVIRENR	160
<i>Pfu_rpoE1</i>	IGLTMROPGLGKFEWIEKEKRKEKEGKK	189
<i>Pab_rpoE1</i>	IGLTMROPGLGKFEWIEKEKRKEKEGKK	188
<i>Pho_rpoE1</i>	IGLTMROPGLGKFEWIEKEKRKEKEGKK	188

(Positionen mit Aminosäureaustausch: 28, hiervon 15 ähnlich und 13 stark abweichend).

8.1.2.6 Untereinheit E'' (rpoE2)

P.furiosus PF0255, *P.abysii* rpoE2 / PAB7428, *P.horikoshii* PH1908a

(trotz Überlappung mit dem C-terminalen Anteil von E' höchstwahrscheinlich korrekt translatiert)

<i>Pfu_rpoE2</i>	MSEKACRNCHYITSEDRCPCVCGSRDLSEEWFDLVIIDVENSEIAKKGAKVPGKYAIRVR	61
<i>Pab_rpoE2</i>	MTEKACRNCHYITTEDQCPVCGSRDLSEEWFDLVIIDVENSEIAKKGAKVPGKYAIRVR	60
<i>Pho_rpoE2</i>	MTEKACRNCHYITTEDQCPVCGSRDLSEEWFDLVIIDVENSEIAKKGAKVPGKYAIRVR	61

(Positionen mit Aminosäureaustausch: 8, hiervon 7 ähnlich und 1 stark abweichend).

8.1.2.7 Untereinheit F (rpoF)

P.furiosus PF1036, *P.abysii* PAB0732, *P.horikoshii* PH1127

<i>Pfu_rpoF</i>	MIGRCKLGEKYITIAEAKELLKREVEEVKAGIEEPLYEARLALEHAERFAKLEADKAKEAVEELMNAFEWMSDRICK	80
<i>Pab_rpoF</i>	MIGRCKLGEKYITIAEAKELLKREVEEVKAGIEEPLYEARLALEHAERFAKLEADKAKEAVEELMNAFEWMSDRICK	80
<i>Pho_rpoF</i>	MIGRCKLGEKYITIAEAKELLKREVEEVKAGIEEPLYEARLALEHAERFAKLEADKAKEAVEELMNAFEWMSDRICK	80
<i>Pfu_rpoF</i>	IVDIMPEDSMDLRVIFAKEEYQPTQEMKQILDILDKYRE	120
<i>Pab_rpoF</i>	IVDIMPEDNLDLRVIFAKEEYQPTQEMKQILDILDKYRE	121
<i>Pho_rpoF</i>	IVDIMPEDNLDLRVIFAKEEYQPTQEMKQILDILDKYRE	121

(Positionen mit Aminosäureaustausch: 30, hiervon 21 ähnlich und 9 stark abweichend).

8.1.2.8 Untereinheit H (rpoH)

P.furiosus PF1565, *P.abysii* rpoH / PAB7151, *P.horikoshii* PHS044

```
Pfu_rpoH  MAGKKEFSIFDHLVLPPEHRVLSSEEEKKALLEKYKITPAQLPQIKASDPAVKALGAKPGDVIEIKRKSPTAGVYYYYRVVV 80
Pab_rpoH  MAGKSEFNIFKHVLPPEHRILSSEEEKKALLEKYRITPAQLPQIKASDPAVKALGAKPGDIEIKRKSPTAGVYYYYRVVV 80
Pho_rpoH  MSGKNEFNIFKHVLPPEHRILSSEEEKKALLEKYKITPAQLPQIRASDPAVKALGAKPGDIEIKRKSPTAGVYYYYRVVV 80

Pfu_rpoH  ED 82
Pab_rpoH  ED 82
Pho_rpoH  ED 82
```

(Positionen mit Aminosäureaustausch: 10, hiervon 7 ähnlich und 3 stark abweichend).

8.1.2.9 Untereinheit K (rpoK)

P.furiosus PF1642, *P.abysii* rpoK / PAB7132, *P.horikoshii* PH1631a

```
Pfu_rpoK  MFKYTRFEKARIIGARALQISMGAPVLIDVPIGITPLEAAILEFEKGVIPITVIRPS 57
Pab_rpoK  MFKYTRFEKARIIGARALQIAMGAPVLIDVPIGITPLEAAIMEFEKGVIPITVIRPS 57
Pho_rpoK  MFKYTRFEKARIIGARALQIAMGAPVLIDVPIGITPLEAAIMEFEKGVIPITVIRPS 57
```

(Positionen mit Aminosäureaustausch: 3, hiervon 2 ähnlich und 1 stark abweichend).

8.1.2.10 Untereinheit L (rpoL)

P.furiosus PF0050, *P.abysii* rpoL / PAB2316, *P.horikoshii* PH0041a

```
Pfu_rpoL  MKIEVIKKEENLLEFYLEGEDHTFANLLVEITLRENPHVKFVAYTIEHPITMARKPRFRVVDGKITPEKALEEAACKIEF 80
Pab_rpoL  MKIEVIKKEQDLLEFYLEGEDHTFANLLVEITLRENPHVKFVAYTIEHPITMARKPRFRVVDGKITPEKALEEAACKIEF 80
Pho_rpoL  MKIEVIKKEEDMLEFYLEGEDHTFANLLTEVLREDPHVKFVAYTIEHPITMARKPRFRVVDGKVTPEKALEEAARKIEF 80

Pfu_rpoL  RAKEVLEAWERAVKS 95
Pab_rpoL  RAKEVLDAWERAVKS 95
Pho_rpoL  RAKEVLDAWERAVKS 95
```

(Positionen mit Aminosäureaustausch: 17, hiervon 10 ähnlich und 7 stark abweichend).

8.1.2.11 Untereinheit N (rpoN)

P.furiosus PF1643, *P.abysii* rpoN / PAB7131, *P.horikoshii* PH1632

```
Pfu_rpoN  ~~~~~~MGASPVIVPVRFCFCGKVIKDYEFKRRVE 31
Pab_rpoN  ~~~~~~MIVPVRFCFCGKVIKDYEFKRRVE 26
Pho_rpoN  MIEQCWSEIQEESPTSQIGQLRVQGLRDKRVIVEYIPLILKTFLLSKTSGVGLIIPVRCFCFCGKVIKDYEFKRRVE 80

Pfu_rpoN  AGEDPEKVLDDLGLERYCCRRMLLSHVELIDDIMHYRVY 70
Pab_rpoN  AGEDPEKVLDDLGLERYCCRRMLLSHVELIDDIMHYRVY 65
Pho_rpoN  AGEDPEKVLDDLGLERYCCRRMLLSHVELIDDIMHYRVY 119
```

(Positionen mit Aminosäureaustausch: 57, hiervon 3 ähnlich und 54 stark abweichend , s.u.).

Untereinheit N mit korrigiertem Start von *Pho_rpoN*:

```
Pfu_rpoN  MGASPVIVPVRFCFCGKVIKDYEFKRRVEAGEDPEKVLDDLGLERYCCRRMLLSHVELIDDIMHYRVY 70
Pab_rpoN  ~~~~~MIVPVRFCFCGKVIKDYEFKRRVEAGEDPEKVLDDLGLERYCCRRMLLSHVELIDDIMHYRVY 65
Pho_rpoN  ~~~~MGLIIPVRCFCFCGKVIKDYEFKRRVEAGEDPEKVLDDLGLERYCCRRMLLSHVELIDDIMHYRVY 68
```

(Positionen mit Aminosäureaustausch: 8, hiervon 3 ähnlich und 5 stark abweichend).

8.1.2.12 Untereinheit P (rpoP)

P.furiosus PF2009, *P.abysii* PAB3072, *P.horikoshii* PHS056

```
Pfu_rpoP  ~~~~~~MVEAVYRCQKCGREVKLDLSITRDLRCPYCGSKILYKPRPKVPRRVKAI 49
Pab_rpoP  ~~~~~~MVEAVYRCQKCGREVKLDLSITRDLRCPYCGSKILYKPRPKVPRRVKAI 49
Pho_rpoP  MGQYLRVWLISQSLRPVRLLEELWSDVEMPEAVYRCQKCGREVKLDLSITRDLRCPYCGSKILYKPRPKVPRRVKAI 77
```

Untereinheit P mit korrigiertem Start von *Pho_rpoP*:

```
Pfu_rpoP  MVEAVYRCQKCGREVKLDLSITRDLRCPYCGSKILYKPRPKVPRRVKAI 49
Pab_rpoP  MVEAVYRCQKCGREVKLDLSITRDLRCPYCGSKILYKPRPKVPRRVKAI 49
Pho_rpoP  MVEAVYRCQKCGREVKLDLSITRDLRCPYCGSKILYKPRPKVPRRVKAI 49
```

(Positionen mit Aminosäureaustausch: 6, hiervon 3 ähnlich und 3 stark abweichend).

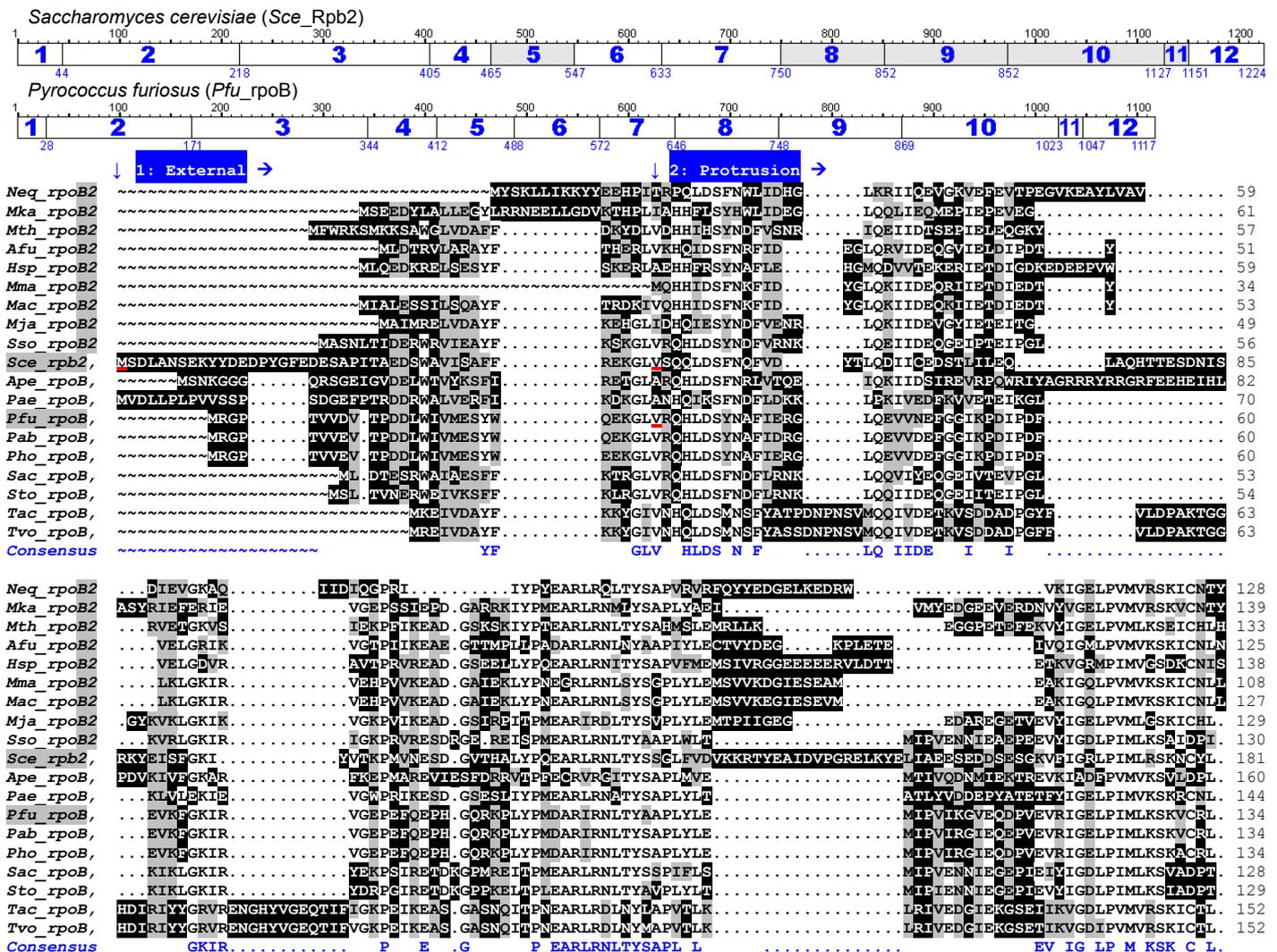
8.1.3 Alignments mit Archaea und Eukarya

8.1.3.1 Untereinheit B und B' / B'' (rpoB, rpoB1, rpoB2)

U.a. bei vielen methanogenen *Archaea* ist die große B-Untereinheit in zwei Teile (B' und B'') geteilt. Es mußte die Entscheidung getroffen werden, ob zwei separate Alignments mit B + B' einerseits und B + B'' andererseits gezeigt werden sollte, oder aber eine Gesamtübersicht mit B + B' + B''. Ich habe mich für Letzteres entschieden, weil sich an dem kombinierten Sequenzvergleich einige Details im Übergangsbereich vom C-Terminus von B'' zum N-Terminus von B' zeigen lassen. Dass auf diese Weise für ca. die Hälfte der untersuchten Organismen jeweils zwei Aminosäure-Sequenzen im Alignment enthalten sind, wirkt sich ziemlich negativ auf den Konsensus aus. Da dieses Alignment vor allem dazu gedacht ist, die Struktur der RNA-Polymerase bei *P.furiosus* besser zu verstehen, und das Auffinden von konservierten Regionen hauptsächlich über den Vergleich mit einem aussagekräftigen Konsensus geschied, wurde der „bessere“ Konsens aus den o.g. beiden einzelnen Alignments in Form einer kombinierten Konsensus-Zeile hinzugefügt (Consensus). In vielen Fällen wären die Alignments ohne die Gene von *S.cerevisiae* homogener gewesen. Da aber eine der Hauptaussagen dieser Darstellungen in dem Bezug von archaeellen Mustern auf die Verhältnisse bei *Eukarya* liegt, sind die teilweise unvorteilhafte Verteilung hingenommen werden.

Das Zinkfinger-Motiv wurde rot hinterlegt. Die Segmentgrenzen wurden rot unterstrichen.

Ape = *Aeropyrum pernix*, Tac = *Thermoplasma acidophilum*, Tvo = *Thermoplasma volcanium*, Mka = *Methanopyrus kandleri* AV19, Mja = *Methanocaldococcus jannaschii*, Mth = *Methanothermobacter thermoautotrophicus*, Mac = *Methanosarcina acetivorans* C2A, Mma = *Methanosarcina mazei* Goe1, Pfu = *Pyrococcus furiosus* DSM 3638, Pab = *Pyrococcus abyssi*, Pho = *Pyrococcus horikoshii*, Afu = *Archaeoglobus fulgidus*, Hsp = *Halobacterium* sp. NRC-1, Sac = *Sulfolobus acidocaldarius*, Sso = *Sulfolobus solfataricus*, Sto = *Sulfolobus tokodaii*, Pae = *Pyrobaculum aerophilum*, Neq = *Nanoarchaeum equitans* Kin4-M., Sce = *Saccharomyces cerevisiae* (rpb2 als homologes Gen bei *Eukarya*).



3: Lobe

Table of amino acid sequences for region 3: Lobe, including Neq_rpoB2, Mka_rpoB2, Mth_rpoB2, Afu_rpoB2, Hsp_rpoB2, Mma_rpoB2, Mac_rpoB2, Sso_rpoB2, Sce_rpb2, Ape_rpoB, Pae_rpoB, Pfu_rpoB, Pab_rpoB, Pho_rpoB, Sac_rpoB, Stc_rpoB, Tco_rpoB, Tvo_rpoB, and Consensus. The table shows conserved residues across species and a consensus sequence at the bottom: L GEDP DPGYFIINGSERVIV EDLAPN LVE AK S GYRA ER ... L V.

Table of amino acid sequences for region 4: Protrusion, including Neq_rpoB2, Mka_rpoB2, Mth_rpoB2, Afu_rpoB2, Hsp_rpoB2, Mma_rpoB2, Mac_rpoB2, Mja_rpoB2, Sso_rpoB2, Sce_rpb2, Ape_rpoB, Pae_rpoB, Pfu_rpoB, Pab_rpoB, Pho_rpoB, Sac_rpoB, Stc_rpoB, Tco_rpoB, Tvo_rpoB, and Consensus. The table shows conserved residues across species and a consensus sequence at the bottom: P PFV MRAAG D DIV AV DPE L LE ... T E ... A IG R.VA GQ E R RA D YLLPHLG.

4: Protrusion

Table of amino acid sequences for region 5: Fork, including Neq_rpoB2, Mka_rpoB2, Mth_rpoB2, Afu_rpoB2, Hsp_rpoB2, Mma_rpoB2, Mac_rpoB2, Mja_rpoB2, Sso_rpoB2, Sce_rpb2, Ape_rpoB, Pae_rpoB, Pfu_rpoB, Pab_rpoB, Pho_rpoB, Sac_rpoB, Stc_rpoB, Tco_rpoB, Tvo_rpoB, and Consensus. The table shows conserved residues across species and a consensus sequence at the bottom: D ... R KA L MA EL LG R EDDKDHYANKRL LAGLD LFR AF L KD YQLE R R ... VR D LIT R.

5: Fork

Table of amino acid sequences for region 6: External 2, including Neq_rpoB2, Mka_rpoB2, Mth_rpoB2, Afu_rpoB2, Hsp_rpoB2, Mma_rpoB2, Mac_rpoB2, Mja_rpoB2, Sso_rpoB2, Sce_rpb2, Ape_rpoB, Pae_rpoB, Pfu_rpoB, Pab_rpoB, Pho_rpoB, Sac_rpoB, Stc_rpoB, Tco_rpoB, Tvo_rpoB, Sso_rpoB2, and Consensus. The table shows conserved residues across species and a consensus sequence at the bottom: HA ATGNWVGG ... RTGVSQLLDRT Y STLSHLRRV SP L R QPHEARDLH TQWR CP ETPEGPNCGLVKNLAL G E.

