

**Functional proteomics
of the lytic bacteriophages
Cp-1 and Dp-1
of *Streptococcus pneumoniae***

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

Fakultät für Chemie und Biowissenschaften
Karlsruher Institut für Technologie (KIT) – Universitätsbereich

vorgelegte
DISSERTATION

von
Dipl.-Biol. Roman Häuser
aus Pforzheim

Dekan: Prof. Dr. Stefan Bräse

Referent: Prof. Dr. Jonathan Sleeman

Korreferent: Prof. Dr. Jörg Kämper

Tag der mündlichen Prüfung: 20.10.2010

Diese Arbeit wurde in der Arbeitsgruppe von PD Dr. Peter Uetz am Institut für Toxikologie und Genetik (ITG) des Karlsruher Institut für Technologie (KIT) und am J. Craig Venter Institute (JCVI) in Rockville, MD, USA, angefertigt.

I. Abstract

Bacteriophages

Bacteriophages (phages) are viruses that infect and kill bacteria. Hence, they play an important role in medicine, ecology, evolution, food production, and many other microbiological applications and processes. A few model phages have been investigated in great detail and thousands of completely sequenced phage genomes are now available. However, there are hardly any comprehensive, functional analyses of phage genomes in the literature although modern methods allow for proteome-wide investigations.

In this work I focused on the phages Cp-1 and Dp-1 that infect *Streptococcus pneumoniae*, a major human pathogen, by using genomic and proteomic approaches to get deeper insights into their biology. Cp-1 belongs to the *Podoviridae* family and Dp-1 to the *Siphoviridae*. Both are rarely isolated *S. pneumoniae* phages that have a virulent (lytic) reproduction cycle and thus are of general interest for medical application in phage therapy. However, little is known about their biology and their interplay with the host is poorly understood.

In this work I comprehensively annotate the Dp-1 genome. It is 56,506 bp in length and encodes for 72 putative proteins. 69 proteins are homologous to known protein sequences in databases. However, by using a homology-based annotation strategy only 43 Dp-1 proteins can be functionally annotated. This emphasizes that functional predictions by homology comparisons are often limited and that there is an urgent need to functionally analyze phage proteomes in a comprehensive manner.

This work describes the most comprehensive analysis of phage protein interaction networks. To learn more about the interplay of all Cp-1 and Dp-1 proteins, I systematically analyzed their proteomes by Yeast Two-Hybrid screens. The intra-viral protein-protein interaction (PPI) network of Cp-1 reveals 17 individual PPIs and that of Dp-1 156 individual PPIs. In this work I demonstrate that the identified interactions can be used to model the Cp-1 and Dp-1 virions by linking a significant number of uncharacterized phage proteins to virion morphogenesis.

Bacteriophage reproduction is always dependent on a host cell. A major aim of this project was thus to analyze the interactions of phage proteins with the host proteome, given that the phage is absolutely dependent on host activities to reproduce. Therefore, I screened all Cp-1 and Dp-1 proteins against a *S. pneumoniae* Yeast Two-Hybrid library to identify binary phage-host PPIs. Such a comprehensive analysis has never been done before for any phage. This analysis identified 11 interactions for Cp-1 proteins and 38 interactions for Dp-1 proteins with *S. pneumoniae* proteins. Although the Cp-1 and Dp-1 interaction patterns are remarkably different, their proteins bind predominantly with host proteins that have regulatory roles, that are involved in DNA-related processes, and that are essential for the host. The detected PPIs indicate that Cp-1 and Dp-1 are able to modify several host pathways

by attacking or recruiting specific key components or to reprogram its gene expression through blocking host repressor proteins.

YbeB

An independent aim of this work was to study the function of *Escherichia coli* YbeB and its orthologs, a conserved hypothetical protein family of hitherto unknown function. YbeB and its orthologs are nearly universally conserved from bacteria to man. In this work I show that YbeB family members bind in a conserved manner to ribosomal protein L14 in most tested species, including *E. coli*, *Treponema pallidum*, *Synechocystis PCC 6803*, *Streptococcus pneumoniae* as well as human mitochondria and *Zea mays* chloroplasts. This interaction is thus conserved from bacteria to man. The human YbeB-ortholog, C7orf30, is targeted exclusively to the mitochondrial compartment and it interacts with the mitochondrial L14 *in vivo*. I show that *E. coli* L14 is the conserved and specific docking site of YbeB on the large ribosomal subunit and that L14's interaction epitope includes highly conserved amino acid residues that normally are involved in contacting the 16S rRNA of the small ribosomal subunit. *In vivo* analyses of an *E. coli ybeB* gene deletion strain reveal that the knock-out of *ybeB* results in premature reporter protein synthesis. *In silico* docking of YbeB with L14 supports these findings: once YbeB is bound to L14 on the large ribosomal subunit the assembly of the large and small ribosomal subunit is blocked by YbeB. The results of this work provide evidence that L14 is the primary and specific docking site of YbeB and its orthologs and that YbeB functions as a negative modulator of protein translation *in vivo*.

II. Zusammenfassung

Bakteriophagen

Bakteriophagen (Phagen) sind Viren, die Bakterien infizieren. Sie spielen deshalb eine wichtige Rolle in der Medizin, Ökologie, Evolution, Nahrungsmittelherstellung und vielen anderen mikrobiellen Applikationen und Prozessen. Ein paar wenige Modellphagen wurden bereits detailliert charakterisiert und tausende komplett-sequenzierte Phagengenome sind mittlerweile bekannt. Jedoch gibt es kaum umfangreiche, funktionelle Analysen von Phagengenomen in der bisherigen Literatur, obwohl moderne Methoden Proteom-weite Untersuchungen hier ermöglichen würden.

In dieser Arbeit präsentiere ich Ergebnisse funktionell-genomischer und -proteomischer Analysen der Phagen Cp-1 und Dp-1, die *Streptococcus pneumoniae*, ein human-pathogenes Bakterium, infizieren. Cp-1 gehört der *Podoviridae*- und Dp-1 der *Siphoviridae*-Familie an. Beide Phagen sind zwei von nur wenigen bekannten *S. pneumoniae*-Phagen, die einen virulenten (lytischen) Reproduktionszyklus haben und daher von Interesse für die medizinische Anwendung in der Phagentherapie sind. Allerdings ist nur wenig über ihre Biologie und dem Zusammenspiel mit ihrem Wirt bekannt.

Diese Arbeit zeigt die Ergebnisse einer detaillierten Genomanalyse von Dp-1. Das Genom umfasst 56,506 bp und kodiert für insgesamt 72 Proteine. 69 Proteine sind homolog zu bekannten Proteinsequenzen in Datenbanken. Allerdings können durch eine Homologie-basierte Annotationsstrategie nur 43 Dp-1-Proteine funktionell annotiert werden. Dies zeigt, dass funktionelle Vorhersagen durch Homologievergleiche oft nur eingeschränkt möglich sind und dass es wichtig ist, Phagenproteome experimentell zu analysieren.

Ich zeige daher die bisher umfangreichsten Protein-Protein-Interaktions-Netzwerke von Phagen. Um mehr über das Zusammenspiel der Cp-1- und Dp-1-Phagenproteine lernen zu können, habe ich die kompletten Phagenproteome mittels systematischer Yeast-Two-Hybrid-Screens analysiert. Das intravirale PPI-(Protein-Protein-Interaktions)-Netzwerk von Cp-1 besteht aus 17 individuellen PPIs, das von Dp-1 aus 156 PPIs. Ich zeige, dass die identifizierten Interaktionen hilfreich sind, um Modelle ihrer Virione zu erstellen und dass vermutlich ein umfangreiches Set an Phagenproteinen am Morphogeneseprozess involviert ist.

Die Reproduktion von Bakteriophagen erfolgt immer in Abhängigkeit einer Wirtszelle. Deshalb war es eine besondere Bestrebung dieser Arbeit herauszufinden, welche Wirtsprozesse von Cp-1 und Dp-1 auf der Protein-Interaktionsebene attackiert werden können, da eine Notwendigkeit besteht, den Wirt für eine effiziente Reproduktion zu optimieren. Dafür screente ich alle Cp-1- und Dp-1-Proteine gegen einen *S. pneumoniae*-Prey-Array. Solch eine umfangreiche Untersuchung wurde für Phagen bisher noch nicht durchgeführt. Die Analyse identifizierte 11 Interaktionen für Cp-1- und 38 Interaktionen für Dp-1-Proteine mit *S. pneumoniae*-Proteinen. Obwohl die Cp-1- und Dp-1-Interaktionsmuster auffällig unterschiedlich sind, scheinen ihre Proteine bevorzugt mit Wirtsproteinen zu binden, die eine

regulatorische Rolle haben, die in den DNA-Metabolismus involviert sind, oder die essentiell für den Wirt sind. Die detektierten PPIs zeigen an, dass Cp-1 und Dp-1 vermutlich mehrere Wirtsprozesse modifizieren können, z.B. indem sie spezifisch Schlüsselkomponenten attackieren oder rekrutieren oder indem sie die Genexpression durch Blockieren von Repressorproteinen umprogrammieren.

YbeB

Ein unabhängiges Ziel dieser Arbeit war, die Funktion des *Escherichia coli*-Proteins YbeB zu studieren. Es handelt sich hierbei um ein konserviert-hypothetisches Protein von bisher unbekannter Funktion. YbeB und seine Orthologe sind fast universell konserviert, mit Vertretern in Eubakterien bis hin zum Menschen. In dieser Arbeit zeige ich, dass YbeB-Familienmitglieder in den meisten getesteten Spezies mit dem ribosomalen Protein L14 interagieren, einschließlich *E. coli*, *Treponema pallidum*, *Synechocystis PCC 6803*, *Streptococcus pneumoniae*, humanen Mitochondrien und *Zea mays*-Chloroplasten. Weiterhin zeige ich, dass das humane YbeB-Ortholog C7orf30 exklusiv in Mitochondrien lokalisiert und dass es mit mitochondrialem L14 *in vivo* interagiert. Desweiteren wird gezeigt, dass *E. coli* L14 die konservierte und spezifische Bindungsstelle für YbeB auf der großen ribosomalen Untereinheit ist und dass das L14-YbeB-Interaktionsepitop hoch-konservierte Aminosäurereste enthält, die normalerweise die Interaktion mit der 16S rRNA der kleinen ribosomalen Untereinheit vermitteln. *In vivo*-Analysen einer *E. coli ybeB*-Gendelektionsmutante heben hervor, dass der Knock-Out von *ybeB* in der verfrühten Synthese von Reporterproteinen resultiert. *In silico*-Docking unterstützt die experimentellen Befunde: Wenn YbeB über L14 an die große ribosomale Untereinheit bindet, dann kommt es zur sterischen Blockade bei der Assemblierung zwischen den ribosomalen Untereinheiten. Die Ergebnisse dieser Arbeit beweisen, dass L14 die primäre und spezifische Bindungsstelle von YbeB und seinen Orthologen auf der großen ribosomalen Untereinheit ist und dass YbeB als negativer Modulator in der Proteintranslation fungiert.