6: External 2

7: External 1 →

Table with protein names (Neq_rpoB2, Mka_rpoB2, Mth_rpoB2, etc.) and their corresponding amino acid sequences. The sequences are aligned across columns. A consensus sequence is provided at the bottom: G V Y N G L G E L I R R R G I S ... V E V N D G R R R P L V E N G P

Table with protein names (Sce_rpb2, Ape_rpoB, Pae_rpoB, etc.) and their corresponding amino acid sequences. The sequences are aligned across columns. A consensus sequence is provided at the bottom: T E H E K L K G E D L V K G I E Y L D A E E E E N A Y A P T H T H L E

8: Hybrid binding →

Table with protein names (Sce_rpb2, Ape_rpoB, Pae_rpoB, etc.) and their corresponding amino acid sequences. The sequences are aligned across columns. A consensus sequence is provided at the bottom: I P I L G V A A S I I P P D H N Q S P R N T Y Q S A M K Q A G L G V A N Y R I R L D S H S Y L L Q Y P O V P T Q T R M L D I M G Y N E R P A G O N M V V A I M S Y G Y N E D A L I N K G S I D M G L A R (Metal B-Motiv)

9: Wall →

Sequence alignment for region 9: Wall. Includes labels for various species (Sce_rpb2, Ape_rpoB, etc.) and a consensus line at the bottom.

↓ 10: Hybrid binding →

Sequence alignment for region 10: Hybrid binding. Includes labels for various species and a consensus line at the bottom.

↓ 11: Anchor →

↓ 12: Clamp →

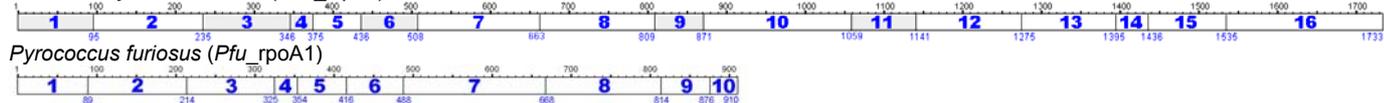
Sequence alignment for regions 11: Anchor and 12: Clamp. Includes labels for various species and a consensus line at the bottom.

Sequence alignment for region 13. Includes labels for various species and a consensus line at the bottom.

8.1.3.2 Untereinheit A' (rpoA1)

Die beiden A-Untereinheiten enthalten bei *M.jannaschii* jeweils ein Intein. Diese Proteine sind offensichtlich durch einen autokatalytischen Mechanismus in der Lage, sich korrekt aus dem „Wirtsprotein“ herauszuschneiden, wobei dessen verbleibenden Teile so zusammengefügt werden, dass trotzdem ein funktionsfähiges Protein entsteht (quasi eine extreme posttranslationale Modifikation). Bei rpoA1 ist das Intein 451 Aminosäuren lang. Bemerkenswert ist seine (im Alignment rot markierte) Position im hochkonservierten sensiblen Bereich direkt vor dem katalytischen Zentrum der A'-Untereinheit (ab Position 461). Auch die Zinkfinger-Motive wurden rot markiert.

Saccharomyces cerevisiae (Sce_Rpb1)



↓ 1: Clamp core →

```

Mka_rpoA1 ~~~~MTLMEPRVQDEPKYVVKRIIFGILFDPKIRKMSKVEVTSPEITYDEDDGYPIE.GGVMDTRMGVIDPGLRRTGGQAGRPGFGHIELARPVLHPYAEPRIKQVLR 105
Neq_rpoA1 ~~~~MTLMEPRVQDEPKYVVKRIIFGILFDPKIRKMSKVEVTSPEITYDEDDGYPIE.NGVMDPRMGVIDPGLRRTGGQAGRPGFGHIELARPVLHPYAEPRIKQVLR 95
Pae_rpoA1 ~~~~MSLREEFDAIEK.KVIRKSIKFGVLSPEIRKMSVMEVTSPEITYDEDDGYPIE.GGLADRRRLGVAEPGSRCEETGGQHDVPGFGHIELVKPVIHVGFARITIVVLR 104
Sso_rpoA1 ~~~~MSE.KNLIKIKFGILSPDEIRKMSVTAITPDVYDEDDGYPIE.GGVMDPRLGVIEPGQKPTGNTLGNPGFGHIELVLPVVIHVGFIVKHVYEFLEK 95
Sac_rpoA1 ~~~~MSE.KTIIRGVKFGVLSPEIRKMSVTAITSEVYDEDDGYPIE.GGVMDPKLVIEPGQKPTGNTLAGPGFGHIELIKPVIHIYGVKHIYDFLR 95
Sto_rpoA1 ~~~~MVRIMSE.KTIIRKIKFGILSPDEIRKMSVTAITSEVYDEDDGYPIE.GGVMDPRLGVIEPGQKPTGNTLNGPGFGHIELVLPVVIHVGFIVKHVYDFLR 99
Ape_rpoA1 ~~~~MIGAMSLR.LSEFRRT.NLIDKILFGVLSPEIRQLAKVIVVRCGLYEDDGYPIE.GGLDHPHFGVIEPGERPVGNRSREPGFGHIELARPVLPVYTDVYVYKILQ 107
Tac_rpoA1 ~~~~MMGHS.KRISISIKFALLSPDEIRKLSQVKVITADTYDDDDGYPIE.HGLMDLHMGVIEPGLRATGGKVDPEPGFGHIELAPVHVHGVFVKIKMFLD 97
Two_rpoA1 ~~~~MMGHS.KRISISIKFALLSPDEIRKLSQVKVITADTYDDDDGYPIE.HGLMDLHMGVIEPGLRATGGKVDPEPGFGHIELAPVHVHGVFVKIKMFLD 97
Mth_rpoA1 ~~~~MRGHL.KKISIQINFGILSPEDIRKMSVTAITPDYDEDDGYPIE.NGLMDPRLGVIDPGLRRTGAKGEGEPGFGHIELARPVVIHVGFADTIHKLR 97
Afu_rpoA1 ~~~~MVP.KRISAIKFEVLSPEIRKMSVTAITPDYDDDDGYPIE.GGLMDPRLGVIDPGLRRTGGKAGEPGFGHVELAAPVVIHVGYAKRIARLLN 95
Mac_rpoA1 ~~~~MANVPPHP.KRIASIKFGLMSPKEIRKMSAETIITADTYDDDDGYPIE.MGLMDLKLGVIDPGLRRTSGRAGEPGFGHIELVAPVHVGFNKKVIRKLLR 100
Mma_rpoA1 ~~~~MANVPPHP.KRIASIKFGLMSPKEIRKMSAETIITADTYDDDDGYPIE.MGLMDLKLGVIDPGLRRTSGRAGEPGFGHIELVAPVHVGFNKKVIRKLLR 100
Mja_rpoA1 ~~~~MERYEIP.KEIGSIFGLLSPDIYRQMSVAKIVTPDYDEDDGYPIE.HGLMDLKLGVIDPGLRRTGGRIAGEPGFGHIELAKPVIHIYFAKTIYKILK 99
Pab_rpoA1 ~~~~MOSVK.KVIESIEFGILSPQEIIRKMSAVEVVPDYDDDDGYPIE.GGVMDKRMGVIDPGLRRTGAKAGEPGFGHIELARPVVIHVGFARITIVVLR 97
Pho_rpoA1 ~~~~MHSVK.KVIESIEFGILSPQEIIRKMSAVEVVPDYDDDDGYPIE.GGVMDKRMGVIDPGLRRTGAKAGEPGFGHIELARPVVIHVGFARITIVVLR 97
Pfu_rpoA1 ~~~~MK.KVIESIEFGILSPQEIIRKMSAVEVVPDYDDDDGYPIE.GGVMDKRMGVIDPGLRRTGGRAGEPGFGHIELARPVVIHVGFARITIVVLR 94
Hsp_rpoA1 ~~~~MSAGOAP.KEIGSIFGLMSPPEIRKMSATKVTADTYDDDDGYPIE.MGLMDPRLGVIDPGLRRTGQRSGGNGFGHIELAPVVIHVGFARITIVVLR 99
Sce_rpb1 ~~~~MVGQYSSAPTRIVKIQVGLSPEEVRVLSVAKIREFPETMDETOFRAKGGGLNDPRLGSLIDRLKQOTQEGNNEPGFGHIDLAKPVHVGFARITIVVLR 104
Consensus ~~~~.K I I F G L S P E I R K M S . T D I Y D D G P I . G G M D R G V I D P G L R C T G G E C P G F G H I E L A P V H V G F K I L 62
  
```

... 2: Clamp head →

```

Mka_rpoA1 ATCPKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 175
Neq_rpoA1 ATCPKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 178
Pae_rpoA1 ATCPKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 183
Sso_rpoA1 ATCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 173
Sac_rpoA1 STCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 173
Sto_rpoA1 STCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 177
Ape_rpoA1 ATCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 185
Tac_rpoA1 ATCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 175
Two_rpoA1 ATCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 175
Mth_rpoA1 STCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 168
Afu_rpoA1 STCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 167
Mac_rpoA1 ATCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 172
Mma_rpoA1 ATCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 172
Mja_rpoA1 AVCPKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 172
Pab_rpoA1 STCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 170
Pho_rpoA1 STCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 170
Pfu_rpoA1 STCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 167
Hsp_rpoA1 STCRKCGRVIKLEDELESVIREVREGIRPVK...SVAAEVIDRAEKRECPHGAGBSRKITIQKPTIIMDD... 202
Sce_rpb1 CVCMKCGKLLL...DEHNEIMRQALAKDKSKR...FAALITWIKTKMVG...ETDVPSPDPPTQVLSRGCGCNTQPTIRKDGILKLVGSWKKDRATGDAD... 196
Consensus TC CGR L Y ..... A CPHG Q IK EKP E 82
  
```

↓ 3: Clamp core →

```

Mka_rpoA1 .....RLVMDIRAWLEKIPDEDEVEILGFHPERARPEAVLTVLPVPPVPMRPSIILETGERAEDDLTHKLV 242
Neq_rpoA1 .....KEEWVKLTPLDVROWLEKIPDEDEVEILGFHPERARPEAVLTVLPVPPVPMRPSIILETGERAEDDLTHKLV 248
Pae_rpoA1 .....KLPPEVIRERLERISKVPSEDELELLGNPSVARPEVAAILKVLVPPVTHVRSPTILETGERSEDLLTHKLV 250
Sso_rpoA1 .....KLTSPDIRERLERIKIPDSVDEILGYDPTTSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 240
Sac_rpoA1 .....KLTSPDIRERLERIKIPDSVDEILGYDPTTSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 244
Sto_rpoA1 .....KLTSPDIRERLERIKIPDSVDEILGYDPTTSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 244
Ape_rpoA1 .....KLTSPDIRERLERIKIPDSVDEILGYDPTTSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 251
Tac_rpoA1 .....KLTSPDIRERLERIKIPDSVDEILGYDPTTSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 242
Two_rpoA1 .....KLTSPDIRERLERIKIPDSVDEILGYDPTTSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 242
Mth_rpoA1 .....KLTSPDIRERLERIKIPDSVDEILGYDPTTSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 236
Afu_rpoA1 .....KLTSPDIRERLERIKIPDSVDEILGYDPTTSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 235
Mac_rpoA1 .....KLTSPDIRERLERIKIPDSVDEILGYDPTTSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 240
Mma_rpoA1 .....KLTSPDIRERLERIKIPDSVDEILGYDPTTSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 240
Mja_rpoA1 .....NEE...KLTSPDIRERLERIKIPDSVDEILGYDPTTSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 243
Pab_rpoA1 .....DEQNGEYKHRMPTDIRERLERIKIPDKDLLELLGLHPEKSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 246
Pho_rpoA1 .....DEQNGEYKHRMPTDIRERLERIKIPDKDLLELLGLHPEKSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 246
Pfu_rpoA1 .....DEQNGEYKHRMPTDIRERLERIKIPDKDLLELLGLHPEKSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 243
Hsp_rpoA1 AEQTDIDISRLINEIDKLSGEFRPRREDREATAIGADLTEDMNLKMPDIRDFEIPEDLEALGVNDSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 312
Sce_rpb1 PEI.....RVLSSTEEILNIFRHSIVKDFTSLGFNEVFSRPEWMILTTLVLPVPPVTHVRSPTILETGERSEDLLTHKLV 267
Consensus ..... L P I R R L E I P D D L G P R P E W M I T V L V P P V T R P S I L E T G R A E D D L T H K L V 129
  
```

↓ 4: Active site →

Mka_rpoA1 DIVRVNRLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 352
Neq_rpoA1 DIVRVNRLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 357
Pae_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 359
Sso_rpoA1 DIVRVNRLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 349
DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 349
Sto_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 353
Ape_rpoA1 DIVRVNRLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 360
Sac_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 351
Tvo_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 351
Mth_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 345
Afu_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 344
Mac_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 349
Mma_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 349
Pab_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 352
Pho_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 355
Jfu_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 352
Hsp_rpoA1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 421
Sce_rpb1 DIIRINRNLKESLEAGAPEPIEDWDLWLLQYHVTTYFDNEIPGVVVARHRTGCKPLKGVORLKGKGEFRONLAGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 376
Consensus DIIRIN RL E EAGAPQLIEDLW LLQYHV TY DNE G PPA HR. SGRPLKTL QRLKGKGEFRGNLSGKRVNFSARSVITPDPPELDPFLGVPEITAKELTV 220

↓ 6: Active site (Metal A) →

Mka_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....DGSRRITDENKERKLAEDLEPGDVVERHLMGDGPFVFNRPQLHICSELCGRVRVMPGKTFRL 447
Neq_rpoA1 PERVTEWNIIDLRLKLVLLGNPES.....LDGANVVID.....PNGR.....RKIMEENKEELAERLEPGWVVERHINDYALFNRPQSLHRMSMMGHVRIILPGKTFRL 452
Pae_rpoA1 PERVTEWNIIDLRLKLVLLGNPES.....PGANYVVRK.....PEGRR.....IDRLRVKDRKE.LAERLAPGWIVERHLDIGDIVLFRNRPQSLHRMSMMGHVRIILPGKTFRL 455
Sso_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....PDGRR.....IDRLRVKDRKE.LASTIAPGYIIVERHLDIGDIVLFRNRPQSLHRMSMMAHVRVRLPGKTFRL 445
Sac_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....PDGRR.....IDRLRVKDRKE.LASSITAGYVVERHLDIGDIVLFRNRPQSLHRMSMMAHVRVRLPGKTFRL 445
Sto_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....PDGRR.....IDRLRVKDRKE.LASTIAPGYIIVERHLDIGDIVLFRNRPQSLHRMSMMGHVRIILPGKTFRL 449
Ape_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGALFVYK.....ASRKK.....YDLRFKDYK.LAESLEPGDVVERHLDIGDIVLFRNRPQSLHRMSMMGHVRIILPGKTFRL 457
Tac_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....PDGR.....RIRKIDONAEANADRIDIGWTVVERHLDIGDIVLFRNRPQSLHRMSMMGHVRIILPGKTFRL 453
Tvo_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....PDGR.....RIRKIDONAEANADRIDIGWTVVERHLDIGDIVLFRNRPQSLHRMSMMGHVRIILPGKTFRL 453
Mth_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....PDGR.....KIRINNETKEVVEENLKPQYIIVERHLDIGDIVLFRNRPQSLHRMSMMAHVRVRLPGKTFRL 440
Afu_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....PDGR.....RIRVLESNKEELAERLEPGWVVERHLDIGDIVLFRNRPQSLHRMSMMAHVRVRLPGKTFRL 437
Mac_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....SDGR.....RIRKIDONAEANADRIDIGWTVVERHLDIGDIVLFRNRPQSLHRMSMMAHVRVRLPGKTFRL 444
Mma_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....SDGR.....RIRKIDONAEANADRIDIGWTVVERHLDIGDIVLFRNRPQSLHRMSMMAHVRVRLPGKTFRL 444
Pab_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....DPEGR.....RIRIMESNKEELAERLEPGWVVERHLDIGDIVLFRNRPQSLHRMSMMAHVRVRLPGKTFRL 450
Pho_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....DPEGR.....RIRIMESNKEELAERLEPGWVVERHLDIGDIVLFRNRPQSLHRMSMMAHVRVRLPGKTFRL 450
Pfu_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....DPEGR.....RIRIMESNKEELAERLEPGWVVERHLDIGDIVLFRNRPQSLHRMSMMAHVRVRLPGKTFRL 447
Hsp_rpoA1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....PDGR.....RIRVLESNKEELAERLEPGWVVERHLDIGDIVLFRNRPQSLHRMSMMAHVRVRLPGKTFRL 516
Sce_rpb1 PERVTEWNIIDLRLKLVLLGNPKH.....PGANYVVRK.....DSCDRIDLRLKRVKRAED.....DQYGVVERHLDIGDIVLFRNRPQSLHRMSMMAHVRVRLPGKTFRL 470
Consensus P VT NI R V NGPPGANYV...P GR... R KE LA L GW VERHL DGDIVLFRNRPQSLHRMS MAH VRVLPYTFRL 282
(RQP (TS) LH-Motiv)

↓ 7: Pore 1 →

Mka_rpoA1 P.AACVMGVCENADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...D 556
Neq_rpoA1 NPAT...PTPNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...D 536
Pae_rpoA1 HLAVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...D 539
Sso_rpoA1 NLLVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...D 529
Sac_rpoA1 NLLVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...D 529
Sto_rpoA1 NLLVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...D 533
Ape_rpoA1 NLLVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...D 541
Tac_rpoA1 NLAVC...TPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...E 537
Tvo_rpoA1 NLAVC...TPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...D 537
Mth_rpoA1 NLLVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...K 522
Afu_rpoA1 NPAVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...E 518
Mac_rpoA1 NPAVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...K 526
Mma_rpoA1 NPAVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...K 526
Mja_rpoA1 NLLVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...G 538
Pab_rpoA1 NLAVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...G 534
Pho_rpoA1 NLAVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...G 534
Jfu_rpoA1 NLAVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...G 531
Hsp_rpoA1 NLLVC...PPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...D 598
Sce_rpb1 NLSV...TPYNADFDGDEMNLHVQSEEARAEAEALMVEK.HIITPRYGGPII.GAQQDYISGAYLLT.....KTFLLTKEEVAQITIGVA...D 557
Consensus NL VC...PPYNADFDGDEMNLHVQ EEA AEA LM V I SPR GGP II .G I V D L C G Y R K L T L R D T F I E L D Q V L N M I Y W V P D 336
(katalytisches Zentrum)

Mka_rpoA1 FEEELHELFVCCOPREGEPIVTGRQLLSVFLPDLDRVLEENAVCPDDEBHGK.....VVIEKGRIVRCFLDEKVTWHGIMTIVHRYAA 641
Neq_rpoA1 NYGKDKIKLVKAKREGDYLTGKEVFSALLPDDLDFEIKTK.....NGEFV.IKNG.LLKGVIDKEITLGVKRGKILRELF 613
Pae_rpoA1 S...QDDP...PEPAILHPVELWTGKQIISHLPKDLN.VVOPTAFKS...K.CONA.....YTCPDEWIIIVNGNLVKGVLDKKISGAEQVDSLWHRIA 624
Sso_rpoA1 V...KIDL...CEPAILAPREYTGKQVISAFLPKDFN.FHGQANISSGERRL.CKNE.....DCHPDSYVIKNG.LLLEGVLDKKAIGNQOPELHLWLI 616
Sac_rpoA1 F...VCEL...CEPAILAPKYTGKQVISAFLPKDFN.FHGQANISSGERRL.CKNE.....DCHPDSYVIKNG.LLLEGVLDKKAIGNQOPELHLWLI 616
Sto_rpoA1 F...MGEI...CEPAILAPREYTGKQVISAFLPKDFN.FHGQANISSGERRL.CKNE.....VCPHDSYVIKNG.LLLEGVLDKKAIGNQOPELHLWLI 620
Ape_rpoA1 Y...KIDL...PEPAILHPVELWTGKQIISHLPKDLN.FKGSKNTAGDLA.CSDE.....LCTHDSYVIKNG.LLLEGVLDKKAIGAEENLHLHITA 628
Tac_rpoA1 P...DIL...PEARIENGEKYYVGRDIFSTILLPKELN.VFRSKLCSGSSERCEBED.....DPSDHYVEIVDGRMIEGCHIDEAAVSPHSGA.IIDKIF 623
Tvo_rpoA1 P...DIL...PDATVINGKYYVGRDIFSTILLPKELN.VFRSKLCSGSSERCEBED.....EKDHYVEIVDGRMIEGCHIDEAAVSPHSGA.IIDKIF 623
Mth_rpoA1 A...GLPLPDSRGRD...WTGKEIFSMVLPDDL.NVYRAEVCRCR.ECELEMEC.....ENDAYVVIENGKLLSGVIDEKAIGAFAGK.ILDHIV 604
Afu_rpoA1 ...SIDSEIPVQDEY...TGKEIFSFILPD.ITLVYRAEVCRCR.ECELEMEC.....ENDAYVVIENGKLLSGVIDEKAIGAFAGK.ILDHIV 599
Mac_rpoA1 S...RARNLPEPAGHDEGKPFWTGKQIIFSLPDLN.MEFAAQVCSHC.DTCKKEEC.....PNDAYVVIENGKLLSGVIDEKAIGAFAGK.ILDHIV 615
Mma_rpoA1 S...SARSLEPEPAGHDEGKPFWTGKQIIFSLPDLN.MEFAAQVCSHC.DTCKKEEC.....PNDAYVVIENGKLLSGVIDEKAIGAFAGK.ILDHIV 615
Pab_rpoA1 I...K.DELWEPDK.VENGWPLYSGKIFSKALPKELN.VVYRNLKDEP.EKCEAEKLEIEKLPDPEEVRLAYDCGVVIQNGKLLSGAIDKKAIGRDEG.IILDIV 625
Pho_rpoA1 V...DIKELPEPDKY.ENGKPLWSGKIFSLPDLN.VVYRNLKDEP.EKCEAEKLEIEKLPDPEEVRLAYDCGVVIQNGKLLSGAIDKKAIGRDEG.IILDIV 638
Jfu_rpoA1 V...DIKELPEPDKY.ENGKPLWSGKIFSLPDLN.VVYRNLKDEP.EKCEAEKLEIEKLPDPEEVRLAYDCGVVIQNGKLLSGAIDKKAIGRDEG.IILDIV 638
Pfu_rpoA1 V...DIKELPEPDKY.ENGKPLWSGKIFSLPDLN.VVYRNLKDEP.EKCEAEKLEIEKLPDPEEVRLAYDCGVVIQNGKLLSGAIDKKAIGRDEG.IILDIV 636
Hsp_rpoA1 T...RIDELPAACTDEDCDYWTGQIFSELLPDDL.LLEFCTTGTDT.....VVIEDCOLLEGCHIDEAAVSPHSGA.IIDKIF 673
Sce_rpb1 G.....VIETPATIKKPLWSGKIFSLPDLN.VVYRNLKDEP.EKCEAEKLEIEKLPDPEEVRLAYDCGVVIQNGKLLSGAIDKKAIGRDEG.IILDIV 634
Consensus .. TGK IFS LP L L C D V I N G I G D K A G G I I 360

8: Funnel

Table with 3 columns: Species (e.g., Mka_rpoA1, Neq_rpoA1), Amino Acid Sequence, and Position (751-999). Includes a Consensus line at the bottom.

9: ...

Table with 3 columns: Species (e.g., Mka_rpoA1, Neq_rpoA1), Amino Acid Sequence, and Position (860-999). Includes a Consensus line at the bottom.

9: Cleft

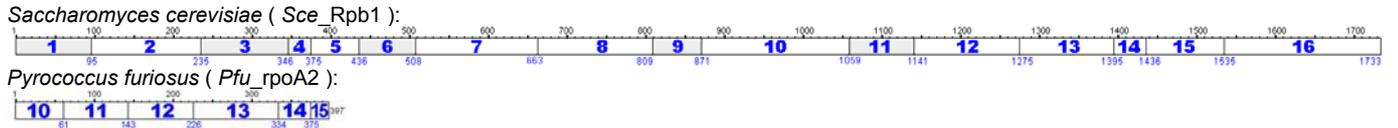
10: Foot

Table with 3 columns: Species (e.g., Mka_rpoA1, Neq_rpoA1), Amino Acid Sequence, and Position (953-1000). Includes a Consensus line at the bottom.

Table with 3 columns: Species (e.g., Hsp_rpoA1, Sce_rpb1), Amino Acid Sequence, and Position (971-1683). Includes a Consensus line at the bottom.

8.1.3.3 Untereinheit A'' (rpoA2)

rpoA2 aus M.jannaschii enthält ebenso wie rpoA1 ein Intein (in diesem Fall 471 Aminosäuren lang). Interessanter Weise ist auch hier eine ansonsten hochkonservierte Region Ziel der Insertion (ab Position 76).



```
Sc_rpb1, MVGQQYSSAPLRTVKEVQFLFSPEEVRRAISVAKIRFPETMDETQTRAKIGGLNDPRLGSDIRNLKCOQCQEGMNECPGHFGHIDLAKPVFHVGFIAKIKKVCCEVCMHC 110
Sc_rpb1, GKLLLDHNEMLRQALAKDSKRRFAAIWTLCKTKMVCETDVPSEDDPTQLVSRGGCGNTQPTIRKDGKLVGSKKDRATGDADEPELRLVLTSTEEILNIFKHI SVKDF 95
Sc_rpb1, SLGFNEVFSRPEWMILTCPLPVPVPPVRSISFNESQRGEDDLTKFLADILKANISLETLEHNGAPHAIEEAESLQFHVATYMDNDIAGQQALQKSGRPVKSIRARLK 330
Sc_rpb1, GKEGRIRGNLMGKRVDFSAITVSGDPNLELDQVGVPKSIKTLTYPEVVTYPYNDRLTQLVNRNGNEHPGAKYVIRDSGDRIDRLYSKRAQIQIQYQWVKVERHIMDND 440
Sc_rpb1, PVLFNRPQSLHKMSMAHRVKVPIYSTFRNLNSVTSFYNAFDGDEMNLHVPQSEETRAELSQLCAVPLQIVSPQSNKPCMGIVQDTLCCGIRKTLTRDTFIEDQVLNML 550
Sc_rpb1, YWVPDWDGVIPTPAIKPKPLWSGKQILSVAIPNGIHLQRFDEGTTLLSPKDNGLMIIIDGQIIIFGVVEKKTGVSSNGGLIHVVTRKGPQVCAKLFNGIQKVVNFLLHN 660
Sc_rpb1, GFSTGIGDTIADGPTMREITETIAEAKKVLVDVTKEAQANLLTAKHGMLTRESFEDNVVRFNLNEARDKAGRLAEVNLKDLNNVKQVMVAGSKGSFNIAQMSACVGGQSV 770
Sc_rpb1, EGKRIAFGVDRITLPHFSKDDYSPEKSGFVENSYLRLGTLPEQFFHAMGGREGLIDTAVKTAETGYIQRLVKALEDIMVHYDNTTRNSLGNVIQFIYGEDGMDAAHIEK 880
```

10: Foot

```
Sc_rpb1, QSLDITGGSDAAFEKRYRVDLLNTDHTLDPSLLESG...SEILGDLKLOVLLDEEMKQLVKDRRFLREVFVDFGEANWLPVN...IRRIIQNAQOTF 971
Tac_rpoA2, ~~~~~MKSMASSLWRDTSKNIAAILLEKLPADYAVDVPNNVREDGYITINK.KNFTYHVWISGVVKYSPDVEAIVRKRKSGIKSITITBEKVEKIEPLSFMEF 95
Mth_rpoA2, ~~~~~MILTEVIWRDTAKNMSLLSKSPAKYAVDFEVPKGTIEGVVTSK.KRFTYHVWISSVAPYSNESDVIQKRSKGSIMEIKIKQIEPISIMEF 95
Sto_rpoA2, ~~~~~MODIICKIEDYSS.....KNGILPDPVVEYVARIAD...E 34
Mka_rpoA2, ~~~~~ 1
Neq_rpoA2, ~~~~~ 1
Pae_rpoA2, ~~~~~ 1
Ape_rpoA2, ~~~~~ 1
Hsp_rpoA2, ~~~~~ 1
Afu_rpoA2, ~~~~~ 1
Mac_rpoA2, ~~~~~ 1
Mma_rpoA2, ~~~~~ 1
Mja_rpoA2, ~~~~~ 1
Pab_rpoA2, ~~~~~ 1
Pho_rpoA2, ~~~~~ 1
Pfu_rpoA2, ~~~~~ 1
Sso_rpoA2, ~~~~~ 1
Sac_rpoA2, ~~~~~ 1
Consensus, ~~~~~ 1
```

11: Cleft

```
Sc_rpb1, HIDHTKPSDL.....TIKDIVLGVKDLQENLLVLRGKNEIHONAQRDVAVTFCCLLRSLATRRV...LOEYRLTKQAFDVLNSIEAQLRSVVIPGEMVGVIAA 1069
Tac_rpoA2, RVGKTLAEMGSF...EVAERQVTEIKKCYC...ENLSEDOVQVDDARAMGTLPLPSVAEIA.....RRRETEGCKAYKNILKRIEETIENGLIDPYEAVGIIAA 193
Mth_rpoA2, RSSGKRIDELLYT...ATAERETAETREKRYEYKRVSSQVLDVLAEANKLGNIPESVAEIT.....RRRETEGCKAYKNILKRIEETIENGLIDPYEAVGIIAA 193
Sto_rpoA2, KLKPEPELQEMVRLFSRIEARNQGLDDDLDAVEEDQYRILKVEITVKRKRFRFPEKLIEDIAEVMKKH...ELSDDELDLIRRVRAVDRARVEAGEAVGVAA 137
Mka_rpoA2, ~~~~~MEEVNVMVSDYIKSYOIERSLGQIRPKLDDKVLQDIREALLNTD...INLTFSDVDKIIDLAVKDYQSEALVPEAGEAVGVAA 78
Neq_rpoA2, ~~~~~MNVEELLEEVCKEEDVELPDSVHREIVEKVSDD...BELKDADEETVRKVLRAIVRNKRCQIEIRGDAVGVIAA 71
Pae_rpoA2, ~~~~~MSILPEKILQEFEBMAKK.....YKIRKCKEQLLKKLEEKVLSLIDPGEPIGIITA 53
Ape_rpoA2, ~~~~~MISREBELSKLSQVLPQPLVKEVEEAVR...DLDEKALRVLVVRVKYVTSLIDPGEAIGVTA 62
Hsp_rpoA2, ~~~~~MEGVKFEKTLLESLEAARYI...LPESLYKELVETVEKED.....GLSEEDKISVVKETIRIYLRSLACPEAVGVIAA 71
Afu_rpoA2, ~~~~~MTEFDVVDGVDHVDIEDLPRRLKDEVLAT.....AEERGVTKSQANETATAVEAGYLDTRVDPLDPVGVVSA 68
Mac_rpoA2, ~~~~~MSKYDYSILLKDEFPFPVAVKEEIKAELEKH.....GLKKTAEAKKVVKEKCFQAYLANLMEPEGAAGVIAA 64
Mma_rpoA2, ~~~~~MSEATIDSMILKDLFPSPNIIITLREDVKA.....GVSKKEEETIERMEEYATSCIEPCDAAGVIAA 64
Mja_rpoA2, ~~~~~MMSISEATVDSMIKDLPLPSNIIKTRDDVKA.....GVSKKEEETIERMEEYATSCIEPCDAAGVIAA 67
Pab_rpoA2, ~~~~~MDMEALKQKIE...GLDIPQSLKDEIKSEKSK...EKD...LTEEMVDEIIDEVVMYAKRALVPEAGEAVGVIAA 65
Pho_rpoA2, ~~~~~MVSSTIKSLIEK.KGKDLPEVSKQELYEKLIK...YNEKYKLTAEVETIIDEVVKYERALVPEAGEAVGVIAA 71
Pfu_rpoA2, ~~~~~MVSSTIKSLIEK.KGANLPEENVKSELYEKLIK...YNEKYKLTAEVETIIDEVVKYERALVPEAGEAVGVIAA 71
Sso_rpoA2, ~~~~~MEEMTD.EKDKPYLEE.KVKQASNIIQKIVDLDLKNLIIKREIT...VTREIDIKIDILAIKEYSEGLIPEAGEAVGVIAA 75
Sac_rpoA2, ~~~~~MID.EKIKCYIDK.RLNEIKDKIPDKIHEDIRAAIMDINGVELTEEDIDRIIDILIREYQSLIPEAGEAVGVIAA 73
Consensus, ~~~~~ P . . . . . I Y L PGEAV VAA 13
```

12: Jaw

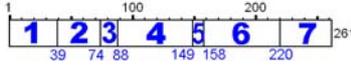
```
Sc_rpb1, QSIGEPATQMTLNTFHFAGVAASKRVTSQVPRKKEILNVA.KNKTPSLTVYLEPGEA...ADQEQAKLRSAIEHTTKSVTIASIEIYDPPDRSTVPIPEDEEIIQLHF 1174
Tac_rpoA2, QSIGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAR.RIPSTPSMTIYLPE...YEYNDEVVMVVKRLENTSISDVADIIT...DIGELTVTVRPPDKK... 290
Mth_rpoA2, QSIGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAR.RIPSTPSMTIYLKPE...EETNDEVVMVVKRLENTSISDVADIIT...DIGELTVTVRPPDKK... 292
Sto_rpoA2, QSVGEPGTMTRTFHFAGVAELNVTGLPRLIEIVDAR.KITSTPTMSIYEFGLDLYEDEFVRRKAN...KIGKSTLNDVKNSISI...QVADMSVVAELDEEKIQ... 237
Mka_rpoA2, QSIGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAR.KIPSTPTMTIYLDE...YKNDREKALEIARKIEYTKIENVSSISV...DIYNNMSITIQDDEMLK... 178
Neq_rpoA2, QSMCEPATQMTMTRTFHFAGVAEIDITGMPALERVLSLVYHCFPTVWIEIRLKEP...YRSREKAEQVARKIEETREBELTEYIEV...NPDECATIEILDEERIE... 172
Pae_rpoA2, QSIGEPGTMTRTFHFAGVSELQVLTGLPRLIEIINAK.KKPTPTPSMTIYLEKN...DEKTAELVNLKIEVKLSIAKTSISI...DVINKRIIIELDQOLVK... 150
Ape_rpoA2, QSIGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAR.KIPSTPTMTIYLKPE...YKNDREKAEVAKKIQOVTLTLAKAEV...DYVAGTFTIHDQOLK... 162
Hsp_rpoA2, QSIGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAK.QTPSPLMIIYLKDE...YAKDIEKAEVARKIEYTTLEKIINDIEM...DLGDRVVAIVINEBYME... 171
Afu_rpoA2, QSIGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAR.KTPSTPTMTIYLKPE...YAEERREAEVAVVVKIEATKILALQD...IST...NVADMVVRIDLNEDTLQ... 167
Mac_rpoA2, QSIGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAR.KIPSTPTMTIYLKPE...YAKDREKAEVARKIEATYVVDVAD...IEV...DIRREFTIIVPEDEKALE... 163
Mma_rpoA2, QSIGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAR.KIPSTPTMTIYLKPE...YAKDREKAEVARKIEATYVVDVAD...IEV...DIRREFTIIVPEDEKALE... 167
Mja_rpoA2, QSIGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAR.KIPSTPTMTIYLKPE...YAKDREKAEVARKIEATYVVDVAD...IEV...DIRREFTIIVPEDEKALE... 170
Pab_rpoA2, QSIGEPSTQMTLNTFHFAGVREMNVTGLPRLIEIVDAR.KNPSTPMNTIYLDEE...HYDREKAEVARRIEGTTLENLARTTTL...DLINMEFTIVPEDEKALE... 171
Pho_rpoA2, QSIGEPSTQMTLNTFHFAGVREMNVTGLPRLIEIVDAR.KNPSTPMNTIYLDEE...HYDREKAEVARRIEGTTLENLARTTTL...DLINMEFTIVPEDEKALE... 171
Pfu_rpoA2, QSVGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAK.KIPSTPTMTIYLKPE...YKNDREKAEVARRIEGTTLENLARTTTL...DLINMEFTIVPEDEKALE... 171
Sso_rpoA2, QSVGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAK.KIPSTPTMTIYLKPE...YKNDREKAEVARRIEGTTLENLARTTTL...DLINMEFTIVPEDEKALE... 175
Sac_rpoA2, QSVGEPGTMTRTFHFAGVREMNVTGLPRLIEIVDAR.KIPSTPTMTIYLKPE...YKNDREKAEVARRIEGTTLENLARTTTL...DLINMEFTIVPEDEKALE... 173
Consensus, QSIGEPGTMTRTFHAGVREMNVTGLPRLIEIVDAR.KIPSTPTMTIYLKPE...YKNDREKAEVARRIEGTTLENLARTTTL...DLINMEFTIVPEDEKALE... 71
```


8.1.3.4 Untereinheit D (rpoD)

Saccharomyces cerevisiae (*Sc_e_rpb3*):



Pyrococcus furiosus (*Pfu_rpoD*):



Das Zinkfinger-Motiv wurde rot markiert. Bei *S.cerevisiae* und 9 *Archaea* findet sich ein komplettes Motiv, bei 2 *Archaea* Teile des Motives und bei 6 *Archaea* ist das Motiv vollständig abwesend.

Sequence alignment for domains 1: Dimerization, 2: Domain 2, and 3: Zink loop. Species include Neq_rpoD, Mja_rpoD, Tac_rpoD, Tvo_rpoD, Pae_rpoD, Pfu_rpoD, Pab_rpoD, Pho_rpoD, Hsp_rpoD, Mka_rpoD, Sac_rpoD, Sto_rpoD, Sso_rpoD, Mth_rpoD, Afu_rpoD, Mma_rpoD, Ape_rpoD, Sc_e_rpb3, and a Consensus line. Red boxes highlight zinc finger motifs.