III. Danksagung/Acknowledgements

Hiermit möchte ich mich recht herzlich bei allen Personen bedanken, die diese Arbeit unterstützt haben.

Zuerst möchte ich mich bei PD Dr. Peter Uetz bedanken, der mir diese Arbeit in seinem Labor ermöglicht hat.

Ich bedanke mich außerdem bei Herrn Prof. Dr. Jonathan Sleeman und Herrn Prof. Dr. Jörg Kämper für die Erstellung der Gutachten dieser Arbeit.

Ein weiteres Dankeschön gilt all meinen Kollaborationspartnern, die an dieser Arbeit mitgewirkt haben (Thorsten Stellberger, Dr. Björn Titz, Dr. Markus E. Diefenbacher, Florian Naeve, Prof. Daniel Nelson, PhD Caren Stark, PhD Mourad Sabri, Siham Ouennane, Prof. Sylvain Moineau, PhD Seesandra V. Rajagopala, PhD Andrey Tovchigrechko und PhD Maria Guadalupe Vizoso-Pinto).

Weiterhin möchte ich meinen Instituten, dem Institut für Toxikologie und Genetik und dem J. Craig Venter Institute, danken, besonders Dr. Sandra Schneider und Prof. Uwe Strähle.

Zudem bedanke ich mich bei dem Karlsruhe House of Young Scientists, das mir durch ein Stipendium meinen Auslandsaufenthalt am J. Craig Venter Institute ermöglicht hat, sowie der Landesstiftung Baden-Württemberg.

Ich danke allen Einrichtungen und Wissenschaftlern, die mir Plasmide, Bakterienstämme und Chemikalien zur Verfügung gestellt haben.

Vor allem möchte ich mich bei meiner Ehefrau, meinen Eltern, Schwiegereltern und Freunden bedanken, die mich während der arbeitsintensiven Promotionszeit stets unterstützten und motivierten.

IV. Lebenslauf/Curriculum Vitae

PERSÖNLICHE DATEN

Geburtsdatum 29.05.1979 in Pforzheim
Staatsangehörigkeit deutsch
Familienstand verheiratet

AUSBILDUNG

seit 10/2006 Promotion im Fach Biologie, Karlsruher Institut für Technologie (KIT), Institut für Toxikologie und Genetik (ITG) in der Gruppe von PD Dr. Peter Uetz für funktionelle Proteomik
Titel der Dissertation: "Functional proteomics of the lytic bacteriophages Cp-1 and Dp-1 of *Streptococcus pneumoniae*"

02/2008 - 11/2008 Auslandsaufenthalt im Rahmen der Promotionsarbeit, J. Craig Venter Institute, Rockville, MD, USA (gefördert durch ein Stipendium des Karlsruhe House of Young Scientists)

10/2000 - 10/2006 Studium der Biologie (Grund- und Hauptstudium), Karlsruher Institut für Technologie (KIT), Studienschwerpunkte im Hauptstudium: Genetik und Zoologie (Hauptfächer), Biochemie und Organische Chemie (Nebenfächer)
Titel der Diplomarbeit: "Functional analysis of conserved hypothetical proteins of bacteria", Abschluss als Diplom-Biologe (Note 1,1)

1999 Abitur, Johanna-Wittum-Schule (Ernährungswissenschaftliches Gymnasium), Pforzheim. Buchpreis des Fonds der Chemischen Industrie

SONSTIGES

1999-2000 Zivildienst, Evangelische Kirchengemeinde Eisingen

V. Publications

Authorships of Roman Häuser

- Titz B., Rajagopala S.V., Ester C., **Häuser R.**, and Uetz P. (2006). Novel conserved assembly factor of the bacterial flagellum. *J. Bacteriol.* 188(21):7700-6
- Titz B., **Häuser R.**, Engelbrecher A., and Uetz P. (2007). The *Escherichia coli* protein YjjG is a house-cleaning nucleotidase *in vivo*. *FEMS Microbiol. Lett.* 270(1):49-57
- Titz B., Rajagopala S.V., Goll J., **Häuser R.**, McKevitt M.T., Palzkill T., and Uetz P. (2008). The binary protein interactome of *Treponema pallidum*-the syphilis spirochete. *PLoS One.* 3(5):e2292.
- Stellberger T., **Häuser R.**, Baiker A., Pothineni V.R., Haas J., and Uetz P. (2010). Improving the yeast two-hybrid system with permuted fusions proteins: the Varicella Zoster Virus interactome. *Proteome Sci.* 15;8:8.
- Sabri M.*, **Häuser R.***, Ouellette M., Liu J., Dehbi M., Moeck G., Garcia E., Titz B., Uetz P., and Moineau S. Genome annotation and intra-viral interactome of the *Streptococcus pneumoniae* virulent phage Dp-1. Submitted to *J. Bacteriol.*
- Häuser R.**, Titz B., Naeve F., Tovchigrechko A., Stellberger T., Diefenbacher E. M., and Uetz P. Functional analysis of YbeB/Iojap/DUF143 – a family of nearly universally conserved accessory translation factors that bind via L14 to the large ribosomal subunit. Submitted to *BMC Biology*.
- Häuser R.**, Vizoso Pinto M. G., and Uetz P. Protein-protein interactions of the bacteriophages Cp-1 and Dp-1 with their host *Streptococcus pneumoniae*. Manuscript in preparation.

*equal contribution

Book chapters

- Häuser R.**, Stellberger T., Rajagopala S.V., and Uetz P. Array-Based Yeast Two-Hybrid Screens: A Practical Guide. *Methods in Molecular Biology*, Springer. In press.
- Häuser R.**, Stellberger T., Rajagopala S.V., and Uetz P. Matrix-Based Yeast Two-Hybrid Screen Strategies and Comparison of Systems. *Methods in Molecular Biology*, Springer. In press.

VI. Table of contents

1 Introduction	1
1.1 Bacteriophages – a brief overview	1
1.1.1 Relevance of phages – what they are (not) good for	1
Phages in medicine	1
Phages in ecology	2
Their relevance in food production and industrial processes	2
Phages and evolution	2
Their relevance in basic research and phage-derived tools used in molecular biology	3
Phages in genomics and proteomics	4
1.1.2 Phage systematics	5
1.1.3 A phage’s reproduction cycle	6
1.1.4 <i>Streptococcus pneumoniae</i> – a major human pathogen	8
<i>Streptococcus pneumoniae</i>	8
Genome sequencing and strain comparison	9
1.1.5 Bacteriophages of <i>Pneumococci</i>	10
Host specificity	11
Phage receptors	12
1.1.6 The lytic phages Cp-1 and Dp-1	12
Cp-1	12
Dp-1	14
Their lytic enzymes	15
1.2 Conserved hypothetical proteins: new hints and new puzzles	16
1.2.1 YbeB – a highly conserved ribosome-associated protein	16
1.2.2 A view on the ribosome	17
Translation – what we still do not know	17
Ribosome biogenesis (<i>E. coli</i>)	18
From the genetic code to a protein	19
Comparison of translational systems	20
1.3 Interactome screening with the Yeast Two-Hybrid (Y2H) System	22
1.3.1 The Yeast Two-Hybrid principle	22
1.3.2 Applications	23
1.3.3 Matrix-based Yeast Two-Hybrid Screens (one-on-one)	24
Why matrix-based screens?	24
1.3.4 (Mini-) pool screening strategies	26
1.4 Motivation	28
Major project: Functional proteomics of the lytic bacteriophages Cp-1 and Dp-1 of <i>S. pneumoniae</i>	28
Minor project: Functional analysis of the conserved hypothetical protein YbeB and its interaction with the ribosomal protein L14	29
2 Materials and Methods	31
2.1 Materials	31
2.1.1 Chemicals and enzymes	31
2.1.2 Instruments, consumable materials, and kits	34
2.1.3 Primers	35
2.1.4 Plasmids	37
2.1.5 Strains	38
2.2 Methods	39
2.2.1 DNA/RNA methods and assays	39
2.2.1.1 Primer design	39
High-throughput primer design	39

Standard primer design	39
2.2.1.2 Common protocols for Polymerase Chain Reaction (PCR)	39
Taq DNA polymerase PCR.....	39
PrimeSTAR DNA polymerase PCR	40
2.2.1.3 Gateway® cloning	41
Gateway® PCR.....	41
Gateway® BP reaction	42
Gateway® LR reaction	43
2.2.1.4 Classical cloning by restriction/ligation.....	43
Cloning of pBAD-lacZ-HA	44
Cloning of localization and BiFC constructs	44
2.2.1.5 Directed mutagenesis of L14 by fusion PCRs	45
2.2.1.6 PCR product clean up	46
2.2.1.7 Plasmid isolation and purification	46
2.2.1.8 Agarose gel electrophoresis	47
2.2.1.9 RNA isolation	48
2.2.1.10 Determination of DNA and RNA concentration	48
2.2.1.11 Reverse transcription of mRNA, gene-specific	49
2.2.2 Protein methods and assays	50
2.2.2.1 Protein expression.....	50
2.2.2.2 Purification of His ₆ -tagged proteins.....	50
Protein extraction.....	50
His ₆ -protein isolation (batch protocol).....	50
Dialysis	51
2.2.2.3 Bradford assay	51
2.2.2.4 Polyacrylamide gel electrophoresis (PAGE)	52
Tris-glycine SDS-PAGE.....	52
Tricine-PAGE (gradient gel)	53
2.2.2.5 Coomassie Blue staining of protein gels.....	53
2.2.2.6 Western Blot (semi-dry)	54
2.2.2.7 Protein immunodetection and ECL (enhanced chemiluminescence)	54
2.2.2.8 Membrane stripping.....	55
2.2.2.9 Protein cross-linking with glutaraldehyde	55
2.2.2.10 Uridine kinase (Udk) enzyme assay and thin layer chromatography (TLC)	56
Udk enzyme assay	56
Thin layer chromatography (TLC)	56
2.2.2.11 Protein gel filtration.....	57
2.2.2.12 Pull downs	57
2.2.3 Bacterial methods and assays.....	58
2.2.3.1 Bacterial media	58
LB (lysogeny broth) medium.....	58
MOPS minimal medium	58
2.2.3.2 Preparation of chemocompetent <i>E. coli</i> cells.....	60
CaCl ₂ -method.....	60
TSS-method	60
2.2.3.3 Transformation of chemocompetent cells by heat shock.....	60
2.2.3.4 Selection conditions for plasmid-transformed <i>E. coli</i> strains	60
2.2.3.5 Flip-out of kanamycin cassette from <i>E. coli</i> gene deletion mutant.....	61
2.2.3.6 Determination of OD600	61
2.2.3.7 <i>ΔybeB</i> phenotyping growth curve assays.....	62
2.2.3.8 <i>ΔybeB</i> phenotyping agar plate assays	62
2.2.3.9 Competitive fitness assay of <i>ybeB</i> gene deletion mutant.....	62
2.2.3.10 ONPG reporter assay of <i>ybeB</i> gene deletion mutant	63
2.2.3.11 Rescue assay	63