Sequence alignment for domains 4: Domain 2, 5: Dim, 6: Loop, and 7: Dimerization. Species include Neq_rpoD, Mja_rpoD, Tac_rpoD, Tvo_rpoD, Pae_rpoD, Pfu_rpoD, Pab_rpoD, Pho_rpoD, Hsp_rpoD, Mka_rpoD, Sac_rpoD, Sto_rpoD, Sso_rpoD, Mth_rpoD, Afu_rpoD, Mma_rpoD, Ape_rpoD, Sc_e_rpb3, and a Consensus line.

Sequence alignment for domains 7: Dimerization and 8: Tail. Species include Neq_rpoD, Mja_rpoD, Tac_rpoD, Tvo_rpoD, Pae_rpoD, Pfu_rpoD, Pab_rpoD, Pho_rpoD, Hsp_rpoD, Mka_rpoD, Sac_rpoD, Sto_rpoD, Sso_rpoD, Mth_rpoD, Afu_rpoD, Mma_rpoD, Ape_rpoD, Sc_e_rpb3, and a Consensus line.

Sc_e_rpb3, SNEDVMMTGAEQDPYSNASQMGNTSGGGYDNDW 318

8.1.3.5 Untereinheit E' (rpoE1)

Sce_rpb7, ~MFEIKDLISLNITLPPSEF.GPRMKQYLKTKLLEVEGESCCTCKFG.....YILCVLDVDNIDIQGRILPDDGSAEENVKVRAVVFKPFKGEVVDGTVVSCSOHG 98
Ape_rpoE1 MYEYVVEEWTIEMEPASTIYSSSEDLDKAAELELR.LEGRIDSEMG.....VIVAVLDAKIVGDGHIPPVSGDESIIYEVRYEVLAFKPVQLIEIVRGVVFTRADDFG 99
Neq_rpoE1 ~MFEISEVEDVVRVDPDLDKETHIEBAIKNALIKOYEGKINKOLGW.....VLSILDILDKSEGL..CYNDGCIYYKVRFKVLWKPILKEVTLGITSNIVDFG 97
Taco_rpoE1 ~MYMLVDEKYIARVPPML.GEDYVAVREVTKRSLEGRLVDVINIKTSQNEGKYLVIISVLYTESIGEGHI..VHGDGGEVYQEVKRYALVYVPEMQEIVDGEVVSVKFG 106
Tvo_rpoE1 ~MYMLVDEKYIARVPPML.GEDYQKAVEEVKRSLEGRLVDVINIKTSQNEGKYLVIISVLYTESIGEGHI..VHGDGGEVYQEVKRYALVYVPEMQEIVDGEVVSVKFG 106
Pae_rpoE1 MFRLVLEEDVVRVPPTEF.GKPLEDTALSOLKSYEGKVLRE.....VG...YIVAVLVNKKVRECKI..VFGDGCYHRAVFMILFMPDLCEVVEGVEVNAAREMG 97
Mka_rpoE1 ~MYELVELETVIRIPPTOF.TSDVEDAVLKALENDVEGKLRDDTCEEGIG...EVVVEDEILDVENEGI..FPGDGASVHRVRFALVFRPVEREVVVEGVEVTRVREYG 102
Sso_rpoE1 ~MYKLKARISIVRIPPNEF.GKPLNEITALNELRQOYQKLIKD.....LG...LVLAAILNVKTSSEGL..VFGDGCATYHEVDFEMLIFVYVPPVQEVVEGEVLQVDNYG 96
Sto_rpoE1 ~MFKLIKAKCIIIRIPPEYF.GGPLDETALQILRQEQKMKID.....LG...LVLAVLDKVSSEGI..IFDGCATYHEVDFEMILFVPIIQEVVEGEVVOVDNYG 96
Hsp_rpoE1 ~MYKARLKDILTEVPEYL.GAVGPDIVRLLQDKLEGKRDDE.....VG...SIVTVLEVHDLGEGAV..LPNRPGVYIEAEFDVAFPEMQEIVDGEVVEVVSFG 96
Afu_rpoE1 ~MYARIKLRDVRVPPSKL.GEDIEKVINSLLEQEQEGRIDRE.....VG...MIEGIESIEEIEGEGRI..IEGDGGEYFDVVFNAICFRPLHQEVVEGEVVIETVEFG 96
Mth_rpoE1 ~MYVYTKIVDVRIPPRDF.EPELEEVAVEVLENEVYVGRMDKN.....LG...OLIVVRELEDIGIKV..IMGDGCATYHEVDFEMLIFKPELHEIVEGEVVIETVEFG 96
Mja_rpoE1 ~MYKILETADVVKVPPPEF.GDKLKEIVKILMEKYEGRDKD.....VG...EVLISIVDKDIGEGV..VHGDGSAHYHVEVFEVLVYIPELMELIEGEVVDVVEFG 96
Pab_rpoE1 ~MYKLKARISIVRIPPNEF.TMDPEKAAMLVLRDYEGKYDKD.....EG...VILSIVEVKEVGDGHI..VFGDGCATYHEVDFDLVWEPKQOEVVEGTVVDAVVEFG 96
Pho_rpoE1 ~MYKIVTVKDVRIPPRME.TMDPEKAAMLVLRDYEGKYDRD.....EG...VILSIVEVKEVGDGHI..VFGDGCATYHEVDFDLVWEPKRDVVEGTVVDDVVEFG 96
Pfu_rpoE1 ~MYKIVTVKDVRIPPRME.TMDPEKAAMILLREYEGKYDKD.....EG...VILSILEVKDIKDGHI..IFDGCATYHEVDFDLVWEPKIEEVVEGTVVDDVVEFG 96
Mac_rpoE1 ~MYKLMKIDVRIPPRLL.GEEMPTIKNALREKLEGQVDDK.....LG...SEVAVYNDIEVGECHI..LVGDGAVYDVFEMIFEMIPQLQEIIEGEVVEAVVVEFG 96
Mma_rpoE1 ~MYKLMKIDVRIPPRLL.GEEMPTIKNALREKLEGQVDDK.....LG...SEVAVYNDIEVGECHI..LVGDGAVYDVFEMIFEMIPDLQEIIEGEVVEAVVVEFG 96
Consensus ~MY D VR PP .G L EG G.... GEG I.. GDG Y V F L F P EVVEGEV FG 34

Sce_rpb7, FEVQVQPMKVYVTKHLMPCD..LTFNAGSNPPSYQSSSEVITIKSRIRVKIEGCTISQVSSIIHAIG.....SIKEDYLCAL~~~~~ 171
Ape_rpoE1 LVVNLGPLDGLVHRNQIMDEG.VELLPPDRSGFVGLASKREIRVGDVVRARVVEVS...RPSRTRQITIGLTMROPMLGKEEWWAEEAKK~~~~~ 186
Neq_rpoE1 AYVNIQPDGLLHISQIMDD.IVDVYS.KDKVLACRKTTKKILKIEDIVKAKIVGTSIK....PIGTKNLTMROGLGKLOWLVEEKAQSE~~~~~ 181
Taco_rpoE1 AFIKFSSPEGLLHISQIMDDR.IDIDLENKRFIGKETKDLKVGDEVRVRIVAMNLASSISDS...KIGFTMKQVGLGKLDWLAKESQ~~~~~ 192
Tvo_rpoE1 AFVRFSSPEGLLHISQIMDDR.IDVDLENKRFIGKDTKRDLVKVGDEVRVRIVALNLASSVEDS...KIGFTMKQVGLGKLEWLEKEAQA~~~~~ 192
Pae_rpoE1 MIVRIQPDGLLHISQIMDDPNIFDASTKSYIGERTKRRVSVGDDVVRARIVCVSETPREGDILLIRVILTMRSPLGKLEWVKEKKPAAKA~~~~~ 191
Mka_rpoE1 AFVRLGPLDGLLHISQIIDE.YMYVDGAREALVGETGRELKRGDVIKVIITGVSNREPRRDS...KIALTVKRPGLGKEEWE~~~~~ 183
Sso_rpoE1 IFVNLGPMDGLVHISQITDD.TLKYDNRGILIEGKSKKVIQKGDVVRARIVASVATV.TCRIP...RIALTMROPMLGKLEWITQKK~~~~~ 180
Sto_rpoE1 YVYNIQPDGLLHISQITDD.NLRFDONRCHLIGERSKKIQKGDVVRARIVSV.STS.GGRMP...RIALTMROPMLGKLEWISQELTRASK~~~~~ 183
Hsp_rpoE1 AFVRIQPDGLLHISQISIDE.YLAFDEENQQLASRESNQVLTGDSVVRARIVSISIDENRPRE...KIGLTMROPGLGKHLQWLOADRECADAAASE~~~~~ 190
Afu_rpoE1 AFVSIQPDGLLHISQVITDD.YMVYDEKMKRLVGRETKKVLKIGDVRARIVSISLKERPEPKS...KIGLTMROPMLGKLEWIEEETKMKRESGV~~~~~ 189
Mth_rpoE1 AFVRIQPDGLLHISQVITDD.YIITYDPKRGSLTGKESGRRLDEGDDVVRARIVALSILKGRREGV...KVGLTMROPGLGKEEWEIEEKRSK~~~~~ 184
Mja_rpoE1 SFVRLGPLDGLLHISQIMDD.YVSYDPKREALIGKETGVLEIGDVRARIVASILKARRKCS...KIALTMROPMLGKLEWIEEKAQNOE~~~~~ 187
Pab_rpoE1 AFVRIQPDGLLHISQIIMDD.YVVYDERNKQFVGEKPKLLKIGDVRARIVSISLKERPEPKS...RIGLTMROPMLGKLEWIEEIEKREKREK.CKK~~~~~ 188
Pho_rpoE1 AFIRIGPMDGLVHISQIIMDD.YVVYDERNKQFVGEKPKLLKIGDVRARIVSISLKERPEPKS...RIGLTMROPGLGKEEWEIEKREKREK.CKK~~~~~ 188
Pfu_rpoE1 AFIRIGPDGLLHISQIIMDD.YVVYDERNKQFVGEKPKLLKIGDVRARIVSISLKERPEPKS...RIGLTMROPGLGKEEWEIEKREKREK.CKK~~~~~ 189
Mac_rpoE1 YVYCMGPMDGLLHISQITDD.FISYDAKNARLVTRNCGKSTAEAGDVRARIVAVSINEREPKES...KIGLTMROPMLGKLOWLEBARRKQPOEAPKKEAA 194
Mma_rpoE1 YVYCMGPMDGLLHISQITDD.FISYDAKNARLVTRNCGKSTAEAGDVRARIVAVSINEREPKES...KIGLTMROPMLGKLOWLEBARRKQPOETPGEEAA 194
Consensus FV GP DGL H SQI DD. D G K L GD VRARI S ... IGLTMROP LGK EW E ~~~~~ 73

8.1.3.6 Untereinheit E'' (rpoE2)

Zu rpoE2 gibt es bei Eukarya (S.cerevisiae) kein Homologes Protein. Das Zinkfinger-Motiv wurde rot markiert.

Ape_rpoE2 ~~~~~MFRGSRVPECGSTQFTENWSGMLVIIDPENSIAIARELGIERPVRRKAILICKKVMII 56
Taco_rpoE2 ~~~~~MKVQYRACKKCKRRLT..PEKTPVHCEDEKTTTEWEGELLITEPELSAIAKRTGITEPGYAIKVRQ~~~~~ 64
Tvo_rpoE2 ~~~~~MKVQYRACKKCKRRLT..PEKTPPIHPEKTTTEWEGELLITEPELSAIAKRAGISEPGYAIKVRQ~~~~~ 64
Mka_rpoE2 ~~~~~MSKTKAVRCRCYLVVEEDTEIIPACHGDEFTENWRGIAVILDE.SQIADRNLNAKIPGKYALRVEE~~~~~ 64
Mja_rpoE2 ~~~~~MRACLKCRKRYLT..NDEIPICHSPT..SENWIGLLIVINPEKSEIAKKAQIDIKKYALSVKE~~~~~ 59
Mth_rpoE2 ~~~~~MTEKACTRCKRRLT..SDERCPVQNVPA.STNWGLLIIIDPDKSDIARELNIITLPGYALRVR~~~~~ 60
Mac_rpoE2 ~~~~~MSEKVCRHOLRV..L.EGOTCPVCGTSDLAEWVGLVLIIDTERSEIAKRLGVDIPDRFALKVR~~~~~ 61
Mma_rpoE2 ~~~~~MPEKVCRHOLRV..L.EGOTCPVCGTSDLAEWVGLVLIIDTERSEIAKRLGVDIPDRFALKVR~~~~~ 61
Pfu_rpoE2 ~~~~~MSEKACRHOHYI..T.SEDRCPVCGSRLSEEWFDLVIIIDVENSEIAKRIKAKVPGKYAIRVR~~~~~ 61
Pab_rpoE2 ~~~~~MTEKACRNOHYI..T.TEDQCPVCGSRLSEEWFDLVIIIDVENSEIAKRIKAKVPGKYAIRVR~~~~~ 60
Pho_rpoE2 ~~~~~MTEKACRNOHYI..T.TEDQCPVCGSRLSEEWFDLVIIIDVENSEIAKRIKAKVPGKYAIRVR~~~~~ 61
Afu_rpoE2 ~~~~~MAETACRNFELNV..DTNIRNCGSTDLTKEWYGVVIVDPEKSEIAKRLIKIPGKYALRVC~~~~~ 62
Hsp_rpoE2 ~~~~~MASNRICHDOHRIVPEDEMCPYCSNSITEDWAGVVIITHPDISEIADKMEVHEGCEFALKVR~~~~~ 65
Sso_rpoE2 ~~~~~MAKRSVLEKACRNCALVDTKRETPVCGSTSETDEWDGMIIINSE.SEIAKITEAPKRWYAIITK~~~~~ 66
Sto_rpoE2 ~~~~~MPGKNEKACRNCALVDPDTKCPVCGSISFSDDSGIVIIIDPE.SETAKLECATVWRVYAIIVK~~~~~ 67
Neq_rpoE2 MPKPKR...KVQLCKKAILLMDVEKCPYCGCTEFTEFEGIIIVNKEKSEARVAGIDKEGKFAIKI~~~~~ 65
Pae_rpoE2 MSTRKRTLSGKACRNCALVMPEDAQPENCGSEFVFRWRGMIVVIDPKSCIAKRLGLEKPGYALRVE~~~~~ 73
Consensus ~~~~~ KAC C CP CG W G II D E SEIAK G PG YA V ~~~~~ 25

8.1.3.7 Untereinheit F (rpoF)

Obwohl auch MK1659 aus *M.kandleri* gut in dieses Alignment paßt, wurde die Sequenz weggelassen, weil sonst Rpb4 aus *S.cerevisiae* nicht mehr integriert werden kann (das *M.kandleri*-Gen verursacht durch einige zusätzliche Gaps eine Verschlechterung des Alignments). Bei *A.pernix* und *N.equitans* fanden sich keine homologen Gene.

```

Sce_rpb4, MNVSTSTFQTRRRRLKKEVEEENAATLQLGQEFOLKQINHQEVEELIALNLSEARLVKEALVERRRAFKRQKHKHKLKHENANDETTAVEDEDDDLDEDDVN,AD 109
Pae_rpoF, ~~~~~~MSVKKIKRSEETPNAA 16
Sso_rpoF, ~~~~~~MS,SVYIIVEEHHYIPYSV 16
Sto_rpoF, ~~~~~~MKILSFSKLEEBEHHYVPSF 20
Tvo_rpoF, ~~~~~~MEKKYISIBE 10
Tac_rpoF, ~~~~~~MMEKKYISIBE 12
Mja_rpoF, ~~~~~~MYQAKRERMICKKILGPRVTVSE 24
Mac_rpoF, ~~~~~~MIVKVELNEEILLTAE 16
Mma_rpoF, ~~~~~~MIVKVELNEEILLTAE 16
Hsp_rpoF, ~~~~~~MTISKEIVTEPLTVSE 17
Mth_rpoF, ~~~~~~MPEMIGKVKLESEPVMAE 19
Afu_rpoF, ~~~~~~MVEMG,KEVRDEEYITISE 18
Pfu_rpoF, ~~~~~~MIGRGLGKEYITIAE 16
Pab_rpoF, ~~~~~~MIGRKLGEKYITIAE 16
Pho_rpoF, ~~~~~~MIGRKLGEKYITISE 16
Consensus ~~~~~~M E Y E 4
    
```

```

Sce_rpb4, DDDEMHSETREKELESIDVILPOTTGGNKKDKNIMYOYLINFSRFRDQETVGAVIQLLKSTGLHPFEVA,QLCSLACDTADEAKTITPSLNKKI,SDDELERIKELSNL 217
Pae_rpoF, ALAILREIATTTQ,.....L.....SEDRKHLIDYLEKVKRRPRESESRIMRQOLVEKFGFTRIT,ATQLVNLSDSDEELKTIISHLEREFSECELKEIKILTGO 111
Sso_rpoF, AKKLLTDVIRSGGSSNL,.....LORTYDYLSNVEKCDAESAQVIEELSNIIVSREDVRAT, .LASICPTTSDEVRSTLVMDTNKTYTSEDIQKIIDIIRKY 110
Sto_rpoF, AKKILODLISNGTSSTI,.....LORTFEYLSNVAKCSSEDAIKIMDELKEITQREDVRAV, .IASICPTTTEEVRSTLVLDSEKTYTTEQIQKIINIIVKSH 114
Tvo_rpoF, WYNILSKK,.....EQL,.....TDMEKENLEYACKFSRVDPSPVIDRVKCDISCHAK,LEPKVIVKILDLRPGTADITSTILSTYNVMI,SDEDLNALADYVKKL 101
Tac_rpoF, VIDLLSNR,.....ENL,.....TDLEKENLAYAEKFAKIDPKAIKKVREDILSIVD, .MPEKALVKILDLRPGTADITSTILSTYGVMI,SDENLNALADYVKKL 103
Mja_rpoF, AAETIMYNRAQIG,.....EL,.....SYEQCCALDYLOKFAKIDKEEAKLVEELISLG, .IDEKTAVKIADILPEDDLDLRAIYKREL, .BENAEILEIVRKY 114
Mac_rpoF, VKDIISRADERREQGIE, .V,.....AYSRRKALHAEQFSKINGEKSRELVNKLEL, .EKMKPVIAVRIADILPQSRDELRSIYAKEKYTL, .SNEELDQILDYVFAA 115
Mma_rpoF, VKDKLTAIAEERREKQIE, .V,.....SYSRRKALNHAEQFSRISGVKSRELVNKLEL, .EKMKPVIAVRIADILPQSRDELRSIYAKEKYTL, .SNEELDQILDYVFAA 115
Hsp_rpoF, AKALLSEV,ETERALDEDREL,.....REFLSRALEHANRFALTEPAESREIVDELLAUDVLDDEAVAKIVDLLRPTRETELRSVEANERYAM, .SGDELDEILDVVAKY 118
Mth_rpoF, VKEILKFK, .G,.....EDHEL,.....TYEQNIYLDHVTRFSRLDPEKSRELVEELMDIPNKKRRH, .AVKIADIMPVDISDLRTIFAKBRVPI, .KADDLPGILEVIDKY 113
Afu_rpoF, AKEIMEET, .AKKROEPAEL,.....DEETRRALKHURLFAKLPADKAKELVDELVNLPOVGSREIATVKLADIMPRIPEDEVRIIYAKBRVTL, .TBEQIQEILDVIDKY 117
Pfu_rpoF, AKELLKRRREEVKAGIEEPL,.....YYEARTALEHAERFAKLPADKAKAEVEELMNAFVMSDRIACKIVDIMPEDSMDLRVIFAKEEYOP, .TQEMKQILDILDKY 118
Pab_rpoF, AKEIILKKYEEGKKAGIEEPL,.....YYEARTALEHAERFAKLSAEKARAEVEELISAFDWMNDRIACKIVDIMPEDNLDLRVIFAKEEYOP, .TQEMKQILDILDKY 118
Pho_rpoF, AKEIILKKYEEGKKAGIEEPL,.....YYEARTALEHAERFAKLPADKAKAEVEELMSEEFWMNDRIACKIVDIMPEDNLDLRVIFAKEEYOP, .TQEMKQILDILDKY 118
Consensus AK IL L..... L F K V EL A KI DI P ELR I . E IL K 29
    
```

```

Sce_rpb4, EYLY 221
Pae_rpoF, ~~~~ 111
Sso_rpoF, IKS~ 113
Sto_rpoF, MES~ 117
Tvo_rpoF, RE~ 103
Tac_rpoF, RE~ 105
Mja_rpoF, H~ 115
Mac_rpoF, ME~ 117
Mma_rpoF, ME~ 117
Hsp_rpoF, V~ 119
Mth_rpoF, RVE~ 116
Afu_rpoF, RH~ 119
Pfu_rpoF, RE~ 120
Pab_rpoF, RET~ 121
Pho_rpoF, RE~ 121
Consensus ~~~~ 29
    
```

8.1.3.8 Untereinheit H (rpoH)

↓ 1: Jaw →

```

Sce_rpb5, MDQENERNISRLWRAFRTVKEMVVKDRGYFITOEVELPLEDFKAKYCDMSGRPQRKMMSFOANPTEESISKFPDMGSLWVEFCDEPSVGVKTMKTFVIHQEKNFOQIGIF 110
Mma_rpoH, ~~~~~~MFF 3

```

↓ 2: Assembly →

```

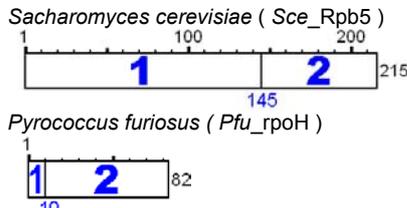
Sce_rpb5, VYONNITPSAMKLVPSIPPATVETTFNEAALVVNTHHELVLPKHIRLSSDEK.RELLKRYVLRKESQ.LPRIQRADPVALVILG.....LKRGEVVKIIRKSETSGRYASYSR 212
Sto_rpoH, ~MVETTYILRLSLVN....IMRS..SSKIRIDPSQHELVPKHEVLSDEE.AYKILKELGKIKPEQ.LPWRASDPVAKVGA.....KPGDIIKIIRKSSLSGEVVVYR 94
Mma_rpoH, ~IMTGKKAIRWRNLIKFFECFIFGG..KDLTKFSLLDHDPVPKHEIMSEGE.LKSVLSKYVIEKEQ.LPKIKVQDPVCKEIGA.....VVGDVVKIRKSSOTAGEADYYR 103
Hsp_rpoH, ~~~~~~MVDVSOHELVPEHIVL.DDEETLDGVLADYDIDRTE.LPKIKYKDPALPDNAEI.....GDVVEIVRDSSTTDEAVVYR 71
Pae_rpoH, ~~~~~~MSRISLGVKEVIVPREEAKELLKRMKLRPQO.LPWIRASDPVLAOKAGA.....KPGDVLKIVRESPTAGEAVVYR 70
Pab_rpoH, ~~~~~~M.AGKSEFNIRKHLVPEHRILSDEEK.KALLEKYEITPAQ.LPQIKASDPAVKALGA.....KPGDIIETIRKSSPTAGVYVYR 77
Pho_rpoH, ~~~~~~M.SCKNEFNIRKHLVPEHRILSDEEK.KALLEKYEITPAQ.LPQIRASDPAVKALGA.....KPGDIIETIRKSSPTAGVYVYR 77
Pfu_rpoH, ~~~~~~M.AGKKEFSIDHVLVPEHRVLSDEEK.KALLEKYEITPAQ.LPQIKASDPAVKALGA.....KPGDVIEIRKSSPTAGVYVYR 77
Tvo_rpoH, ~~~~~~M.MAKFNVLHDHNLVPEHRIVSEEEK.KTILKELGIEK.EFLPRISENDPVIAKALEAHEGK.IKQGTVIKIIIRNSPTVGHSVYR 79
Tac_rpoH, ~~~~~~MV.VKMAFNVLHDHNLVPEHRIVSEEEK.KNLLKELNIEK.EFLPKISENDPAIKALEAVHGGK.IKQGTIIKIKIRKSSPTAGVYVYR 83
Mka_rpoH, ~~~~~~MA.ERLDPDVLNHLVLPKHEVLDDEEIEKILEELGVEK.DDLPRIHTNDPVVVALSEKIKGRIKPGSLVKIVRDSPTAGKTVYR 85
Neq_rpoH, ~~~~~~M.MKKKOPKIKRHLVPEYORLTKEE.VEKLLKELNIDSIDYLPKIYVDDPAIQHLNP.....KPGDVLRIIRKSSPTAGKTVYR 75
Mth_rpoH, ~~~~~~M.MKREILKHQVLPHEVILNISE.AKRVLKELDAHPEQ.LPKIKTDPVAKAIGA.....KPGDVIKIIIRKSSPTAGVYVYR 73
Ape_rpoH, ~~~~~~M.SSK.TRKSSRVLEHEVPEHRILLSDEE.AVQVLKMLGKIKPEQ.LPKISVNDPIARTLKA.....KPGDIIETIRKSSPTAGEAVYR 80
Sso_rpoH, ~~~~~~M.MRG.SSNKRIDPRIHVLVPKHEVLDDEE.AYKILKELGKIRPEQ.LPWIRASDPAVARINA.....KPGDIIIRIRKSSOLYGEVVYR 79
Mja_rpoH, ~~~~~~M.MKVLDHMLVLPKHEIVPEKEE.VEELKRYNLIKIQO.LPKIYEDDPIQIEIGA.....KPGDVVVRIRKSSPTAGVYVYR 71
Afu_rpoH, ~~~~~~M.MKFKLDHMLVLPKHEVLSKEE.AEELKIKIGICEQ.LPKIKADDPVIAKEIGA.....KPGDVIKIIIRKSSPTAGEAVYR 73
Mac_rpoH, ~~~~~~M.MTKFSLLDHESVPKHEIISDEE.LKSVLSKYVIEKEQ.LPKIKVQDPVSKIEIGA.....VVGDVVRIIRKSSOTAGEADYYR 74
Consensus, ~~~~~~H LVP H L EE LK I Q LP I DP GA.....K GD I RKS TAG YR 31

```

```

Sce_rpb5, ICM ~~~~~ 215
Sto_rpoH, LVISC ~~~~~ 99
Mma_rpoH, LVIE ~~~~~ 107
Hsp_rpoH, LVIE ~~~~~ 75
Pae_rpoH, LVVPC ~~~~~ 75
Pab_rpoH, VVVE ~~~~~ 82
Pho_rpoH, VVVE ~~~~~ 82
Pfu_rpoH, VVVE ~~~~~ 82
Tvo_rpoH, VVASEVFK ~~~~~ 87
Tac_rpoH, VVASEVFK ~~~~~ 91
Mka_rpoH, VVTPPE ~~~~~ 92
Neq_rpoH, VVIERPKIVEKEBEEBE 91
Mth_rpoH, LVOD ~~~~~ 77
Ape_rpoH, EVVAYQKGVK ~~~~~ 90
Sso_rpoH, LVISC ~~~~~ 84
Mja_rpoH, LVIRRII ~~~~~ 78
Afu_rpoH, LVV ~~~~~ 76
Mac_rpoH, LVIE ~~~~~ 78
Consensus, V ~~~~~ 32

```



8.1.3.9 Untereinheit K (rpoK)

↓ 1: Tail →

```

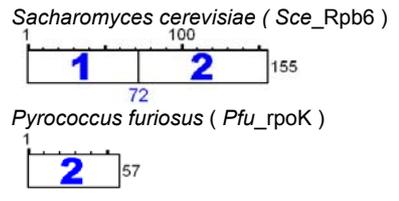
Mma_rpoK, ~~~~~~MSNLSCFSTIIGTHRSVFRKFGDLLVKN.....YTRFERARIVGARALQIAMGAPVLVE 54
Ape_rpoK, ~~~~~~MAERYTSYPRIVDDVKIIGPPY.....LTRYERARIIGVRAEQLSIGAPPLVD 47
Sso_rpoK, ~~~~~~MGLERDEITISDLHNEVHISLWNR.....LTRYEARVISARALQIAMGAPALID 52
Pae_rpoK, ~~~~~~MTTSELIERINKLIDVVKAINKREFFPPR.....LTKYERARIIGARALQIAMGAPLVD 56
Sce_rpb6, MSDYEEAFNDGNGENFDFDVEHFSDEETYEKQPFQDCGETTDANGKTIIVGGNGPDDFOQHEQIRRTKLEKAIIPKORATTPYMTKYERARILGTRALQISMNAPVVD 110
Mja_rpoK, ~~~~~~MMLTKFEARILGARSLQISSGAYVIE 28
Neq_rpoK, ~~~~~~MMSKYEKARIIGRAMQIAGAPILIK 27
Hsp_rpoK, ~~~~~~MMSDSQHSRYEKARIIGARALQVAYGAPLVD 32
Mth_rpoK, ~~~~~~MMHMASKLTRFERARILGARALQISMGARPLVE 33
Mka_rpoK, ~~~~~~MEDLTRFERARIVGARALQIAMGAPVLVE 29
Pab_rpoK, ~~~~~~MFKYTRFEKARIIGARALQIAMGAPVLID 29
Pho_rpoK, ~~~~~~MFKYTRFEKARIIGARALQIAMGAPVLID 29
Pfu_rpoK, ~~~~~~MFKYTRFEKARIIGARALQISMGAPVLID 29
Tac_rpoK, ~~~~~~MKYTKFEKARIIGARALQIAMGAPVLID 28
Tvo_rpoK, ~~~~~~MKYTKFEKARIIGARALQIAMGAPVLID 28
Afu_rpoK, ~~~~~~MGIKVKPPEYTRFEKARIIGARALQIAMGAPVLVE 36
Mac_rpoK, ~~~~~~MVKR.....YTRFERARIVGARALQIAMGAPVLVD 31
Sto_rpoK, ~~~~~~MSINKIFESWKNR.....LTRYEARIVISARALQIAMGAPLID 40
Consensu, ~~~~~~T FE ARIIGARALQIAMGAPVL D 22

```

```

Mma_rpoK, ~DDGRID....PLGVAEELKAGVIPITVRR..KKS..... 83
Ape_rpoK, ~PER....AGSRNPLDIARVEVDNGILPVSIVRPLPGGGGOSLSRLV.EIAREILLGSEYV 103
Sso_rpoK, ~INN....ISSSDVISIABEEFRRGVLPITIRRLPNGKILLSLRKS..... 95
Pae_rpoK, ~IQE....VGSVDPVLIAMEELRGLDFFIIVREMPDGKTRIRLKELL.ELERTI..... 106
Sce_rpb6, ~LE....GETDPLRTAMKELAEKRIPLVIRRLPQGSFEDDSVEELIVDL 155
Mja_rpoK, ~TK....CDSLKIAYEEIKQCKVPLKEITREVKA..... 57
Neq_rpoK, ~IDKETIKKINYSPIEIAKMEFEKALPIKIVIR..KIPVKPHIN...EV..... 70
Hsp_rpoK, ~TDQTE...E....PIIYAAEEYDADALPITVRR..EGT..... 60
Mth_rpoK, ~IKESLD....PVDIARKELEKVMPLQVRR..DK..... 61
Mka_rpoK, ~V.EGDD....PLEIAKREFDEGVVPPVVIR..R..... 56
Pab_rpoK, ~VEEG.IT....PLEAALMEFEKGVIPITVIR..PS..... 57
Pho_rpoK, ~VEEG.IT....PLEAALMEFEKGVIPITVIR..PS..... 57
Pfu_rpoK, ~VPPG.IT....PLEAALMEFEKGVIPITVIR..PS..... 57
Tac_rpoK, ~VEKNID....PVDIAMLEFFENVIPITIKK..ASKILN..... 61
Tvo_rpoK, ~IEKNVID....PVDIATLEFFENVIPITIKR..VSRVILN..... 61
Afu_rpoK, ~TD.KTE....PLEIALEEENRGVIPITVRR..RRNEFVWLERYDLE..... 75
Mac_rpoK, ~DDGRID....PLGVAEELKAGVIPITVRR..KKS..... 60
Sto_rpoK, ~TNN....LPENDVISIABEELKRGVLPITVRRVYVYPNKTELVSVKVY..EFK..... 86
Consensu, ~.....P I A E G V PITV R. .... 33

```



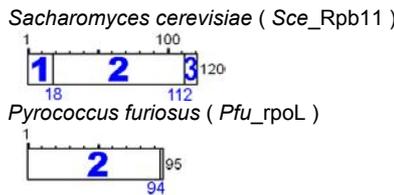
8.1.3.10 Untereinheit L (rpoL)

↓ 1: Tail → ↓ 2: Dimerization →

<i>Ape_rpoL</i>	~LRLERLRIIRESRGSYVIVYGEDHTLCTLLVLEAIKRVSNKELAYEAVHPMEDIIQVYVKY.....EDDVDIKVEILRKASDYLLLEVI	84
<i>Sce_rpb11</i>	~MNPADRFELEFTEGEGSKLIDPDKAPNAVITTEKEDHTLGNLIRALLINDRRVLFVAAAYKVEHEFFARF.....KLRITQTEGYDPKDALRQACNSIINKL	98
<i>Mka_rpoL</i>	~MELPEVEVVVKYDKDEVLELEGGEDHTLGNLIRWALNRDGTIATYRLEHPILGKEHKVDEERYVPPKMRIRAVDEADAREALERATEELLELV	96
<i>Neq_rpoL</i>	~MVTIPEVKILEETDNLIKVQKGEHTLLENALKEHAYTINGVKKAAYLEIEHPLRKNNY.....FITETDGSIKARDALIQALKKKEEL	85
<i>Pae_rpoL</i>	~MSPVDRLEILKDDKYLELKAQGETYTLSEPLVEVLSNDPVEYVQFDVDHPLOSNAYE.....K.....LKVKRNPEAIQRAVNAIISDL	83
<i>Pab_rpoL</i>	~MKIEVIKKEODLLEFYLEGGEDHTFANLLTEVLRNPNPKVFVAYTIEHPILMAR.....KPRFRVVTDGKTPPEKALEEAAKTIISDRA	82
<i>Pho_rpoL</i>	~MKIEVIKKEEDMLEFYLEGGEDHTFANLLTEVLRNPNPKVFVAYTIEHPITMAR.....KPRFRVVTDGKTPPEKALEEAAKTIISDRA	82
<i>Pfu_rpoL</i>	~MKIEVIKKEENLLEFYLEGGEDHTFANLLTEVLRNPNPKVFVAYTIEHPITMAR.....KPRFRVVTDGKTPPEKALEEAAKTIISDRA	82
<i>Sso_rpoL</i>	~MEIRILKSESNYLELETEGGEDHTLGNLIRAGTLRRTISGVSFASYQPHPLSDKII.....VKLLTDGSIIPKDALIKALENIRGMT	80
<i>Sto_rpoL</i>	~MEIKILRSGENYLELQIDGEEHTVGNLLKGYLLKVPGVKFAYSKPHPLDSII.....LKIMTDGSIIPKDALIKALENIRGMT	80
<i>Hsp_rpoL</i>	~MDLRVIEKDDAELTIEVAGENHTFMNVLKGLLELDEGVTAASYDNPEQSGGQT.....EPILLTKTDGSDVPVDALQAAAGHTSOKL	83
<i>Mth_rpoL</i>	~MEVILDKRNEMEIVEEGEHTLGNLRSILMEDERVKAAAYSIDHPI.....VGEPOLYTRAGS...PKKSLKAAAEITPRDR	75
<i>Afu_rpoL</i>	~MEVKIIDIGEDYVRLVIGECHTYLNLLQHYLVEDEVIVARYNIEHPIL.....ICEPEIYIKTSGVDPTFAVKRANEKIITAC	79
<i>Mac_rpoL</i>	~MEINILNKTNNELEVELRGEHTLGNLLKDLIKDERVEAAFYDKMHSV.....ISDPVLYIKTSGVDPTFLVKETAALITTAOC	79
<i>Mma_rpoL</i>	~MEINILNKTNNELEVELRGEHTLGNLLKDLIKDERVVTAFYDKMHSV.....ISDPVLYIKTSGVDPTFLVKEDVVAIVSEC	79
<i>Mja_rpoL</i>	~MEIKILERKDNVLELITNEDHSLNLLKDLITKTEGKVMASYSIDHPLHPETG...RMISNPKTIITIEGADPEVLEGLERDILKMC	88
<i>Tac_rpoL</i>	~MORERTAESLRLVSKKNSITVEMINVDNTILRLTVEEILKDDQVDEARYVIKHPVLDNPO.....IYVVKVSKGKPOSAIKRAVRKLSKLY	87
<i>Tvo_rpoL</i>	~MORERTNLSLRVSKKNSITVEMINVDNTILRLTVEEILKDDQVDEARYVIKHPVLDNPO.....IYVVKVSKGKPOSAIKRSIRKLSKLY	87
Consensus	~.....GEHTLNLLLVAYHP.....PALA	18

↓ 3: Tail →

<i>Ape_rpoL</i>	~GDFRRRVLKALERRGGGG.....	102
<i>Sce_rpb11</i>	~GALKTNFEFEWNLCFLAADDAA.....	120
<i>Mka_rpoL</i>	~EEAKDEEFGA...LEEKES.....	112
<i>Neq_rpoL</i>	~LNFKDWYSN...L.....	96
<i>Pae_rpoL</i>	~EELERGFPS.....	92
<i>Pab_rpoL</i>	~KEVLDWEKA...IK.....	95
<i>Pho_rpoL</i>	~KEVLDWEKV...VKS.....	95
<i>Pfu_rpoL</i>	~KEVLDWEKA...VKS.....	95
<i>Sso_rpoL</i>	~SHYLDDEIKGL...TK.....	92
<i>Sto_rpoL</i>	~NKFTIEVKS...EK.....	93
<i>Hsp_rpoL</i>	~QSFADTFEAA...V.....	94
<i>Mth_rpoL</i>	~DEFRRLLIESL.....	85
<i>Afu_rpoL</i>	~NTL...LEQL.....	86
<i>Mac_rpoL</i>	~DEF...IDVF...SKAANA.....	92
<i>Mma_rpoL</i>	~DEF...IDVF...SKAANA.....	92
<i>Mja_rpoL</i>	~DTL...LDEL...KEKK.....	99
<i>Tac_rpoL</i>	~EDLGTQFOKE...FORYESDHMIKAVE.....	111
<i>Tvo_rpoL</i>	~EDLGAQVEKE...LEKYRSNHIIKTTTE.....	111
Consensus	~.....	18



8.1.3.11 Untereinheit N (rpoN)

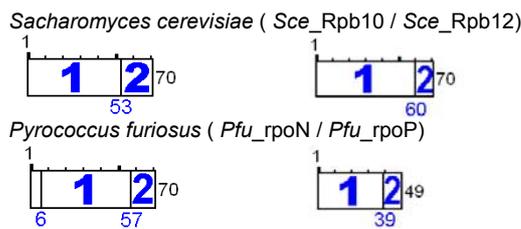
Das Zinkfinger-Motiv wurde rot markiert.