2.2.3.12	Detection of β -galactosidase reporter protein levels in <i>ΔybeB</i> and wild-type cells	63
2.2.3.13	GFP fluorescence reporter assay of <i>ybeB</i> gene deletion mutant	64
2.2.4	Cell culture and cell culture related assays	64
2.2.4.1	Cultivation.....	64
	Passaging cells	64
	Seeding cells	65
	Freezing and thawing cells.....	65
2.2.4.2	Cell transfection	65
2.2.4.3	Cell preparation for fluorescence microscopy	65
2.2.4.4	Fluorescence microscopy	66
2.2.4.5	Co-localization and BiFC (bi-molecular fluorescence complementation) assay	66
2.2.4.6	LuMPIS	67
2.2.5	Yeast methods and Yeast Two-Hybrid assays	69
2.2.5.1	Yeast media.....	69
	YEPD medium	69
	5 x medium concentrate	69
	Selective plates.....	69
	Dropout mix (-His, -Leu, -Trp).....	69
	Amino acid solutions	69
2.2.5.2	Yeast transformation	70
2.2.5.3	Setup of a screen array	70
2.2.5.4	Bait and prey yeast selection conditions	71
2.2.5.5	Bait self-activation test.....	71
2.2.5.6	Matrix-based Yeast Two-Hybrid screens – one-on-one tests	72
	Preparations.....	72
	Mating procedure	72
	Selection of diploids.....	72
	Interaction selection	73
2.2.5.7	Yeast Two-Hybrid mini-pool screens	73
2.2.5.8	Yeast colony PCR	73
2.2.5.9	Yeast Two-Hybrid retests	74
2.2.6	Bioinformatics and statistics	75
	Multiple alignments	75
	Web logos	75
	ConSurf.....	75
	Protein docking	75
	Local protein and nucleotide sequence blast.....	76
	Data storage.....	76
	Sequence-based domain predictions	76
	Annotation of the Dp-1 genome.....	76
	Dp-1 genome context vs. detected protein-protein interactions.....	77
3	Results	79
3.1	Functional proteomics of the lytic bacteriophages Cp-1 and Dp-1.....	79
3.1.1	Functional annotation of the Dp-1 genome.....	79
	General characteristics of the Dp-1 genome and open reading frame prediction	79
	Homology-based prediction of protein functions.....	81
	Annotations that need some further explanations	82
	tRNAs	86
	Genomic context, prediction of gene modules, and a predicted transcriptional map.....	86
3.1.2	Some remarks on Dp-1 protein systems.....	89
	Queuosine biosynthesis enzymes	89
	DNA replication and recombination systems	91
	Structural components of Dp-1	93

Dp-1 DNA transcription	93
Lytic cluster	94
Other notable gene products	94
3.1.3 Intra-viral protein-protein interaction networks of Cp-1 and Dp-1	95
The binary Dp-1 interactome	95
The binary Cp-1 interactome	98
Comparison of the genomic contexts and the intra-viral PPI networks reveals a modular organization of phage interactomes	99
3.1.4 Protein-protein interactions of Cp-1 and Dp-1 with their host <i>S. pneumoniae</i>	103
Proteome-wide identification of phage-host PPIs by Y2H screens	103
Dp-1 and Cp-1 attack the host interaction network mainly at different positions	106
Comparison of host and phage PPI network properties	107
Comparison of T7 phage genetic host links with Dp-1 and Cp-1 virus-host PPIs	110
Verification of phage-host PPIs by LuMPIS pull down assays	110
3.1.5 Characterization of the Cp-1 lysozyme - host uridine-cytidine kinase interaction	113
The Cpl1 cell wall binding module mediates the interaction with Udk	113
Cpl1 binds to Udk in vitro	115
Cpl1 does not affect the Udk enzyme activity in vitro	116
3.2 Functional analysis of the conserved hypothetical protein YbeB	118
3.2.1 High evolutionary distribution of YbeB homologs by low sequence conservation	118
3.2.2 Systematic examination of conserved docking sites of YbeB on the large ribosomal subunit	124
Interaction partners of YbeB in <i>T. pallidum</i> : which translational components interact with YbeB?	124
Which interaction is conserved?	124
Cross-species interactions indicate a high L14-YbeB interaction co-conservation	127
The YbeB-L14 interaction is conserved in man and plants	128
The human YbeB/L14 orthologs co-localize into mitochondria and interact in vivo	129
Mapping the YbeB binding site on <i>E. coli</i> L14 – YbeB caps conserved and critical L14 amino acid residues	131
3.2.3 YbeB functions as a negative modulator of translation <i>in vivo</i>	133
Absence of YbeB causes a growth defect under competitive conditions	133
The loss of YbeB results in premature protein synthesis in vivo	134
3.2.4 <i>In silico</i> -protein docking of YbeB/L14 reveals that YbeB interferes with 50S-30S assembly on the 3D-level	137
4 Discussion	139
4.1 Intra-viral PPIs of Dp-1 and Cp-1 and interactions with their host <i>S. pneumoniae</i>	139
4.1.1 PPI data quality – the problem with false-negatives and false-positives	139
4.1.2 Virion models: functional protein linkage maps of virion associated proteins of Dp-1 and Cp-1 implicate expected and unexpected proteins to function in virion assembly - a comparative view	142
A Dp-1 virion model	142
A Cp-1 virion model	147
4.1.3 How can a bacteriophage hijack its host? A comparative view	149
Phage proteins can optimize the host cell	149
Do virulent (lytic) bacteriophages kill the host prior to the lytic step?	151
Phages fill up their “genomic gaps” through recruitment of host protein	152
Global reprogramming of the host’s gene expression by attacking regulatory host proteins	154
Specific reprogramming of the host metabolism by protein-binding	159
The physiological meaning of the Cp-1 lysozyme-host uridine-cytidine kinase interaction	160
4.2 Functional and conserved association of the YbeB protein family with the large ribosomal subunit and protein translation	161

The evolutionary distribution of YbeB orthologs link it to the eubacterial origin.....	161
L14 recruits YbeB to an important 50S-30S interface region on the large ribosomal subunit	161
YbeB – physiological function.....	163
5 Conclusions and outlook.....	166
Cp-1 and Dp-1.....	166
YbeB.....	167
References.....	168
Supplementary information.....	182
Supplementary figures.....	182
Supplementary tables.....	186
List of abbreviations.....	193
List of figures.....	194
List of tables.....	195

1 Introduction

1.1 Bacteriophages – a brief overview

Bacteriophages (phages) are estimated to be the most frequent biological entities in the biosphere in terms of their particle number and their diversity. Bacteriophages (from Ancient Greek *phagein* means “to eat” bacteria) are viruses that specifically infect and kill bacteria. Like all viruses, phages are metabolically inert in their extracellular form and can reproduce themselves only by taking advantage of a host cell. Because of their inability to reproduce by themselves they are not defined as living organisms although they function as biological entities. The first clear description of the phage phenomenon, namely the property of forming plaques, was described by Félix d’Herelle in 1917, a French-Canadian microbiologist working at the Pasteur Institute in Paris (d’Herelle, 1917). D’Herelle observed what he called an “invisible microbe” that was present in the bacteria-free filtrates of stool samples from dysentery patients and surmised the recovery of patients was due to development of phages.

1.1.1 Relevance of phages – what they are (not) good for

Phages in medicine

It was quickly realized that phages had the potential to kill bacteria that are causatives of many human, animal, and plant diseases without having adverse reactions. In early phage therapy trials, d’Herelle applied phage lysates to treat fowl typhoid (*Salmonella gallinarum*) and diarrhea caused by *Shigella dysenteriae* in animal tests (d’Herelle, 1926a). Moreover, he was able to cure water buffalos in Indochina successfully from bovine hemorrhagic septicemia (*Pasteurella multocida*) (d’Herelle, 1926b). With evidence for therapeutic effects d’Herelle extended the trials by self-administration. Phage therapy was also under investigation to treat cholera in India in the 1920s and 1930s (d’Herelle et al., 1929; Summers, 1993). These studies were extended later, sponsored by the World Health Organization (WHO) (Marcuk et al., 1971; Monsur et al., 1970). An institute for the study and production of phages was founded in the mid 1930s in the Soviet Republic of Georgia (State Serum and Vaccine Institute in Tbilisi) and remains active today. During the Second World War Soviets and Germans used medical kits based on phage therapy to treat war-wounded. In the Western Hemisphere phage therapy has been limited with exiguous commercialization. Its use all but ceased in the 1940s with the emergence of penicillin and other antibiotics. Disadvantages like low stability of phage preparations, limited action because of high host specificities (in contrast to a broad mode of action of antibiotics), development of bacterial resistant strains, as well as immunological answers to phage therapy led to the stagnation in this field of research. However, over the last two decades there has

been renewed interest in phage therapy, largely due to the growing resistance of many strains of bacteria to existing antibiotics.