↓ 1: Zink Bundle →

<i>Mja_rpoN</i>	~MFPVRCFSCGVIASVYVVEEYKERTIK.....GENPKVLDLDLGKIKYCCRR	47
<i>Afu_rpoN</i>	~MDVPRNENSSDFPIRCFSCGAVIGHLYDEYKOKISE.....GKSPKEALDELGIERYCCRR	56
<i>Mth_rpoN</i>	~MIPVRCFSCGKPVSAFNEYQRRVA.....DGEDPKVLDLDLGKIKYCCRR	46
<i>Neq_rpoN</i>	~VLPPIRCFSCGKPLGHLWEEFNERNK.....GEDPKVLDLDLGKIKYCCRR	47
<i>Ape_rpoN</i>	~MIPVRCFSCGKPLGHLWEEFNERNK.....GKTPCEALDELGVDRYCCRR	47
<i>Tac_rpoN</i>	~MIIPVRCFSCGRVIAVDYGRVTKRVNEIKAEGRDPSPEEIKETFDLDGKIKYCCRR	56
<i>Tvo_rpoN</i>	~MIIPVRCFSCGRVIAVDYGRVTKRVNEIRSEGREPTAELEKTFDLDGKIKYCCRR	56
<i>Hsp_rpoN</i>	~MMPVRCFSCGVIASVYVVEEYKERTIK.....HDGDEDPADVLDLDGKIKYCCRR	50
<i>Sce_rpb10</i>	~MIPVRCFSCGKPVSAFNEYQRRVA.....EDELDEGTAISRLGLKRYCCRR	48
<i>Mka_rpoN</i>	~MIIPVRCFSCGKPIAHLWEKYVELLE.....EPGMEPEALDELGVDRYCCRR	48
<i>Mac_rpoN</i>	~MIPVRCFSCGKVISNYWDEYKRRV.....TDGECSAVLDLDGKIKYCCRR	46
<i>Mma_rpoN</i>	~MIPVRCFSCGKVISNYWDEYKRRV.....TDGECSAVLDLDGKIKYCCRR	46
<i>Pae_rpoN</i>	~MIIPVRCFSCGKPLGHLWEEFNERNK.....LAGEPEGRVLDLDGKIKYCCRR	47
<i>Sto_rpoN</i>	~MIIPVRCFSCGKPIAHLWEKYVELLE.....NAGENPKVLDLDGKIKYCCRR	47
<i>Sso_rpoN</i>	~MIIPVRCFSCGLIDKWEFPITRV.....SNGEDPKVLDLDGKIKYCCRR	47
<i>Pab_rpoN</i>	~MIPVRCFSCGKVIASVYVVEEYKERTIK.....EAGEDPEKVLDDGLGKIKYCCRR	47
<i>Pfu_rpoN</i>	~MIGASPVLPVRCFTCGKVIADKRYEYFKRRV.....EAGEDPEKVLDDGLGKIKYCCRR	52
<i>Pho_rpoN</i>	~MIEQCWSEIQEESQSPSQIGQLRVQGLRDKRVIYEYIPLILKTFLLSKTSGVGPLIIPVRCFTCGKVIADKRYEYFKRRV.....EAGEDPEKVLDDGLGKIKYCCRR	101
Consensus	~.....M PVRCF CG VI RV.....GE P VLDDLG RYCCRR	27

↓ 2: Tail →

<i>Mja_rpoN</i>	~MISYRIGEDGREIIDEIHADERYL.....	73
<i>Afu_rpoN</i>	~METHK.....SVLKEKSRFHGAVG.....	76
<i>Mth_rpoN</i>	~MLISHV.....ETW.....	55
<i>Neq_rpoN</i>	~TLLSAV.....VFPVIAKFKKV.....	65
<i>Ape_rpoN</i>	~TLFISV.....TYLHAAKYRVGIHVPPERLKMWC	77
<i>Tac_rpoN</i>	~MIVSHV.....DLISEIMPS.....	72
<i>Tvo_rpoN</i>	~MISHV.....DLINEIMPS.....	72
<i>Hsp_rpoN</i>	~MLIAHQ.....DLVDVSPYQ.....	66
<i>Sce_rpb10</i>	~MLLTHV.....DLLEKIRVNP...LEKRD.....	70
<i>Mka_rpoN</i>	~MELSHV.....DLLEESIPYTPPRLGMPR.....	72
<i>Mac_rpoN</i>	~MELSHV.....ELVDVSPYQ.....	62
<i>Mma_rpoN</i>	~MELSHV.....ELIDVSPYQ.....	62
<i>Pae_rpoN</i>	~TLMAHV.....EVIDDVLVRRS.....	66
<i>Sso_rpoN</i>	~MLLSHV.....DIINEVHYTRPI.....	66
<i>Sto_rpoN</i>	~MLLSHV.....DIKKEVHYTRPI.....	66
<i>Pab_rpoN</i>	~MLLSHV.....ELIDDIMHYRVY.....	65
<i>Pfu_rpoN</i>	~MLLSHV.....ELIDDIMHYRVY.....	70
<i>Pho_rpoN</i>	~MLLSHV.....ELIDDIMHYRVY.....	119
Consensus	~M SHV..... I Y.....	33



8.1.3.12 Untereinheit P (rpoP)

Das Zinkfinger-Motiv wurde rot markiert.



8.1.3.13 Aminosäure-Austausch: Bewertung der „Ähnlichkeit“ in Alignments

Warum unterscheiden sich Alignments teilweise mehr oder weniger?

Wie schon beschrieben (siehe [Abb.4-09]), beruht die Zuordnung ähnlicher Aminosäure-Positionen (und somit letztlich die Aussage, wie nah zwei Proteine miteinander verwandt sind) auf Eigenschaften-Teilmengen, die an sich schon gewichtet werden wollen (z.B. ob für eine Aminosäure nun entscheidend ist, dass sie klein oder geladen ist). Darüber hinaus hat man festgestellt, dass diese Ähnlichkeitsbewertung bei sehr nah verwandten (also fast identischen) Sequenzen einerseits, und phylogenetisch sehr weit voneinander entfernten Proteinen unterschiedlich sein muß, um zu aussagekräftigen und schlüssigen Ergebnissen (z.B. in Form eines Stammbaums) zu gelangen.

Aus diesem Grund können je nach Art der untersuchten Sequenzen sgn. „Matrices“ verwendet werden – letztlich Tabellen, die für jedes Aminosäure-Paar angeben, wie groß die Ähnlichkeit sein soll. Natürlich kann eine Matrix nicht unbegrenzt manipuliert werden – schließlich beziehen sich die Angaben auf unveränderliche, physikalisch-chemische Eigenschaften der Aminosäuren.

Die heute meistens eingesetzten Matrices wie z.B. die unten genannte "PAM 250"-Matrix wurden gar nicht aufgrund irgendwelcher physikalisch-chemischer Daten sondern auf der Basis der statistischen Auswertung einer großen Zahl miteinander verwandter Proteine ermittelt. Man schaut also nach, bei welchen Aminosäurepaaren die Evolution einen Austausch begünstigt und wo ein solcher Austausch extrem selten vorkommt. Die wird dann in der "Tabelle" berücksichtigt.

Jedenfalls ist dies eine wichtigen Ursache dafür, dass nicht jedes Alignment (trotz gleicher Sequenzen) hundertprozentig identisch ist (allerdings gibt es noch andere Ursachen, z.B. Einstellungen, für wie schwerwiegend eine größere Lücke gehalten wird = „Gap penalty“ etc.).

Die folgende, mit ClustalX erstellte Grafik verdeutlicht das Gesagte: In Form zweier fiktiver Aminosäure-Sequenzen wurde jede Aminosäure mit jeder anderen kombiniert. Es wurde die Standard-Matrix „Gonnet PAM 250“-Matrix – die Voreinstellung von ClustalX – verwendet. Die Höhe der Säulen ist ein Maß für die Ähnlichkeit der verglichenen Aminosäuren.

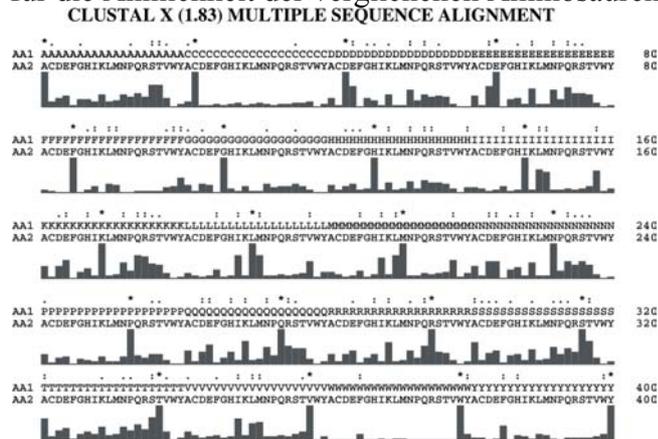


Abb.8-01: Festlegung von "Ähnlichkeit": ClustalX, PAM 250-Matrix

Folgende Ähnlichkeiten wurden (neben der eindeutigen „Identität“ zweier Aminosäuren) in dieser Arbeit verwendet: A (S,T), C (), D (E,N), E (D,K,N,Q), F (), G (), H (), I (L,M,V), K (E,Q,R), L (I, M, V), M (I, L, V), N (D, E, Q, S), P (), Q (E,K,N,R), R (K,Q), S (A,N,T), T (S,A), V (I,L,M), W (), Y (). Als Richtwert für „Ähnlichkeit“ wurde „mindestens 50%“ entsprechend o.g. Standard-Matrix gewählt.

8.2 Liste verfügbarer 3D-Strukturen zum Thema RNA-Polymerase & Transkription

Die Zahl experimentell verifizierter (also per Röntgenstrukturanalyse von Kristallen bzw. NMR-Spektroskopie gewonnenen) 3D-Strukturen wächst stetig. Die mit diesen Techniken ermittelten Angaben über die dreidimensionale Form von Proteinen sind sehr präzise und eine wichtige Referenz z.B. für die große Zahl von Bioinformatik-Tools, die ihre Voraussagen bezüglich möglicher (oder wahrscheinlicher) 3D-Strukturen, hauptsächlich aus der Analyse der Aminosäure-Sequenz (also der Primärstruktur der Proteine) gewinnen. Als leistungsfähiges Werkzeug erweist sich die Kombination dieser Programme / Algorithmen mit Protein-Datenbanken: Mittels Alignments wird teilweise recht erfolgreich der Versuch unternommen, die 3D-Struktur eines neuen Proteins mit Hilfe von Kristallstruktur-Daten ähnlicher (aber nicht identischer) Protein zu kalkulieren.

In dieser Arbeit stellte vor allem das 12 Untereinheiten umfassende 3D-Modell der eukaryotischen RNA-Pol II aus *S.cerevisiae* eine wichtige Referenz für die Diskussion der Daten aus den Far Western Blot-Experimenten dar.

Es wird jeweils der Organismus, der Name des Proteins (bzw. Proteinkomplexes), der / die Urheber, sowie die MMDB- bzw. PDB-ID angegeben, über die man auf der NCBI-Webseite [3.8] schnell den entsprechenden Datensatz finden kann.

Structure Of Tata-Binding Protein

Pyrococcus woesei
B.S.Dedecker & P.B.Sigler, 4-Oct-96
MMDB: 5296 PDB: 1PCZ

Tata-Binding ProteinTRANSCRIPTION FACTOR (II)BBRE+TATA- Box Complex From Pyrococcus Woesei

Pyrococcus woesei
O.Littlefield, Y.Korkhin & P.B.Sigler, 1-Oct-99
MMDB: 11573 PDB: 1D3U

Crystal Structure Of The Lrp-Like Transcriptional Regulator From The Archaeon Pyrococcus Furiosus.

Pyrococcus furiosus
P.M.Leonard, S.H.J.Smits, S.E.Sedelnikova, A.B.Brinkman, W.M.De, Vos, J.Van Der Oost, D.W.Rice & J.B.Rafferty, 1-Feb-01
MMDB: 18395 PDB: 111G

Crystal Structure Of Thermus Aquaticus Core Rna Polymerase- Includes Complete Structure With Side-Chains (Except For Disordered Regions)-Further Refined From Original Deposition-Contains Additional Sequence Information

Thermus aquaticus
L.Minakhin, S.Bhagat, A.Brunning, E.A.Campbell, S.A.Darst, R.H.Ebright & K.Severinov, 18-Dec-00
MMDB: 15252 PDB: 1HQM (wichtigste Referenz für die RNAP bei *Bacteria*)

Complete 12-Subunit Rna Polymerase II

Saccharomyces cerevisiae
K.-J.Armache, H.Kettenberger & P.Cramer, 29-Jan-03
MMDB: 24539 PDB: 1NT9 (wichtigste Referenz für die RNAP bei *Eukarya*)

Structure Of An Archeal Homolog Of The Eukaryotic Rna Polymerase II Rpb4Rpb7 COMPLEX

Methanococcus jannaschii
F.Todone, P.Brick, F.Werner, R.O.J.Weinzierl & S.Onesti, 17-Oct-01
MMDB: 18143 PDB: 1GO3

N-Terminal Domain Of Tfiib, Nmr

Pyrococcus furiosus
W.Zhu, Q.Zeng, C.M.Colangelo, L.M.Lewis, M.F.Summers & R.A.Scott, 27-Mar-96
MMDB: 4839 PDB: 1PFT

RNA Polymerase II-TfiIS Complex

Saccharomyces cerevisiae
H.Kettenberger, K.-J.Armache & P.Cramer, 19-Jun-03
MMDB: 24113 PDB: 1PQV

Wild Type Rna Polymerase Ii*Saccharomyces cerevisiae*

D.A.Bushnell & R.D.Kornberg, 24-Dec-02

MMDB: 22688 PDB: 1NIK

Crystal Structure Of Yeast Rna Polymerase Ii Complexed With The Inhibitor Alpha Amanitin*Saccharomyces cerevisiae*

D.A.Bushnell, P.Cramer & R.D.Kornberg, 22-Oct-01

MMDB: 18935 PDB: 1K83

Rna Polymerase Ii Elongation Complex*Saccharomyces cerevisiae*

A.L.Gnatt, P.Cramer & R.D.Kornberg, 2-Mar-01

MMDB: 16228 PDB: 1I6H

Solution Structure Of The Essential Rna Polymerase Subunit Rpb10 From Methanobacterium thermoautotrophicum*Methanobacterium thermoautotrophicum*

C.D.Mackereth, C.H.Arrowsmith, A.M.Edwards & L.P.Mcintosh, 7-Feb-00

MMDB: 13663 PDB: 1EF4

Solution Structure Of Rna Polymerase Subunit Rpb5 From Methanobacterium Thermoautotrophicum*Methanobacterium thermoautotrophicum*

A.Yee, V.Booth, A.Dharamsi, A.Engel, A.M.Edwards & C.H.Arrowsmith, 25-Feb-00

MMDB: 13628 PDB: 1EIK

Structure Of A Transcribing T7 Rna Polymerase Initiation Complex*Bacteriophage T7*

G.M.T.Cheetham & T.A.Steitz, 1-Sep-99

MMDB: 12809 PDB: 1QLN

Crystal Structure Of A T7 Rna Polymerase-T7 Promoter Complex*Bacteriophage T7*

G.M.T.Cheetham, D.Jeruzalmi & T.A.Steitz, 11-Mar-99

MMDB: 10243 PDB: 1CEZ

Structure Of Escherichia Coli Rna Polymerase Alpha Subunit N-Terminal Domain*Escherichia coli*

G.Zhang & S.A.Darst, 8-May-98

MMDB: 10201 PDB: 1BDF

Solution Structure Of Rna Polymerase Subunit H*Methanococcus jannaschii*

A.Thiru, M.Hodach, J.Eloranta, V.Kostourou & R.Weinzierl, 5-Feb-99

MMDB: 10045 PDB: 1HMJ

T7 Rna Polymerase Complexed With T7 Lysozyme*Bacteriophage T7*

T.Steitz & D.Jeruzalmi, 8-Aug-97

MMDB: 8719 PDB: 1ARO

Thermococcus Celer Rpb9, Nmr, 25 Structures (TFIIS)*Thermococcus celer*

B.Wang, D.N.M.Jones, B.P.Kaine & M.A.Weiss, 19-Aug-97

MMDB: 6818 PDB: 1QYP

Weiter Einträge zum Thema „RNA Polymerase“ inklusive Verweisen auf die verfügbaren 3D-Strukturen im PDB-Format finden sich z.B. auch unter folgender Internet-Adresse: http://www.rcsb.org/pdb/molecules/pdb40_report.html

8.3 Wie mache ich eine lokale BLAST-Suche?

Gene auch in einem noch nicht fertig annotierten Genom suchen und finden!

Das unten angegebene Beispiel „*Nanoarchaeum*“ ist insofern nicht mehr ganz aktuell, als mittlerweile eine analysierte / annotierte Version bei NCBI vorliegt. Wie unter [Abb.4-05] angegeben, hat die „lokale BLAST-Suche“ selbst in diesem Fall noch eine wichtige Funktion, und verhindert evtl. fälschliche Aussagen zum Beispiel bezüglich des „Fehlens eines Gens im Genom von Organismus XYZ“.