In parallel, scientists identified and characterized the factors that are responsible to kill the host cell, so-called lytic enzymes, which are encoded by the phage genomes. Pre-medical studies revealed that such enzybiotics can kill pathogenic bacteria very efficiently in purified form and might be helpful in the future to cure bacterial infections as an alternative treatment strategy to antibiotics (Loeffler et al., 2001).

Phages in ecology

Bacteriophages are the most abundant biological entities on earth. Direct counts reveal that there are up to 10^7 phage particles per ml ocean and lake water and up 10^9 per gram of sediment and top soil (Ashelford et al., 2003; Berg et al., 1889). There are approximately five to ten phage particles per prokaryotic cell (Wommack and Colwell, 2000). Whitman and colleagues estimated 10^{30} prokaryotes on earth and 10^{31} free phages (Whitman et al., 1998). They keep the bacterial population in balance. It is suggested that marine phages kill about half of the bacterial population every day (Fuhrman and Noble, 1995). The dissolved bacterial matter is consumed by heterotrophic bacteria which can be lysed again by phages. Thus, phages are responsible for the production of CO_2 . In the world's oceans it is estimated that phages are accountable for the release of approx. 6 to 26% of photosynthetically fixed organic carbon (Wilhelm and Suttle, 1999) and thereby contribute significantly to the carbon cycle as well as to the balance of ecological systems.

Their relevance in food production and industrial processes

In commercial food production bacteriophages are not welcome: as soon as phages emerge in sauerkraut, cheese, and yogurt production a single phage particle is able to kill the lactic acid bacteria and thus the whole fermentation process (Lu et al., 2003). Moreover, it has been reported that phages can interfere with the industrial acetone butanol fermentation process (Jones et al., 2000). To deal with this problem the industry is interested in isolation and usage of phage-resistant bacterial strains. For instance, for Cheddar cheese making bacteriophage-insensitive *Streptococcus cremoris* strains were isolated (Thunell and Sandine, 1981).

Phages and evolution

Unlike eukaryotes, which evolve principally through modification of their genetic material, bacteria obtained a part of their genetic diversity through annexing of DNA sequences from distantly related organisms. Comparison of completely sequenced bacterial genomes revealed a significant amount of laterally transferred DNA. For instance, *E. coli* K12 contains about 16% of foreign DNA (Ochman et al., 2000). Lateral transfer occurs by uptake of naked DNA or by bacterial conjugation. However, bacteriophages are potent transducers since they cannot only deliver genetic material to a donor cell but also provide DNA integration systems that fuse the donor DNA into the recipient genome. Here

phage-encoded integrases, recombinase, and homologous recombination systems play a beneficial role (Ochman et al., 2000).

Their relevance in basic research and phage-derived tools used in molecular biology

Bacteriophages served as model systems in the early stage of molecular biology because they represented simple and convenient biological systems. Studies on phages revealed the mechanisms of transcription and transcriptional regulation, DNA-related processes, virion reproduction, infection, and assembly. Moreover, it was of general interest to understand mechanisms how phages can manipulate the host cell and how temperate phages are able to integrate their DNA molecule into the host genome. However, with the rise of modern techniques and more general interest in eukaryotic biology, research on bacteriophages went astray. Nevertheless, great elementary discoveries were made and modern molecular biology still benefits from these findings. Phages are still in focus in system structural biology. Completely novel mechanisms are still discovered in phage research. The following tables summarize the most important model phages as well as common molecular tools with phage origin:

Tab. 1 Model phages

ORFs: represents the number of protein-coding open reading frames. Information was taken from NCBI webpage.

Model phage	Family	Host	Lifestyle	Genome (size)	ORF
<i>T7</i>	<i>Podoviridae</i>	<i>E. coli</i>	lytic	dsDNA (39.9 kbp)	60
λ	<i>Siphoviridae</i>	<i>E. coli</i>	temperate	dsDNA (48.5 kbp)	73
<i>T4</i>	<i>Myoviridae</i>	<i>E. coli</i>	lytic	dsDNA (168.9 kbp)	278
$\phi 29$	<i>Podoviridae</i>	<i>B. subtilis</i>	lytic	dsDNA (19.3 kbp)	27
<i>M13</i>	<i>Inoviridae</i>	<i>E. coli</i>	non-lytic	ssDNA (6.4 kb)	9
<i>SPP1</i>	<i>Siphoviridae</i>	<i>B. subtilis</i>	lytic	dsDNA (44,0 kb)	101

Tab. 2 Phage-derived tools and methods

Tool/method	Description	Reference
λ -red system	Application of λ -red homologous recombination system ($\gamma, \beta, \text{exo}$) for construction of gene knock-outs of <i>E. coli</i> genes.	(Datsenko and Wanner, 2000; Murphy, 1998)
λ -integrase	Site-specific recombination system with λ -recombinase finds commercial application for ORFeome construction with the Gateway® System (Invitrogen).	(Landy, 1989)
T7 RNA polymerase	Usage for standard protein expression in <i>E. coli</i> BL21. High level expression of genes.	(Studier and Moffatt, 1986; Studier et al., 1990)
T4 DNA ligase	Commercially available enzyme for cloning reactions by DNA end strand ligation reaction/DNA ring closure (blunt or sticky end ligation).	(Weiss and Richardson, 1967)
Phage display	Exposure of a library-fused coat proteins on virion surface, fishing for binding epitopes followed by liquid chromatography. Application in antibody-antigen, protein-protein, or protein-DNA interaction screening.	(McCafferty et al., 1990; Rebar and Pabo, 1994)
λ DNA standards	Commercially available DNA ladders of restricted λ -DNA.	-
M13 DNA	M13 genomic ssDNA molecule as shot gun sequencing vector for Sanger DNA sequencing method.	(Messing et al., 1981; Sanger and Coulson, 1975)

Phages in genomics and proteomics

Pioneer sequencing studies of whole genome DNA sequences were already published in 1977 for *E. coli* phage ϕ X174 by F. Sanger, 1980 for M13, and 1983 for T7 (Dunn and Studier, 1983; Sanger et al., 1977; van Wezenbeek et al., 1980). Since that time phage genome sequencing projects increased exponentially (Fig. 1). By 2009 577 completely sequenced phage genomes had been submitted to NCBI/GenBank. Although conserved genome features like protein encoding ORFs can be and are annotated by homology-based functional predictions, there is still no comprehensive phage sequence database available online, such as TIGR CMR for bacteria (Davidsen et al., 2010).

In 2002 the age of viral metagenomics started with sequencing of marine (viral) communities (Breitbart et al., 2002). Several studies have been published, e.g., the viral/microbial metagenome from human feces (Breitbart et al., 2003). Comparison of different metagenome datasets revealed that around 60 to 90% of these sequences have no significant sequence similarities (Edwards and Rohwer, 2005). It is assumed that most of these unknown sequences are attributed to bacteriophage genomes (Daubin and Ochman, 2004) underlining their immense diversity. While phage DNA sequences are available in large numbers, only two comprehensive datasets have been published that focused on functional aspects: Bartel and colleagues analyzed the T7 interactome by systematic Yeast Two-Hybrid screens (Bartel et al., 1996) and Qimron and others tested genome-wide effects of *E. coli* ORFs on T7 reproduction efficiency (Qimron et al., 2006) (Fig. 1). However, phage genome sequences turned out to be less informative than bacterial genomes. To find functions for the large fraction of orphan proteins, comprehensive and genome-wide experiments are urgently needed, e.g., transcriptomic and proteomic analyses of the phages and their hosts.

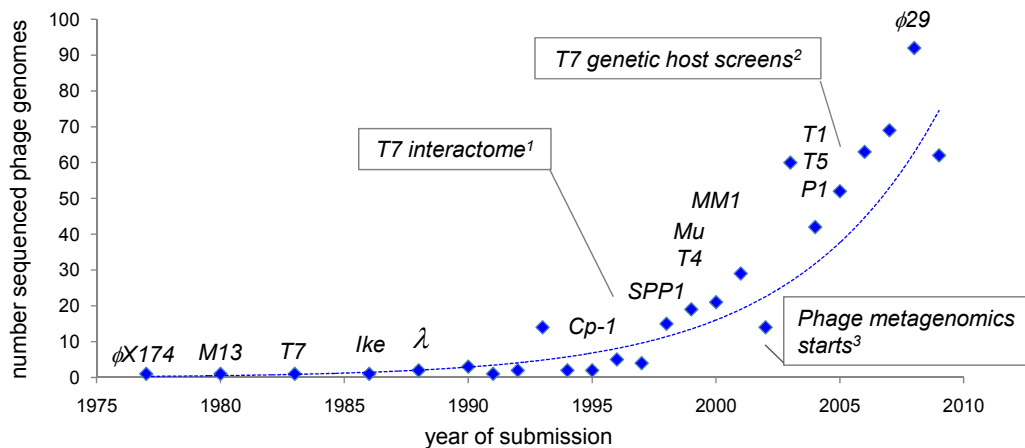


Fig. 1 Comprehensive, functional phage studies are underrepresented compared to phage genome sequencing projects

The graph represents the number of submitted whole genome phage sequences to NCBI from 1977 to 2009. Submission dates of some representative bacteriophages are labeled by the phage name. Consider that submission and publication dates might differ. Publication year of comprehensive phage studies are given ¹(Bartel et al., 1996) and ²(Qimron et al., 2006) and beginning of phage metagenomics era with ³(Breitbart et al., 2002) (metagenome sequences are not represented in the graph).

1.1.2 Phage systematics

Bradley initiated the systematic classification of bacteriophages in 1969 by morphological criteria (Bradley, 1969). He grouped phages based on capsid/capsomere symmetry (e.g., isahedragonal, octahedrogonal) and the kind and length of the tails (contractile/non-contractile, short/long), but also on the type of the genome into several basic groups (Tab. 3). Most of the phage families contain a dsDNA molecule as genetic material. However, phage genomes are known to consist of ssDNA, ssRNA, or dsRNA and the nucleic acid molecules occur in circular or linear forms. Moreover, some rare bacteriophage groups contain an internal lipid membrane (*Tectiviridae* and *Coricoviridae*). Other groups (*Lipothrixviridae*, *Plasmaviridae*, and *Cystoviridae*) are characterized by a lipid outer layer and are thus commonly described as enveloped phages.

Many phage groups have been classified and modern classification by sequence comparisons give a detailed view on phage relationships and evolution beyond morphological aspects (Fig. 2).

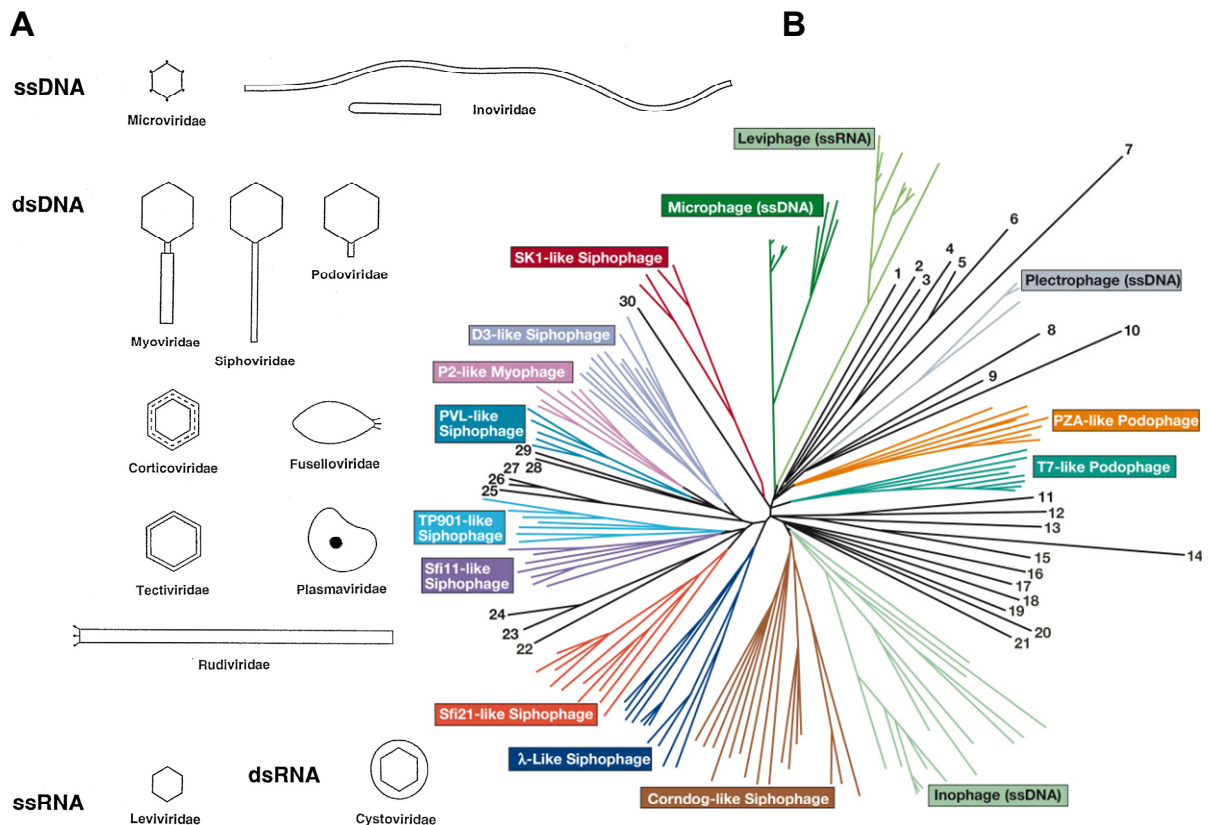


Fig. 2 Phage systematics

(A) Bradley Classification of bacteriophage groups based on general, different morphotypes (Bradley, 1969). Kind of genetic material is indicated. Image taken from www.thebacteriophages.org. (B) A modern phage proteomic tree (Edwards and Rohwer, 2005). The tree includes 167 phage genomes. Phages in black cannot be classified into any clade. Image taken from (Edwards and Rohwer, 2005).

Tab. 3 Bradley Classification of bacteriophages

The table here gives the six classified phage groups after Bradley (Bradley, 1969). Modern designations of virus families and phage examples were added.

Family	Description	Nucleic acid type	Example
<i>Myoviridae</i>	Contractile tail	dsDNA	<i>Enterobacteria phage T4</i>
<i>Siphoviridae</i>	Long non-contractile tail	dsDNA	<i>Enterobacteria phage λ</i>
<i>Podoviridae</i>	Short non-contractile tail	dsDNA	<i>Enterobacteria phage T7</i> <i>Bacillus phage ϕ29</i>
<i>Microviridae</i>	Tailless, large capsomer	ssDNA	<i>Enterobacteria phage ϕX174</i>
<i>Leviviridae</i>	Tailless, small capsomer	ssRNA	<i>Enterobacteria phage MS2</i>
<i>Inoviridae</i>	Filamentous	ssDNA	<i>Enterobacteria phage M13</i>

1.1.3 A phage's reproduction cycle

Beside their systematic classification, bacteriophages can have different lifestyles. However, the lifestyle is not a systematic criterion since phages from different groups can have similar lifestyles. Both, lytic phages as well as temperate phages, first enter the bacterial host cell (infection). For instance, λ phage (*Siphoviridae*) penetrates the *E. coli* cell with the gpJ protein in its tail tip that binds to LamB maltoporin (maltose porin) of the outer membrane (Randall-Hazelbauer and Schwartz, 1973). DNA injection is accomplished by only partially understood mechanisms: forces like capsid pressure, temperature, DNA length, and non-specific DNA-binding proteins such as HU are known parameters that control DNA injection (Lof et al., 2007). By contrast, T4 phage (*Myoviridae*) penetrates the host cell by binding to LPS and porin OmpC (Dawes, 1975; Yu and Mizushima, 1982) and injects its DNA by a cascade that actively triggers the contraction of the tail and thus the genome injection (Rossmann et al., 2004). *Bacillus subtilis* phage SPP1 (*Siphoviridae*) was reported to be absorbed by membrane protein YueB and by poly(glycerolphosphate)teichoic acid of the cell wall (Baptista et al., 2008; Sao-Jose et al., 2004). *B. subtilis* phage ϕ 29 (*Podoviridae*) phage antireceptor protein was shown to consist of a glycosidase domain that might help to degrade the peptidoglycan layer during the penetration as an example that degradation events can take place during the infection process (DiMauro et al., 2007). Next, they start to amplify their genome (replication) and produce large amounts of structural proteins (synthesis). After the virion particles have assembled (assembly) they leave the host cell in a final step (lysis) and the daughter virions can start with a new infection cycle (Fig. 3). Lysis factors are lytic enzymes, commonly called lysins or cell wall hydrolases. They digest the peptidoglycan layer which triggers the burst of the host cell. The lysins can enter the periplasmic or outer space by holins. These are phage-encoded transmembrane proteins that assemble in the (inner) membrane and shuttle the lysozymes through the membranes (Wang et al., 2000).

In contrast to lytic phages, temperate phages can also enter a lysogenic pathway. That is, after infecting the host cell, they are able to integrate their genome into the host genome. Integrated phage genomes are designated as prophages. For many host generations the virus genome is silenced and

mostly prophage gene expression is blocked by phage-encoded repressors. This has the advantage that the phage genomes are passively copied and spread within a bacterial strain. Notably, beneficial transduction effects of prophage-encoded toxins on host pathogenicity were reported by lateral transfer (Faruque and Mekalanos, 2003; Wagner et al., 1999; Wagner et al., 2002; Wagner and Waldor, 2002). Another possible advantage of the lysogenic state arose from ecological studies of phage populations. It was shown that a higher number of pro-phages are detectable over the cool season (Cochran and Paul, 1998). Switching into the lysogenic pathway might represent a “survival” strategy of temperate phages in regions that undergo seasonal changes. Characteristics for temperate phages are self-encoded recombination enzymes (e.g., integrases, transposases for site-specific DNA recombination or high-efficient homologous recombination systems) that mediate the phage genome integration into the host DNA and are absent in genomes of lytic phages. The lysogenic pathway of *Enterobacteria* phage λ has been intensively studied. The decision between the lytic and lysogenic pathway is regulated by six λ -encoded transcriptional regulators whose actions result in defined circuits that trigger the pathway fate by regulating the gene expression or suppression of corresponding λ -genes (Herskowitz and Hagen, 1980).

The pro-phage is able to reenter the lytic pathway. Factors that induce λ pro-phage typically cause DNA damages. For instance, potent prophage inducers are ultraviolet light, x- or γ -irradiation, or chemical compounds like mitomycin C (Kirby et al., 1967; Roberts and Roberts, 1975; Shinagawa et al., 1977). In the case of λ phage the λ cI and host LexA repressors are proteolytically cleaved by host RecA due to the host’s SOS response. This triggers the expression of λ -genes responsible for reentering the lytic pathway. λ -DNA is excised and circulates under the influence of the λ recombinase (Day, 1977; Little, 1984; Roberts et al., 1978).

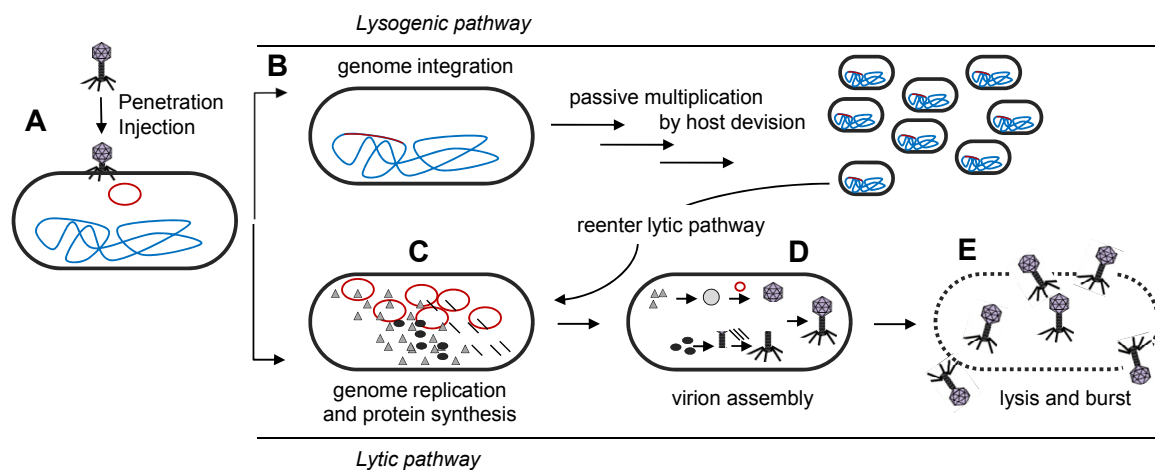


Fig. 3 Scheme of a lytic and temperate phage’s reproduction cycle

(A) Phage penetrates host cell and injects its DNA (in red). (B) Integration of phage DNA into host genome (in blue) in the lysogenic pathway. (C) In the lytic pathway the phage genome is replicated and structural proteins are synthesized. (D) Stepwise self-assembly of daughter virions (e.g., prohead assembly, DNA packaging, tail assembly, head-tail assembly). (E) Cell burst: the final lytic step lyses the host cell and daughter virions are released. Figure was drawn after general information from textbooks.