Lokales BLAST auf dem eigenen PC ist z.B. interessant für die Analyse umfangreiche Nucleotid-Sequenzen, die noch nicht in Form einer annotierten Genbank-Datei vorliegen. Beispiel: *Nanoarchaeum equitans* (dieses vielleicht kleinste bekannte Archaeon lebt offensichtlich als Parasit auf *Ignicoccus* (ebenfalls *Archaea*); ähnlich klein sind nur die zellwandlosen Vertreter der Gattung *Mycoplasma* beziehungsweise die parasitären Rickettsien und Chlamydien bei *Bacteria*; siehe auch *Thermoplasma* bei *Archaea*). Ich wollte z.B. die Gene der DNA-abhängigen RNA-Polymerase von *Nanoarchaeum* für eine vergleichende Darstellung suchen, da mit *Nanoarchaeum* neben *Crenarchaeota*, *Euryarchaeota* und *Nanoarchaeota* ein weiteres Reich (die *Nanoarchaeota*) postuliert wird, und diesem Genom deshalb besonderes Interesse gilt.

Es gibt eine ganze Reihe von Internet-Seiten, auf denen man eine Online-BLAST-Suche durchführen kann.

(z.B. bei NCBI: http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). Nachteil: Nutzlos, sofern das Genom des gewünschten Organismus noch nicht in der Datenbank enthalten ist. Eine Alternative wäre z.B. die Webseite der "Biocomputing Service Group" am DKFZ Heidelberg („HUSAR Heidelberg“) - hier hat man ein eigenes Benutzerverzeichnis auf dem Server und kann mit beliebigen eigenen Sequenzdateien arbeiten. Nachteil: dieser Dienst ist nicht kostenlos - man muß regelmäßig für seinen Benutzer-Account bezahlen.

Viele der angebotenen Online-Funktionen kann man mit geeigneten Programmen auch auf dem eigenen PC realisieren. Für die BLAST-Suche wird hier das kostenlose Programm **Bioedit (Version 5.0.9)** verwendet, welches man sich unter der folgenden Adresse aus dem Internet laden kann: <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>

Die Sequenz von *Nanoarchaeum equitans* ist zum jetzigen Zeitpunkt (05.11.03) noch nicht annotiert, liegt aber schon in einer Rohfassung (nur mit der DNA-Sequenz) im Genbank-Format vor: bei http://www.ncbi.nlm.nih.gov/genomes/static/a_g.htm (einer NCBI-Seite mit einer Liste bisher komplett sequenzierter Archaea) als Suchbegriff "Nanoarchaeum" eingeben ("Search" von "Genome" auf "Nucleotide" stellen). Von den angezeigten Treffern "**AAACL01000001**" ("**Nanoarchaeum equitans Kin4-M N.equitans_v1.0, whole genome shotgun sequence**") anklicken. Die jetzt angezeigte Genebank-Datei enthält noch HTML-Formatierungen, die beim Einlesen in Bioedit stören würden. Deshalb auf "Send to" mit Einstellung "File" oder "Text" klicken und die Datei im reinen Textformat abspeichern (passenden Namen wählen - z.B. Nanoarchaeum.txt).

Bioedit starten. Die Nanoarchaeum-Genomdatei mit "File" → "Open" laden. Unter "Sequence" → "Nucleic Acid" → "Sorted Six Frame Translation" wählen. In der angezeigten Dialogbox unter "Minimum ORF size" z.B. 40 (Aminosäuren) angeben. Auf "Translate" klicken und die resultierende Datei als Textdatei abspeichern.

Unter "Accessory Application" → "BLAST" → "Create a local protein database file" wählen und die eben abgespeicherte Datei mit den Aminosäuresequenzen auswählen. Es öffnet sich kurz ein MS-DOS-Fenster, das sich aber von selbst wieder schließt (die Datenbank ist jetzt fertig).

Unter "**Accessory Application**" → "**BLAST**" → "**Local BLAST**" wählen. In der angezeigten Dialogbox bei "**Program**" "**blastp**" wählen und bei "**Protein Database**" die neu generierte Datenbank wählen, die den selben Namen wie die Textdatei hat. Unter "**Load From File**" kann man eine Textdatei mit der Aminosäuresequenz laden, mit der man eine BLAST-Suche machen will. Oder Sequenz einfach in das Eingabefeld kopieren. Mit "**Do Search**" den BLAST-Lauf starten... eigentlich sollte sich jetzt ein Fenster mit der Liste der besten Übereinstimmungen öffnen. Diese Datei sollte man zur Dokumentation der Suche abspeichern.

Bei dem Alignment des "besten Treffers" (der zuoberst in der Liste steht), findet sich eine Formulierung wie z.B. "187775 to 185637: Frame - 3". Dieses Verweis hilft bei der Zuordnung der zugrundeliegenden Nucleotidsequenz zur gefundenen Aminosäuresequenz.

Um nun schließlich an die komplette Aminosäure-Sequenz zu gelangen, kopiert man einen Teil der unter "**Sbjct**" angegebenen Aminosäuren und sucht damit in der eingangs erstellten und abgespeicherten Datei mit den einzelnen Aminosäuresequenzen (d.h. die Datei, mit der die Datenbank generiert wurde). Sobald man den entsprechenden Eintrag gefunden hat kopiert man ihn am besten in einer neue Datei.

In den meisten Fällen fängt die Sequenz nicht mit einem Methionin an, so dass man noch einige Aminosäuren am Anfang löschen kann. Im Idealfall hat man jetzt die komplette Aminosäuresequenz des gesuchten Proteins vorliegen. Allerdings sollten mit weiteren Tests der Start und das Ende des Genes auf Plausibilität getestet werden (eine Möglichkeit hierfür bietet z.B. ein multiples Alignment mit homologen Proteinen anderer Arten).

8.4 „GenbankAnalyse“ – ein selbst geschriebenes Bioinformatik-Tool

„GenbankAnalyse“ wurde im Rahmen dieser Arbeit von mir mit Delphi 5 (Borland) programmiert. Es stellt eine Ergänzung zu anderen Bioinformatik-Programmen dar und ist auf Cut & Paste in enger Verbindung mit leistungsfähigen Texteditoren wie z.B. Ultraedit ausgelegt. Die von dem Programm generierten Grafiken werden in die Zwischenablage kopiert, von wo sie in ein beliebiges Grafikprogramm (oder z.B. Word) eingefügt werden können.

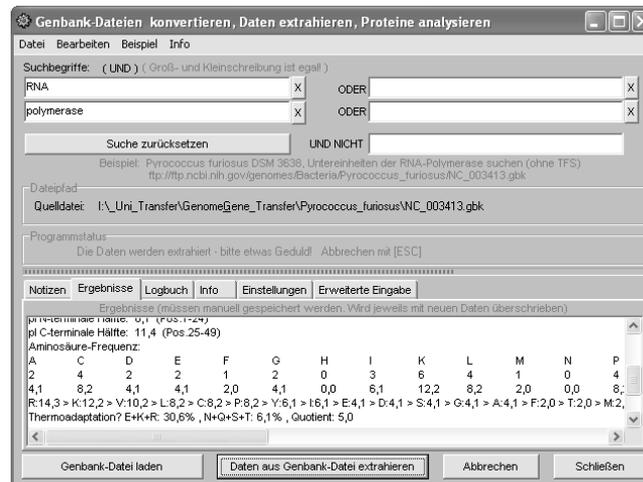


Abb.8-02: Screenshot des programmierten Bioinformatik-Tools "GenbankAnalyse"

Das Programm wurde zunächst entwickelt, um die Vielzahl von transkriptions-relevanten Gene aus archaellen und bakteriellen Genomen zu extrahieren und dabei jeweils gleich die meisten für diese Arbeit benötigten Basis-Analysen vorzunehmen (z.B. Berechnung von theoretischem Molekulargewicht, Ladung, isoelektrischem Punkt (pI), Aminosäure-Verteilung samt Auffälligkeiten, Zuordnung und Formatierung der DNA-Sequenz zur Aminosäure-Sequenz, Extraktion von Stromaufwärts-Region auf Nucleotidebene etc.; allein Tabelle [Abb.4-05] zeigt 211 RNA-Pol-Gene, die nicht alle in den Online-Datenbanken enthalten sind).

Später kamen Funktionen zur Analyse einzelner Proteine und zur Erstellung von kompakten / platzsparenden Grafiken hinzu.

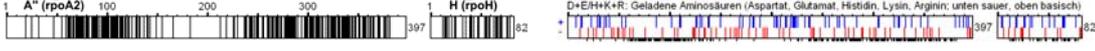
Funktionen des Programms in Stichworten:

- Einlesen von Genom-Dateien im Genbank-Format (*.gbk); Abspeichern der transformierten Daten in 5 verschiedenen Dateien / Listen:
 - a) Gene zeilenweise; Reihenfolge wie im Genom; Zusatzinfo über Abstand zum vorhergehenden Gen
 - b) Liste sortiert nach theoretischem Molekulargewicht
 - c) Statistik-Datei mit globalen Angaben zum Genom: GC-Gehalt, Häufigkeit von Start- und Stop-Codons, Länge der intergenen Region (Überschneidung, Wechsel der Leserichtung etc.). Tab-getrennte Tabelle mit Daten zum Thema „gekoppelte / überlappende Gene“.
 - d) Datei mit Ergebnissen einer Stichwortsuche; Zuordnung der Nucleotid-Sequenz zur Aminosäuresequenz (wenn nötig revers komplementär etc.). U.a. Berechnung der durchschnittlichen prozentualen Aminosäure-Häufigkeit in den gefundenen Protein-Genen (z.B. auch für alle Protein-Gene: bei P.furiosus z.B. 2065) sowie der "Codon-Usage".
 - e) Tab-getrennte Liste für Import in Excel: Abstand zum vorhergehenden Gen und Info über Wechsel der Leserichtung etc.; Start- und Stop-Position (Locus) des Genes; Aminosäure-Anzahl; theoretisches Molekulargewicht; Ladung bei pH 7,62; Isoelektrischer Punkt; Start- und Stop-Codon; Gen-Bezeichnung; „/product“-Angabe; „/note“-Angabe
- Analyse einzelner Proteine: die o.g. Daten; zusätzlich Infos über C-Terminus / N-Terminus; u.a. Analyse auf Indizien für Thermoadaptation

- Generierung von kompakten Säulen-Diagrammen zum Thema „Aminosäure-Häufigkeit“ (unsortiert / sortiert / farblich markiert / optionale „Normalisierung“).

Beispiel: 

- Generierung von Grafiken zum Thema: Alignment-Konsensus; Geladene Aminosäuren etc.

Beispiel: 

- Formatierung von Konsensus-Sequenzen (u.a. Umwandlung in ein ProSite-Suchmuster für die Markierung von Protein-Abschnitten in Cn3D über die Funktion „Find pattern“).
- Komplette Nucleotid-Sequenz aus Genbank-Datei extrahieren (z.B. für Import in Bioedit, falls die Genbank-Datei zu groß ist, um korrekt geladen zu werden).

Dieses Programm steht den Mitarbeitern der AG Thomm als Hilfsmittel u.a. bei der Fortführung des *Pyrococcus*-RNA-Polymerase-Projektes zur Verfügung.

Die Daten für die Implementierung der Berechnung des **theoretischen Molekulargewichtes** wurden folgender Datei entnommen: <http://jjorg.chem.unc.edu/personal/monroe/docs/mwthelp.txt>
Um die Angaben zu **Ladung** und **Isoelektrischem Punkt** besser einschätzen zu können, wird kurz auf das Prinzip bzw. die Daten eingegangen, die zu ihrer Berechnung verwendet werden:

8.4.1 Berechnung der theoretischen Nettoladung und des Isoelektrischen Punktes

Das Berechnen der Nettoladung bzw. des Isoelektrischen Punktes (pI) ist nicht ganz trivial, weil eine Reihe Voraussetzungen und Einschränkungen getroffen werden. Zusammenfassend kann gesagt werden, dass die resultierenden Werte nützlich und in vielen Fällen ausreichend, jedoch auf jeden Fall "theoretisch" sind, und praktische Experimente nicht komplett ersetzen können.

Ich habe mich weitgehend an der Funktion "PepStats" von HUSAR-Heidelberg (siehe 3.8) orientiert, deren Ergebnisse zur Verifizierung des von mir selbst (auf Grundlage von im Internet verfügbaren Informationen) implementierten Algorithmus diene.

Generell erfolgt die Berechnung auf der Basis einer Tabelle mit pK_a-Werten, die entsprechend der Henderson-Hasselbalch'schen Gleichung für jede Aminosäure einen Wert (Dissoziationskonstante) liefert, der zu der (theoretischen) Netto-Ladung des Proteins hinzuaddiert werden kann. Da hierbei die besondere Verteilung der Aminosäuren bzw. die tatsächliche 3D-Konformation nicht berücksichtigt wird, handelt es sich um eine Abschätzung. Im nativen Protein beeinflussen sich die in unmittelbarer Nähe befindlichen Aminosäure-Reste gegenseitig, so dass von der Theorie abweichende Dissoziationskonstanten resultieren.

Es muß beachtet werden, dass verschiedene pK_a-Wert-Tabellen mit leicht voneinander abweichenden Werten kursieren. Ich habe mich nach folgender Tabelle gerichtet, die auch PepStats standardmäßig verwendet (bezeichnender Weise kann man in PepStats auch eine eigene, alternative Tabelle benutzen):

pK data from Bull, "An Introduction to Physical Biochemistry", p128.

Residue	pK
C	8.3
D	3.91
E	4.25
H	6.50
K	10.79
R	12.50
Y	10.95
NH ₂	8.56 (amino terminus)
COOH	3.56 (carboxyl terminus)

Die Berechnung des Isoelektrischen Punktes (pI) erfolgt, indem die Netto-Ladung des Proteins für das gesamte pH-Spektrum ausgerechnet wird, und der Umgebungs-pH ermittelt wird, bei dem die Netto-Ladung gleich 0 ist.

Anhand der Aminosäuresequenz der Untereinheiten der RNAP aus *P.furiosus* und einiger Testproteine erfolgte die Überprüfung der von meinem Programm errechneten Werte.

Während sich mit den o.g. pK_a-Werten bei meinen eigenen Berechnungen ein identischer Wert für den pI ergibt, weicht die Angabe für die Netto-Ladung aus unerfindlichen Gründen etwas von PepStats ab (normalerweise nicht mehr als +/-1), was aber für die Gesamtaussage keinerlei Folgen hat, wenn man bedenkt, dass schon geringe Änderungen des Umgebungs-pH deutliche Abweichungen der Protein-Ladung nach sich ziehen. Die von PepStats errechneten Daten stimmen mit meinen Werten am besten überein, wenn ein pH von 7,62 zugrunde gelegt wird (dieser Wert wurde von mir im folgenden verwendet).

Folgende Beispiele zeigen einen Vergleich der Werte von PepStats und "GenbankAnalyse":
(in Klammern die von mir selbst errechneten Werte)

	Ladung	pI
rpoB:	-10 (-10)	6,70 (6,69)
rpoA1:	-6 (-6)	6,83 (6,82)
rpoA2:	-7 (-7)	5,86 (5,86)
rpoD:	-14 (-14)	4,71 (4,71)
rpoE1:	0 (0)	7,76 (7,76)
rpoF:	-8 (-8)	4,66 (4,65)
rpoK:	2 (2)	10,34 (10,34)
rpoP:	9 (8)	10,73 (10,72)
Test1:	57 (59)	??? (13,92) (ausschließlich basisch)
Test2:	-69 (-74)	2,19 (2,19) (ausschließlich sauer)

(die beobachtete Genauigkeit wurde als ausreichend betrachtet; bei den künstlichen (in natura nicht vorkommenden) Testsequenzen mit ausschließlich basischen / sauren Aminosäuren gibt es Abweichungen; "???" deutet an, dass PepStats hier keine Ergebnis lieferte).

Weitergehende Informationen zum Thema "Berechnung von Nettoladung und Isoelektrischem Punkt" finden sich unter folgenden Internet-Adressen:

<http://bioperl.org/pipermail/bioperl-l/2002-September/009280.html>

<http://fields.scripps.edu/DTASelect/20010710-pI-Algorithm.pdf>

Für die konkrete Programmierung erwiesen sich auch folgende Informationen als nützlich:

<http://www.biochem.uwo.ca/undergrad/381a/ta/tutorial6.doc>

8.5 Liste der mikrobiellen Genom-Dateien im Genbank-Format (Internet)

Allgemein Liste mit komplett sequenzierten Mikroorganismen:

<http://www.ncbi.nlm.nih.gov/genomes/Complete.html>

Liste aktueller, komplett sequenzierter archaeeller Genome:

http://www.ncbi.nlm.nih.gov:80/genomes/static/a_g.html

Aeropyrum pernix:

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Aeropyrum_pernix/NC_000854.gbk

Archaeoglobus fulgidus DSM 4304:

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Archaeoglobus_fulgidus/NC_000917.gbk

Halobacterium sp. NRC-1:

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Halobacterium_sp/NC_002607.gbk

Methanocaldococcus jannaschii:

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Methanocaldococcus_jannaschii/NC_000909.gbk

Methanopyrus kandleri AV19:

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Methanopyrus_kandleri/NC_003551.gbk

Methanosarcina acetivorans C2A:

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Methanosarcina_acetivorans/NC_003552.gbk

Methanosarcina mazei Goe1:

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Methanosarcina_mazei/NC_003901.gbk

Methanothermobacter thermoautotrophicus str. Delta H:

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Methanobacterium_thermoautotrophicum/NC_000916.gbk

Nanoarchaeum equitans Kin4-M:

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NC_005213

Pyrobaculum aerophilum:

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Pyrobaculum_aerophilum/NC_003364.gbk

Pyrococcus abyssi:

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Pyrococcus_abyssi/NC_000868.gbk

Pyrococcus furiosus DSM 3638:

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Pyrococcus_furiosus/NC_003413.gbk

***Pyrococcus horikoshii*:**

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Pyrococcus_horikoshii/NC_000961.gbk

***Sulfolobus solfataricus*:**

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Sulfolobus_solfataricus/NC_002754.gbk

***Sulfolobus tokodaii*:**

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Sulfolobus_tokodaii/NC_003106.gbk

***Thermoplasma acidophilum*:**

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Thermoplasma_acidophilum/NC_002578.gbk

***Thermoplasma volcanium*:**

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Thermoplasma_volcanium/NC_002689.gbk

Bakterielle Genome (Bacteria):

***Aquifex aeolicus*:**

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Aquifex_aeolicus/NC_000918.gbk

***Deinococcus radiodurans*:**

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Deinococcus_radiodurans/NC_001263.gbk

ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria/Deinococcus_radiodurans/AE000513.gbk

Escherichia coli CFT073

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NC_004431

***Helicobacter pylori 26695*:**

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Helicobacter_pylori_26695/NC_000915.gbk

***Salmonella typhimurium LT2*:**

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Salmonella_typhimurium_LT2/NC_003197.gbk

***Staphylococcus aureus subsp. aureus MW2*:**

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Staphylococcus_aureus_MW2/NC_003923.gbk

***Thermotoga maritima*:**

ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Thermotoga_maritima/NC_000853.gbk

8.6 Accession-Nummern einiger in dieser Arbeit erwähnter Gene

***Saccharomyces cerevisiae* (Gene der Pol II):**

Die folgenden Gene finden sich z.B. auch bei Swissprot unter RPB1_YEAST, RPB2_YEAST, ...