A third lifestyle is represented by filamentous phages. They are not lethal to the host cell and thus have a non-lytic lifestyle. They infect the host cell (in case of filamentous M13 phage) via the host F-pili receptor TolA (Sun and Webster, 1986). Virion assembly occurs at the inner membrane. However, the non-lytic process of DNA uptake and virion release is not fully understood.

Lifestyles of (bacterial) viruses share all the general steps as illustrated in Fig. 3. However, phages from different groups evolved different penetration and release strategies. Adaption of their replication systems is another evolutionary strategy, e.g., development of reverse transcriptase for RNA phages. Bacteriophages developed as manifold biological mechanisms since they had to find ways to adapt to versatile host systems and ecological niches.

1.1.4 *Streptococcus pneumoniae* – a major human pathogen

Streptococcus pneumoniae

S. pneumoniae (formerly also called *Pneumococcus* and *Diplococcus*) is a gram-positive, alpha-hemolytic, aerotolerant but anaerobe eubacterium. *S. pneumoniae* does not form spores and is non-motile. The cells are round to ovoid in shape and appear in pairs (diplococci) but also form single cells or cell chains (Fig. 1). They are between 0.5 to 1.25 μm in size. They are mesophilic and grow optimally between 30 and 35°C.

Louis Pasteur and George M. Sternberg discovered *S. pneumoniae* independently in 1881. Frederick Griffith demonstrated in 1928 that a harmless *Pneumococcus* strain can be transformed into a pathogenic one. He pre-incubated a heat-inactivated encapsulated strain with a vital non-encapsulated strain that turned out to become lethal to mice since it developed a capsule (Griffith, 1928). In 1944 Oswald Avery and co-workers demonstrated that the transforming factor in Griffith's experiment was indeed DNA, not protein as it was widely believed at the time (Avery et al., 1944). This work marked the birth of the molecular era of genetics.

S. pneumoniae is pathogenic to humans and causes several diseases including pneumoniae, otitis media, meningitis, bacteremia, and sinusitis. *S. pneumoniae* is one of the deadliest human pathogens. It is estimated to be responsible for approx. 14.5 millions infections and causes 826,000 deaths per year of children (O'Brien et al., 2009).

Although infections can be treated by classical antibiotics, resistant strains appeared frequently over the last years and became a clinical problem. Penicillin and other β -lactam, macrolide, and fluoroquinolone resistant strains have been reported (Albrich et al., 2004; Bronzwaer et al., 2002). Thus, there is an urgent need to develop novel strategies to control pneumococcal diseases.

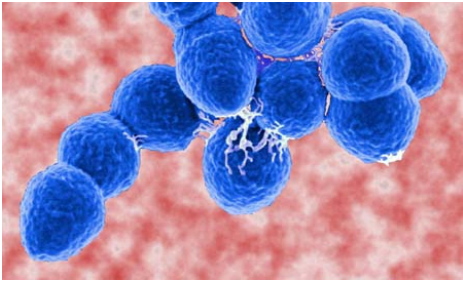


Fig. 4 *Streptococcus pneumoniae* cell chains
Image taken from <http://emssolutionsinc.wordpress.com>.

S. pneumoniae is a natural part of the upper human respiratory tract but it can become pathogenic under certain conditions. The major virulence factor is its capsule (also called murein sacculus) that prevents phagocytosis by the host's immune cells. It can reach a thickness of approx. 400 nm in contrast to the other peripheral compartments, the membrane (9 nm), peptidoglycan layer (20 nm), and cell wall polysaccharides and proteins (20 to 30 nm). It is composed of a complex mixture of polysaccharides. More than 90 different serotypes are known. The capsule consists of high-molecular-weight polymers made up of units of repeating oligosaccharides, which can contain between two and eight monosaccharides. Many serotypes possess acidic components (like D-glucuronic acid or phosphate groups), ribitol, or arabinitol. In other serotypes phosphorylcholine is a part of the capsular polysaccharide. Another important structure is the lipoteichoic acid that is attached to the membrane (teichoic acid covalently linked to lipids) (AlonsoDeVelasco et al., 1995). Some other virulence factors are the potential to form pili which adhere via adhesins to the host epithelium (Nelson et al., 2007), the production of H_2O_2 that damages the lunge epithelium (Duane et al., 1993), and the release of toxic pneumolysin that binds to cholesterol containing host cell membranes and lyses them (Lock et al., 1992).

Genome sequencing and strain comparison

Currently, the TIGR CMR database contains eight complete *S. pneumoniae* genome sequences from different strains. The genome of highly virulent TIGR4 strain (belonging to serotype 4 group) was sequenced by Tettelin and colleagues and that of the non-encapsulated laboratory strain R6 by Hoskins et al. (Hoskins et al., 2001; Tettelin et al., 2001). The genomes are 2.16 and 2.04 Mbp in length and encode for 2,236 and 2,043 putative proteins, respectively for TIGR4 and R6. A characteristic of the genomes is the low GC content (39.7% and 40%, respectively) and thus the abundance of several amino acids including isoleucine, glutamic acid, valine, and lysine is high (Fraser et al., 2000). For TIGR4 64% of its ORFs were assigned to have a biological role based on homology comparisons. 5% of the genome sequence contains insertions that could contribute to genomic rearrangements by uptake of foreign DNA. Moreover, the TIGR4 strain does not contain any pro-phage in its genomic sequence. *S. pneumoniae* has the widest substrate utilization range for sugars and substituted nitrogen compounds of the three completed genomes of near-commensal residents of the human upper respiratory tract (*H. influenzae*, *Neisseria meningitidis*, and *S. pneumoniae*). It contains a large set of transporters that indicate a wide spectrum of metabolite uptake systems. Genome features suggest that

S. pneumoniae possesses pathways for catabolism by the pentose phosphate pathway and is capable to metabolize many sugars like cellobiose, glucose, galactose, fructose, and others (consider Fig. 48 in the results section). Features like putative extracellular sugar hydrolases indicate the potential to digest and mobilize sugar and nitrogen compounds from the extracellular host matrix that could be used for biosynthesis of the murein sacculus. Moreover, components of the respiratory chain and the citrate cycle are completely absent and correlate with the bacterium's property to be an anaerobe organism (Tettelin et al., 2001).

A comparison of the TIGR4 and R6 genome revealed that TIGR4 contains a higher number of insertion sequences and transposable elements. This indicates a higher genetic flexibility and mutation frequency in TIGR4. However, major cellular systems and features of TIGR4 that are notably different include the genes involved in amino acid biosynthesis, cell envelope, cellular processes, central intermediary metabolism, energy metabolism, and also others. These differences might explain the uniqueness of both strains. For instance, pathogenicity due to the murein sacculus of TIGR4 is caused by 12 proteins that are involved in biosynthesis of capsular polysaccharides. In contrast, R6 encodes only for one such an enzyme (Jothi et al., 2008).

1.1.5 Bacteriophages of *Pneumococci*

Pneumococcal (*S. pneumoniae*) phages include temperate and lytic phages with dsDNA genomes and have a great variability in terms of their morphology (Garcia et al., 1997) (Tab. 4). ssDNA, ssRNA, and dsRNA viruses have not been reported yet for *S. pneumoniae*.

Temperate phage EJ-1 was isolated after pro-phage induction by mitomycin treatment of *S. pneumoniae* strain 101/87 (Diaz et al., 1992). Bernheimer isolated 49 (partially encapsulated) *S. pneumoniae* strains from 110 pediatric patients that carried several mitomycin inducible pro-phages. They were named HB-1 to HB-6 (Bernheimer, 1979).

Recently in 2009, Romero and colleagues detected by PCR a great number of *S. pneumoniae* pro-phages and could induce 48 of them (Romero et al., 2009b). Moreover, they sequenced the genomes of 10 other temperate *S. pneumoniae* phages (Romero et al., 2009a). These numbers underline that still many novel bacteriophages can be found in *S. pneumoniae* isolates although the latter ones are all temperate.

Regarding phage therapy, lytic phages are of general interest since they cannot integrate their genome into the host genome and thus are lethal to the host cell. But only two lytic phages (Cp-1 and Dp-1) have been isolated and were more or less intensively characterized. Thus, it is of general interest to isolate novel lytic *S. pneumoniae* viruses. Recent attempts resulted in two novel lytic *Pneumoniae* phages from human samples taken in Saudi Arabia. They have been named ϕ L6B1 and ϕ L6B2 and belong to the *Siphoviridae* and *Myoviridae* family, respectively (Almaghrabi and Clokie, 2010) (data unpublished).

Host specificity

Host specificities are only poorly understood, i.e. it remains unclear which *Streptococcus* strains can be infected by a certain phage.

EJ-1 failed to infect other species of *Streptococcus* (Diaz et al., 1992). It was shown that Cp-1 and Cp-7 are the only *Pneumococcus* phages known so far to infect and replicate in *S. oralis* which shares the same niche with *S. pneumoniae*. In contrast, Dp-1 failed to propagate in *S. oralis* (Ronda et al., 1989). All other phages listed in Tab. 4 are only known to infect *S. pneumoniae*.

Infection assays on four *S. pneumoniae* reference strains with the temperate HB phages revealed no clear pattern of host specificities although differences were found (Bernheimer, 1979). The lytic phages ϕ L6B1 and ϕ L6B2 can infect the capsulated strain D39 but not non-capsulated ones and thus are promising candidates for phage therapy (Almaghrabi and Clokie, 2010) (unpublished data). In contrast, Cp-1 and Dp-1 can be propagated in non-capsulated R6 but fail to infect the capsulated TIGR4 strain (unpublished data, personal communication with Mourad Sabri, Université Laval, Québec).

Tab. 4 Formerly isolated *Pneumococcal* bacteriophages and some properties

If a phage genome was sequenced, the number of corresponding protein-coding ORFs is given. (-) no or (?) no clear information is available in literature.

Family	Name	Lifestyle	Genome (size/kbp)	Number ORFs	References
<i>Siphoviridae</i>	Dp-1	Lytic	dsDNA (56.5)	72	This work, (Lopez et al., 1977; McDonnell et al., 1975) (Garcia et al., 1979)
<i>Siphoviridae</i>	Dp-4	-	dsDNA	-	(Garcia et al., 1979)
<i>Podoviridae</i>	Cp-1	Lytic	dsDNA (19.3)	28	(Martin et al., 1996b; Ronda et al., 1989; Ronda et al., 1981) (Lopez et al., 1984)
<i>Podoviridae?</i>	Cp-5,8,9	-	dsDNA	-	(Lopez et al., 1984)
<i>Podoviridae</i>	Cp-7	-	dsDNA (ca. 17)	-	(Lopez et al., 1984; Ronda et al., 1989)
-	w1,2,4,5,6,7,9	-	-	-	(Tiraby et al., 1975)
<i>Siphoviridae</i>	w3, w8	Temperate (?)	dsDNA	-	(Porter and Guild, 1976; Tiraby et al., 1975)
-	HB-1,2,4,5,6	temperate	-	-	(Bernheimer, 1979)
<i>Siphoviridae</i>	HB-3	temperate	dsDNA (40)	-	(Bernheimer, 1979; Garcia et al., 1997; Romero et al., 1990) (Romero et al., 1990)
<i>Siphoviridae</i>	HB-623 HB-746	temperate	dsDNA	-	(Romero et al., 1990)
<i>Siphoviridae</i>	MM1	temperate	dsDNA (40.3)	53	(Gindreau et al., 2000; Obregon et al., 2003)
<i>Myoviridae</i>	EJ-1	temperate	dsDNA (42.9)	73	(Diaz et al., 1992; Romero et al., 2004)

Phage receptors

In contrast to the vast majority of phages that use mostly outer membrane proteins like porins as main receptors and other co-receptors (some examples are given under 1.1.3) the situation of at least one *Pneumococcus* phage is different. For Dp-1 it was shown that the main receptor is the aminoalcohol choline which is a component of teichoic acid of the cell wall. When substituted by ethanolamine Dp-1 adsorption is disrupted (Lopez et al., 1982). Choline containing phage adsorption sites were found at the equatorial cell zone and termini of growing *Pneumococcus* chains (Lopez et al., 1982). However, there is no information on the receptors of other *S. pneumoniae* phages and no protein receptors have been described in literature so far.

1.1.6 The lytic phages Cp-1 and Dp-1

Cp-1

Cp-1 is the best characterized *Pneumococcus* phage. It was isolated in 1981 (Ronda et al., 1981). “Cp” stands for “complutense”, the name which was given to the Spanish people the phage was isolated from. Its virion shows an irregular hexagonal structure with a short, non-contractile tail of 20 nm (Fig. 5). It belongs to the *Podoviridae* family. SDS PAGE analysis revealed that its virion is composed of at least nine structural proteins. It consists of a linear dsDNA molecule (Ronda et al., 1981).

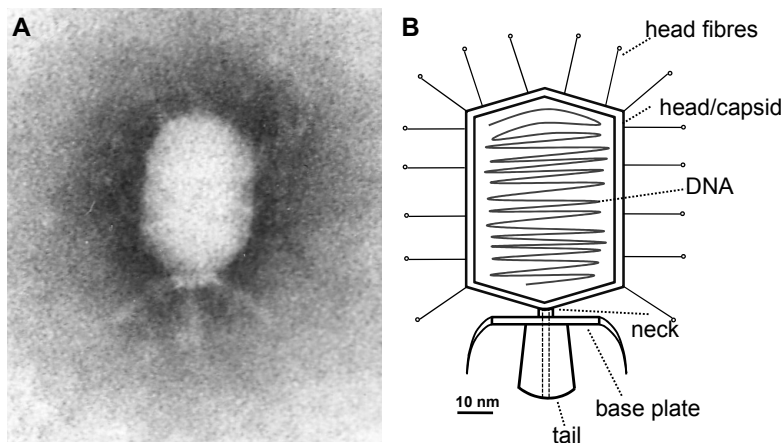


Fig. 5 Cp-1 virion

(A) Micrograph of a mature Cp-1 virion. Image was taken from <http://www.ictvdb.rothamsted.ac.uk>.

(B) Schematic virion structure drawn after (Ronda et al., 1981).

Replication of the linear dsDNA molecule is managed by a protein-primed mechanism which is also found in $\phi 29$ *Bacillus* phage. A terminal protein (TP) is covalently linked to a 5'-dAMP by the hydroxyl group of a threonine residue. The TP-dAMP complex binds to the 3'-ends of the DNA molecule that contains 238 bp long inverted terminal repeats. The genome termini are made of three sequential thymidine residues. TP-dAMP binds to the third thymidine residue (initiation) and jumps or slides back to the first thymidine residue to maintain the length of the daughter strands. Cp-1 DNA polymerase recognizes specifically the DNA-bound TP-dAMP as priming complex followed by linear extension of the daughter strand (Martin et al., 1996a). Since replication of the Cp-1 genome is

INTRODUCTION

dependent of the coupled terminal protein, reconstitution of Cp-1 particles from naked DNA is difficult.

The Cp-1 genome sequence was published in 1996 (Martin et al., 1996b). It is 19.3 kb in length and encodes for putative 28 proteins (Fig. 6). Many of them share high sequence similarities with proteins from *B. subtilis* ϕ 29 phage revealing a close relationship between these two podoviruses. Furthermore, a gene of a putative packaging RNA (pRNA) was identified. It shares high sequence similarity with ϕ 29 pRNA and thus likely functions in DNA packaging. mRNA primer extension analysis revealed a transcriptional map of Cp-1. Five early and nine late promoters were found. After host infection four early and twelve late transcripts were detected (Martin et al., 1996b).

A Cp-1-encoded endoprotease (gp13) was characterized that cleaves off the 48 N-terminal amino acid residues of the major head protein (Martin et al., 1998b). Its lysis system consists typically of a cell wall hydrolase (Cpl1) and a membrane protein (holin, Cph1) (Martin et al., 1998a). An updated annotation list of all Cp-1 ORFs is shown in Tab. 5. Here 46% of all proteins can be functionally annotated by homology predictions or based on experimental evidences. Although there is a close relationship to the well characterized phage ϕ 29, the majority of putative gene products are of unknown function.

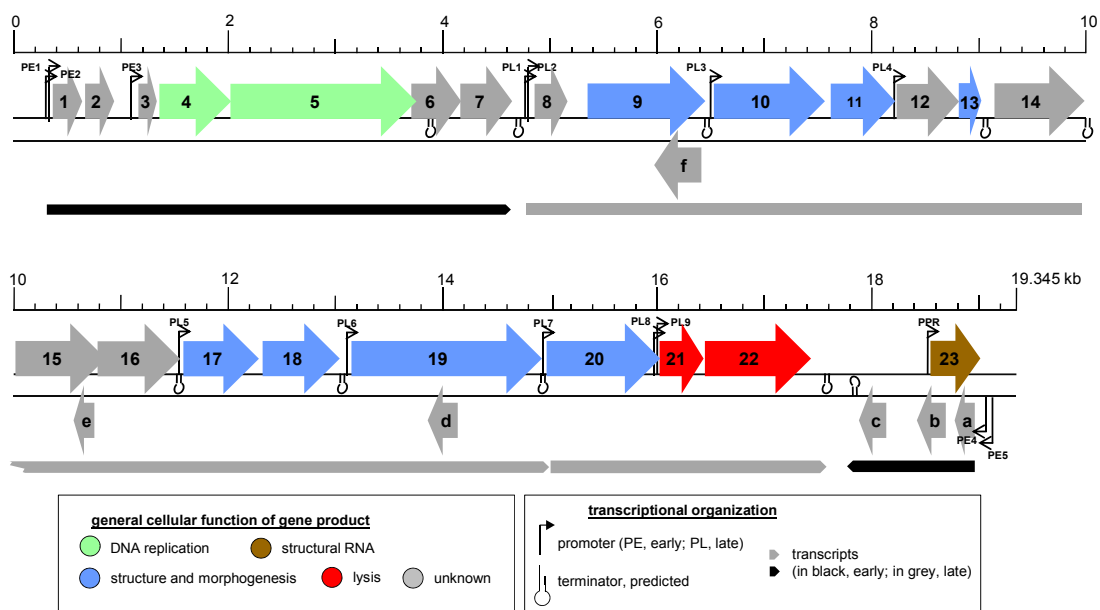


Fig. 6 Genome map of Cp-1 and transcriptional organization

Descriptions of the annotated genes are given in Tab. 5. Explanations for map features are given in the legend (structural RNA= packaging RNA). Only the longest transcripts that were identified by Martin et al. by primer extension analysis are given as arrows below the ORFs. After (Martin et al., 1996b).

Tab. 5 An updated annotation list of Cp-1 gene products

This updated annotation list was compiled after (Martin et al., 1996b). Protein information was integrated from later publications. Hypothetical proteins without experimental or homology-based information are not shown.