Rpb1:	P04050	(Synonyme: RPO21, RPB220, SUA8, YDL140C, D2150)
Rpb2:	P08518 / NP_014794	(Synonyme: RPO22, RPB150, YOR151C)
Rpb3:	P16370 / NP_012243	(Synonyme: YIL021W)
Rpb4:	P20433 / NP_012395	(Synonyme: YJL140W, J0654)
Rpb5:	P20434 / NP_009712	(Synonyme: YBR154C, YBR1204)
Rpb6:	P20435	(Synonyme: RPO26, YPR187W, P9677.8)
Rpb7:	P34087	(Synonyme: YDR404C, D9509.22)
Rpb8:	P20436 / NP_014867	(Synonyme: YOR224C, YOR50-14)
Rpb9:	P27999	(Synonyme: YGL070C)
Rpb10:	P22139 / NP_014853	(Synonyme: YOR210W)
Rpb11:	P38902 / NP_014638	(Synonyme: YOL005C)
Rpb12:	P40422	(Synonyme: RPC10, YHR143BW)

***Escherichia coli* (Gene der bakteriellen RNA-Polymerase):**

β':	P00577	swissprot: RPOC_ECOLI
β:	P00575	swissprot: RPOB_ECOLI
α:	P00574	swissprot: RPOA_ECOLI
ω:	P08374	swissprot: RPOZ_ECOLI
σ^{70}:	P00579	swissprot: RPSD_ECOLI (Standard- σ -Faktor)
σ^{54}:	P24255	swissprot: RP54_ECOLI (Arginin-Metabolismus)

***Sulfolobus acidocaldarius* (Gene der RNA-Polymerase):**

(*S.acidocaldarius* ist noch nicht komplett sequenziert, wird aber an verschiedenen Stellen zitiert bzw. in Alignments mit einbezogen, da an diesem Organismus wichtige Untersuchungen zu Thema "RNA-Polymerase bei Archaea" durchgeführt wurden (Lanzendörfer *et al.*, 1994)).

rpoA/B/C:	X14818	(Operon u.a. mit Untereinheit B, A', A")
rpoD/N:	X80194	(Operon mit rpoD, rpoN und ribosomalen Proteinen)
B:	P11513 / S04716	swissprot: RPOB_SULAC

A':	P11512 / B33926	swissprot: RPA1_SULAC
A'':	P11514	swissprot: RPA2_SULAC
rpoD:	P39471	swissprot: RPOD_SULAC
rpoE:	P39466 / S38658	swissprot: RPOE_SULAC
rpoH:	P11521	swissprot: RPOH_SULAC
rpoK:	P39463	swissprot: RPOK_SULAC
rpoL:	P46217	swissprot: RPOL_SULAC
rpoN:	P39472	swissprot: RPON_SULAC

8.7 Abkürzungen und Begriffe

Alignment	Kurzform für "Multiple alignment"
bp	Basenpaare (DNA)
Da	Dalton
extrem thermophil:	Organismus wächst bei Temperaturen > 75°C
Genbank	Einen der wichtigsten, online verfügbaren Gen-Datenbanken
GTG	Alternatives Start-Codon (bei <i>Archaea und Bacteria</i>)
hyperthermophil:	Organismus wächst bei Temperaturen > 90°C
kDa	Kilodalton
mesophil:	Organismus wächst bei 20-45°C
MG	Molekulargewicht
NCBI	National Center for Biotechnology Information
NusA	transcription termination-antitermination factor NusA
NusG	transcription termination-antitermination factor NusG
ORF	Open reading frame
PA	Polyacrylamid
PAGE	Polyacrylamid Gel Elektrophorese
pI	Isoelektrischer Punkt
PIC	Präinitiationskomplex
PKA	Proteinkinase A (zur radioaktiven Markierung geeigneter Protein)
Pol I / Pol II / Pol III:	Eurkaryotische DNA-abhängige RNA-Polymerase
psychrophil:	(auch „cryophil“) Organismus wächst bei Temperaturen < 20°C
rbs	Ribosomen-Bindestelle
RNA-Pol	In dieser Arbeit ist immer die DNA-abhängige RNA-Polymerase gemeint
RNAP	(siehe oben)
SRS:	Sequence Retrieval System (Gene im Internet / Datenbanken suchen)
thermophil:	Organismus wächst bei Temperaturen > 45°C
TBP	TATA-Box-Bindeprotein
TFB	(archaeeller) Transkriptionsfaktor B (Initiation)
TFE	(archaeeller) Transkriptionsfaktor TFE (Initiation)
TFS	(archaeeller) Transkriptionsfaktor TFS (Elongation)
TTG	Alternatives Start-Codon (bei <i>Archaea und Bacteria</i>)
UE	Untereinheiten (der RNA-Polymerase)

8.8 Abbildungsverzeichnis

Abb.2-01: Stammbaum mit <i>Pyrococcus furiosus</i> , einem hyperthermophiler Vertreter der <i>Archaea</i>	3
Abb.2-02: Anzahl und Größe der RNA-Pol-Untereinheiten bei <i>Eukarya</i> , <i>Archaea</i> und <i>Bacteria</i>	5
Abb.2-03: Schematische 3D-Darstellung: Prinzipielle Arbeitsweise einer RNA-Polymerase	6
Abb.3-01: Prinzip der Far Western Blotting-Methode.....	17
Abb.3-02: Schematische Darstellung des Vektors pET-33b(+) und Variante pET-33b(+) ₂	19
Abb.3-04: Konstruktion und "Multiple Cloning Site" von pET-33b(+) ₂ im Detail	21
Abb.3-05: Beispiel für Far Western Dot-Blot	22
Abb.3-06: Blockierlösung für Far Western Blotting-Experimente nach (Burgess <i>et al.</i> , 2000)	24
Abb.4-01: Anteil der einzelnen RNA-Pol-Untereinheiten am Gesamtkomplex	29
Abb.4-02: Größen der homologen RNA-Pol-Untereinheiten bei <i>P.furiosus</i> und <i>S.cerevisiae</i>	30
Abb.4-03: Gene und allgemeine Eigenschaften der RNAP-Untereinheiten in <i>P.furiosus</i>	30
Abb.4-04: Liste der gegenwärtig verfügbaren, komplett sequenzierten archaeellen Genome	31
Abb.4-05: Tabelle mit den Namen der Gene und den theoretischen Molekulargewichten der RNA-Pol-Untereinheiten bei 17 komplett sequenzierten <i>Archaea</i>	32
Abb.4-06: Tabellarische Darstellung statistischer Rahmendaten der archaeellen Genome.....	34
Abb.4-07: Aminosäure-Austausch in den RNAP-Genen innerhalb der Gattung <i>Pyrococcus</i>	36
Abb.4-08: Aminosäure-Austausch an Positionen, die bei den anderen Schwesterarten identisch sind.....	36
Abb.4-09: "Protein-Scoring-Systeme": Gruppierung ähnliche und unähnlicher Aminosäuren.....	38
Abb.4-10: Eigenschaften von rpoB aus <i>P.furiosus</i>	39
Abb.4-11: Eigenschaften von rpoA1 aus <i>P.furiosus</i>	40
Abb.4-12: Eigenschaften von rpoA2 und rpoH aus <i>P.furiosus</i>	41
Abb.4-13: Eigenschaften von rpoD und rpoL aus <i>P.furiosus</i>	42
Abb.4-14: Eigenschaften von rpoE1, rpoE2 und rpoF aus <i>P.furiosus</i>	42
Abb.4-15: Eigenschaften von rpoK, rpoN und rpoP aus <i>P.furiosus</i>	44
Abb.4-16: Übereinstimmung der RNA-Pol-Untereinheiten aus <i>P.furiosus</i> mit dem Konsensus	44
Abb.4-17: Vergleichende grafische Darstellung von rpoA1 (Rpb1, β')	45
Abb.4-18: Proteinstammbäume für rpoA1 (Rpb1 / β') und rpoL (Rpb11 / α).	46
Abb.4-19: Länge der intergenen (intercistronischen) Regionen im Genom von <i>P.furiosus</i>	48
Abb.4-20: Gegenüber <i>Archaea</i> konservierte Genabschnitte in Rpb1 aus <i>S.cerevisiae</i>	49
Abb.4-21: Gegenüber <i>Archaea</i> konservierte Genabschnitte in Rpb2 aus <i>S.cerevisiae</i>	51
Abb.4-22: Gegenüber <i>Archaea</i> konservierte Genabschnitte in Rpb3 und Rpb11 aus <i>S.cerevisiae</i>	51
Abb.4-23: Gegenüber <i>Archaea</i> konservierte Genabschnitte in Rpb7 und Rpb4 aus <i>S.cerevisiae</i>	52
Abb.4-24: Gegenüber <i>Archaea</i> konservierte Genabschnitte in Rpb5 und Rpb6 aus <i>S.cerevisiae</i>	53
Abb.4-25: Gegenüber <i>Archaea</i> konservierte Genabschnitte in Rpb10 und Rpb12 aus <i>S.cerevisiae</i>	53
Abb.4-26: Nicht aufgelöste Abschnitte im 3D-Modell der <i>S.cerevisiae</i> -RNA-Polymerase.....	55
Abb.4-27: 3D-Struktur inklusive Visualisierung der gegenüber <i>Archaea</i> ähnlichen Regionen	56
Abb.4-28: 3D-Struktur mit gegenüber <i>Archaea</i> konservierten Regionen, Gesamtkomplex.....	57
Abb.4-29: Prozentuale Aminosäurehäufigkeit in den RNA-Pol-Genen von <i>P.furiosus</i>	59
Abb.4-30: Indizien für Thermoadapation. Vergleich von <i>P.furiosus</i> und <i>M.mazei</i>	60
Abb.4-31: Prozentuale Aminosäurefrequenz nach "Normalisierung".....	61
Abb.4-32: Häufigkeit von E+K+R gegenüber N+Q+S+T in den RNA-Pol-Genen nach Umrechnung	61
Abb.4-33: Hitzestabilität und Löslichkeit der RNAP-Untereinheiten von <i>P.furiosus</i> in der Praxis.....	63
Abb.4-34: Thermoadaptation? Aminosäurefrequenz in den Proteingenen von <i>Archaea</i> und <i>Bacteria</i>	65
Abb.4-35: Übersicht: Aminosäurefrequenz vor dem Hintergrund potentieller Thermoadaptation	66

Abb.4-36: E+K+R gegenüber N+Q+S+T in hyperthermophilen, thermophilen und mesophilen Arten.....	67
Abb.4-37: Alignment der TFB-Kopien aus <i>P.furiosus</i> , <i>P.horikoshii</i> und <i>P.abysyi</i>	69
Abb.4-38: Vergleichende Darstellung der Struktur und Ladungsverteilung in TFB und TFB-2	70
Abb.4-39: Gegenüberstellung rekombinanter RNAP-Untereinheiten und der nativen RNAP.....	73
Abb.4-40: In <i>E.coli</i> seltene Codons in für die Transkription relevanten Proteingenen in <i>P.furiosus</i>	74
Abb.4-41: Frequenz der verschiedenen Codons (Codon Usage) in <i>P.furiosus</i>	74
Abb.4-42: Ausgewählte Ergebnissen aus Far Western Dot-Blot-Experimenten	77
Abb.4-43: Ponceau S-Färbung einer Dot Blot-Membran: Optimierung der Proteinmengen.....	78
Abb.4-44: Far Western Dot Blotting-Experimente: TBP- und TFB-Verdünnungsreihe	79
Abb.4-45: Ergebnisse aus Far Western Blotting-Experimenten mit rpoD, rpoL, TFB u. TBP als Sonde.....	81
Abb.4-46: Far Western Blot mit rpoD und rpoK als Sonde gegenüber teilweise deletiertem TFB.....	82
Abb.4-47: Tabellarische Übersicht der beobachteten Bindungsstärken beim Far Western Blotting	83
Abb.4-48: 3D-Struktur des Rpb3-Rpb10-Rpb11-Rpb12 Komplexes (homolog zum DLNP-Komplex)	84
Abb.4-49: 3D-Struktur von Rpb5, Rpb6, Rpb7-Rpb4 (homolog zu rpoH, rpoK und rpoE1-rpoF)	85
Abb.4-50: Schematische Darstellung von Bindungsstärken zwischen den RNAP-Untereinheiten.....	87
Abb.4-51: Befinden sich Rpb3/Rpb11/Rpb12 (rpoD/rpoL/rpoP) in Reichweite von TFIIB (TFB)? [1].....	88
Abb.4-52: Befinden sich Rpb3/Rpb11/Rpb12 (rpoD/rpoL/rpoP) in Reichweite von TFIIB (TFB)? [2].....	88
Abb.4-53: Bildung eines DL- und E'F-Komplexes in Nativgel-Experimenten.....	89
Abb.4-54: Bildung eine DL- und DN-Komplexes in Nativgel-Experimenten.....	89
Abb.4-55: Ein DL-Komplex und mutmaßlicher DLN-Komplex bei Molekularsieb-Chromatographie	90
Abb.4-56: Ausführliches Schema der RNAP-Reinigung	92
Abb.4-57: Western Blot mit Antikörpern gegen Untereinheit F	93
Abb.4-58: Ausschnitt einer PVDF-Membran mit den "geblotteten" kleinen RNAP-Untereinheiten.....	95
Abb.4-59: Zuordnung von RNAP-Untereinheiten auf der Basis von MALDI-TOF-MS Daten	95
Abb.4-60: Auswertung der MALDI-TOF-MS-Daten von einem Gemisch aus RNAP-Untereinheiten	96
Abb.8-01: Festlegung von "Ähnlichkeit": ClustalX, PAM 250-Matrix.....	125
Abb.8-02: Screenshot des programmierten Bioinformatik-Tools "GenbankAnalyse"	130

8.9 Danksagung

An erster Stelle möchte ich mich bei Prof.Dr.M.Thomm dafür bedanken, dass er mir die Bearbeitung des komplexen "RNA-Polymerase"-Themas ermöglicht hat. Ohne sein Interesse, die Anregungen und konstruktive Kritik, sowie seine allgemeine Unterstützung und Geduld hätte diese Arbeit nicht realisiert werden können.

PD Dr.W.Hausner danke ich für seine kontinuierliche Bereitschaft zur Diskussion theoretischer und praktischer Details, was sich in vielen Fällen hilfreich bei der Lösung auftretenden Probleme erwies.

Für angeregte Diskussionen und gute Zusammenarbeit im Rahmen ihrer Diplomarbeiten möchte ich Arne Lange (Lange, 2001) und Oliver von Kampen (von Kampen, 2003) danken.

Souad Naji, die meine Arbeit an der Universität Regensburg fortführt, möchte ich für ihre konstruktiven Ideen (z.B. zur Darstellung von Protein-Komplexen mittels Native-PAGE) und gute Zusammenarbeit bei der Übergabe des Themas danken.

Jens Thomsen und Udo Lange danke ich als meinen wichtigsten Laborkollegen (neben den "fachlichen Dingen") für die gute Atmosphäre und das freundschaftliche "Betriebsklima".

Jutta Kock danke ich für hilfreiche Hinweise zu manchem Detail im praktischen Laboralltag und großzügige Aushilfe bei aktuell fehlenden Reagentien.

Dagmar Haacks gilt besonderer Dank für den "organisatorischen Komplettservice" und die regelmäßigen Aufmunterungen für "die Drei aus Lab 111".

Brigitte Vass danke ich dafür, dass wir uns aufgrund ihrer Arbeit keinerlei Gedanken über die Reinigung von Laborgefäßen machen mußten, und auf diese Weise unsere Experimente zügig durchführen konnten.

Natürlich haben auch all die interessanten Gespräche und die gute Zusammenarbeit mit Isabell Dahlke, Sebasti n Fran ois, Gudrun Vierke, Patrizia Spitalny, Carina Hebbeln und allen weiteren hier nicht namentlich genannten Mitarbeitern am Ifam zum Gelingen dieser Arbeit beigetragen, wof r ich mich neben der hieraus resultierenden insgesamt freundlichen und herzlichen Arbeitsatmosph re besonders bedanken m chte.

Randolf Bodle danke ich f r die kritische Durchsicht eines Teils dieser Arbeit.

8.10 Lebenslauf

Persönliche Information

- Name: Bernd Goede
- Geburtsdatum: 18.05.66
- Geburtsort: Marl
- Nationalität: deutsch
- Familienstand: ledig

Schule

- 1972 – 1974: Grundschule in Oerlinghausen (Bielefeld, NRW)
- 1974 – 1976 : Grundschule am Heidenberger Teich in Kiel
- 1976 – 1985: Gymnasium im BZM in Kiel
- 6/85 Allgemeine Hochschulreife (Abitur)

Zivildienst

- 8/85 bis 3/87 in der Ev.-Luth.Thomas-Kirchengemeinde Kiel-Mettenhof
- 3/87 bis 9/88 „Studium Generale“ an der Christian-Albrechts-Universität zu Kiel (Kunstgeschichte, Literaturwissenschaft, Philosophie, Psychologie)

Ausbildung

- 7/88 bis 3/90: Ausbildung zum „Medizinisch- technischen Laboratoriums-assistenten“ (MTA) an der Abteilung Hygiene, Sozialhygiene und Gesundheitswesen im Klinikum der Christian-Albrechts-Universität Kiel.

Studium

- 9/91 bis 3/97: Biologie-Studium an der Christian-Albrechts-Universität zu Kiel
- 3/97 bis 5/98: Diplomarbeit am „Institut für Allgemeine Mikrobiologie“ der CAU Kiel in der Arbeitsgruppe von Prof.Dr.M.Thomm über das Thema „Untersuchungen zur Funktion des archaellen Transkriptionsfaktors TFB aus *Pyrococcus furiosus*“.

Doktorarbeit

- 9/98 bis 2/04: Doktorarbeit in der Arbeitsgruppe von Prof.Dr.M.Thomm zum Thema „Protein-Protein-Wechselwirkungen der Untereinheiten der RNA-Polymerase aus *Pyrococcus furiosus*“.

8.11 Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst und nur die von mir angegebenen Quellen und Hilfsmittel verwendet habe.

Diese Arbeit war bisher noch nicht Bestandteil eines Prüfungsverfahrens, andere Promotionsversuche wurden nicht unternommen.