ORF id	Annotation	Comment	Reference
4	Terminal protein (TP)	DNA replication by protein priming mechanism	(Martin et al., 1996b)
5	DNA polymerase	DNA polymerase A	(Martin et al., 1996b)
8	Possible scaffolding protein	Prohead formation	(Martin et al., 1996b)
9	Major head protein	Structural protein, head	(Martin et al., 1996b)
10	Connector protein	Structural protein, neck	(Martin et al., 1996b)
11	Lower collar protein	Structural protein, collar	(Martin et al., 1996b)
13	Endoprotease of major head protein	Virion maturation factor	(Martin et al., 1998b)
17	Tail protein N	Homology with N-terminal part of ϕ 29 tail protein	(Martin et al., 1996b)
18	Antireceptor protein, putative	Structural protein	(Martin et al., 1996b)
19	Tail protein C	Homology with C-terminal part of ϕ 29 tail protein	(Martin et al., 1996b)
20	Encapsidation protein (terminase)	DNA packaging	(Martin et al., 1996b)
21	Holin (Cph1)	Lysis	(Martin et al., 1996b)
22	Lysozyme (Cpl1)	Lysis	(Martin et al., 1996b)
23	Packaging RNA (pRNA), putative	DNA packaging	(Martin et al., 1996b)

Dp-1

Dp-1 was the first isolated *Pneumococcus* phage (McDonnell et al., 1975). It belongs to the *Siphoviridae* and is lytic. Its virion consists of a polyhedral head and a long non-contractile tail. A base plate and tail fibers are absent. The head is approx. 67 nm in diameter. An SDS-PAGE analysis of virion particles revealed three major structural proteins and some additional, less-prominent proteins. Furthermore, it was reported that *Dp-1* virions contain a double-layered membrane coat around the head. This is a unique characteristic of *Dp-1* and was not reported for any other siphophage. The membrane lipid composition corresponds to that of the host. Thus, it is assumed to be of host origin. In total, *Dp-1* particles contain 8.5% lipids of their dry weight. Phospholipids, glycolipid A and B, and neutral lipids were identified (Lopez et al., 1977).

A 4.7 kb fragment of *Dp-1* genome was sequenced and published including the genes of the lysis cassette (Sheehan et al., 1997). The whole genome was sequenced in 2000 by The Medicines Company, Ville St. Laurent, Quebec, Canada and patented (Pelletier et al., 2000). It was partially annotated (García et al., 2005) but has never been made available in public genome databases.

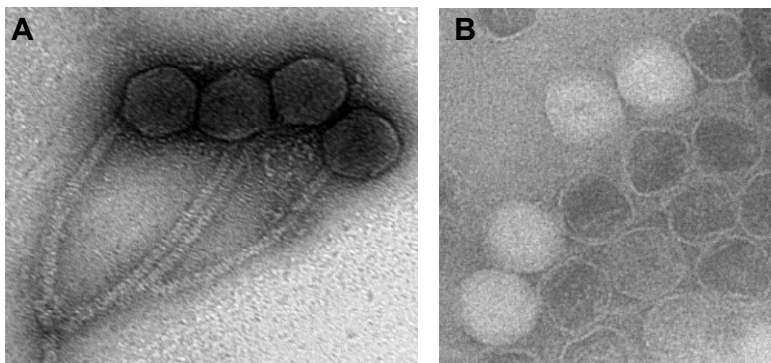


Fig. 7 EM micrographs of *Dp-1* virions

(A) Mature *Dp-1* virions and (B) proheads. Images were kindly provided by Mourad Sabri and Sylvain Moineau, Université Laval, Québec.

Their lytic enzymes

Of general research interest during the last decade was the isolation and characterization of the lytic enzymes of Cp-1 and Dp-1. They consist of two domains: an N-terminal catalytic domain responsible for degradation of the peptidoglycan layer and a C-terminal domain that consists of cell wall (choline) binding repeats. It was shown that the latter are responsible to target the enzymes to the cell wall, necessary in gram-positive bacteria since they lack an outer membrane. Thus, the local concentration can be maintained and release into the environment is avoided. The cell wall binding domain binds to choline in the cell wall and may help to orientate the catalytic domain correctly (Fischetti, 2010; Perez-Dorado et al., 2007).

The catalytic domains of the lytic enzymes can be distinguished by their substrate specificity: e.g., the lytic enzyme from Cp-1 (Cpl1) is defined as lysozyme (old term) or muramidase/glycosidase (later terms) since it acts on the carbohydrate backbone of the peptidoglycan. It hydrolyzes the β -1,4 linked N-Acetylglucosamine and N-Acetylmuramic acid repeats (Garcia et al., 1987). In contrast, the Dp-1 lytic enzyme Pal is an amidase that breaks down the amide bond that connects covalently the sugar and peptide moieties. It cleaves between the acetyl group of the N-acetyl-muramic acid and the L-alanyl residues as indicated by its name N-acetyl-muramoyl-L-alanine amidase (Garcia et al., 1984; Sheehan et al., 1997).

It turned out that purified Cp-1 lysozyme (Cpl1) and Dp-1 muramidase (Pal) can rapidly kill *S. pneumoniae in vitro* (Loeffler et al., 2003; Loeffler et al., 2001). Non-capsulated, encapsulated as well as penicillin resistant strains are killed efficiently. Doses of purified lytic enzymes given to *S. pneumoniae* infected mice (oral, vaginal, and nasal colonization models) were able to cure them (Loeffler et al., 2003; Loeffler et al., 2001; Witzenrath et al., 2009). The lytic enzymes act specifically against the host, the phage is also able to infect, underlining an evolutionary specification of them (Fischetti, 2010; Loeffler et al., 2001). Moreover, phage-derived lytic enzymes are highly efficient. For instance, *S. pneumoniae* lysin (LytC) that cleaves the same glycosidic linkage of pneumococcal peptidoglycan as Cpl1 shows 6,000 units per mg versus 100,000 units per mg of purified Cpl1 (Garcia et al., 1999). Thus, these zymogens are very promising for medical applications because they act very specifically while they have no side-effects on other gram-positives or humans.

They are also promising since the host might not easily become resistant against lytic enzymes: bacteriophages optimized these weapons since millions of years and the host cells had to change the integrity of the cell wall which is hard to attain in contrast to acquire an antibiotic resistance gene (Fischetti, 2010).

The Cpl1 protein crystal structure was the only resolved one so far of Cp-1 and Dp-1 proteins (Perez-Dorado et al., 2007) (Cpl1 3D-structure is shown in the results section Fig. 31D).

1.2 Conserved hypothetical proteins: new hints and new puzzles

Comparative genomics revealed that a substantial number of proteins of hitherto unknown function are distributed among organisms from several phylogenetic lineages. Thus, they are annotated commonly as “conserved hypothetical proteins” (CHPs). Because they are conserved they are expected to have conserved physiological functions. That is an important side-benefit: it is often easier to conduct experiments with bacterial than with eukaryotic proteins and functional characterization of microbial CHPs may facilitate prediction and subsequent experimental study of their human homologs. In fact, most of the highly conserved CHPs are thought to play a role in conserved pathways such as DNA metabolism and protein translation (Galperin and Koonin, 2004). However, they are often tough to analyze since their gene deletions exhibit no obvious phenotypes or there is no point to start at all since functional links are missing. With the availability of many experimental high-throughput datasets novel hints can now be found for these tough nuts.

Koonin and Galparin summarized such highly conserved hypotheticals and proposed the most interesting candidates for further experimental investigations which have human orthologs (Galperin and Koonin, 2004). They split CHPs into two groups: “unknown unknowns” have no functional information at all and “known unknowns” have little known function, e.g., the presence of a known domain signature. For instance, *E. coli* YchF belongs to the “known unknowns” since it is composed of a GTPase domain (Teplyakov et al., 2003). Co-expression with peptidyl-tRNA hydrolase makes it likely to function in protein translation. However, since 2003 no progress was made on its characterization and its exact function is still mysterious. Koonin’s top-10 list of “unknown unknowns” is headed by *E. coli* YbeC. But also in this case no experimental progress was made underlining that these proteins are difficult to analyze.

1.2.1 YbeB – a highly conserved ribosome-associated protein

In other cases “unknown unknowns” changed into “known unknowns”. For instance, *E. coli* YbeB was ranked originally as third protein on the list of “unknown unknowns” by Koonin and Galparin. In the meantime several protein-protein datasets and other high-throughput datasets were published that found binary protein interaction partners for various YbeB homologs (Kazuta et al., 2008; Parrish et al., 2007; Rain et al., 2001; Titz et al., 2008). However, these sets include interaction partners belonging to various cellular functions and thus are not really helpful on the raw dataset level.

One striking finding was done by Jiang and colleagues (Jiang et al., 2007). They showed that a set of *E. coli* CHPs co-migrate with different ribosome fractions. In the case of YbeB it was shown that it co-migrates exclusively with the large ribosomal subunit and thus might play a role in translation. Also its maize homolog Iojap was shown to bind to the large ribosomal subunit of maize chloroplasts (Han and Martienssen, 1995) and large-scale identification of *E. coli* and yeast protein complexes could detect

YbeB in complexes with ribosomal proteins and the large ribosomal subunit again (Butland et al., 2005; Krogan et al., 2006).

The loss of maize Iojap leads to a striking phenotype. As soon as its gene is mutated, the chloroplast's ribosomal machinery breaks down and leads to the loss of ribosomes and thus to pattern-striped albinism (Jenkins, 1924; Rhoades, 1943) (Fig. 8). While this phenotype was the subject of several studies (Byrne and Taylor, 1996; Halter et al., 2004; Zubko and Day, 1998) the effect could not be explained up to date and remains mysterious. In *E. coli* general phenotypes are surprisingly absent when *ybeB* is deleted implying a putative accessory role of YbeB in bacterial translation (Jiang et al., 2007). Although several studies were able to find a physical association of YbeB with the large ribosomal subunit its role in translation is uncertain.



Fig. 8 Iojap-affected maize

Corn seedlings with albino-striped leaves due to the homozygous nuclear mutation of the *ij* gene (Iojap). Images taken from <http://www.forestryimages.org>. (P.F. Byrne, Bugwood.org)

1.2.2 A view on the ribosome

The ribosome is responsible for translating the genetic information from mRNA into a polypeptide chain. It is one of the most complex biological nanomachines since many structural proteins, structural and catalytic RNAs, transfer RNAs, small molecules, regulatory factors, and enzymes have to play together during its biogenesis and to convert the nucleic acid code into functional proteins. The ribosome is composed of a large and a small ribonucleoprotein subunit. In prokaryotes the large subunit sediments with 50S, the small with 30S and assembled subunits with 70S. The small subunit is made of 16S rRNA and 21 ribosomal proteins. The large subunit consists of a 23S rRNA, a 5S rRNA, and 33 proteins (Kaczanowska and Ryden-Aulin, 2007) (Fig. 9).

Translation – what we still do not know

While protein translation is one of the best understood biological processes, there are still many open questions. The core components of translation like ribosomal rRNAs and proteins and the translation factors are well-known. Also maturation factors that are responsible for ribosome biogenesis are mostly characterized. Novel insights into regulatory mechanisms show that there exist manifold possibilities to regulate translation initiation. Recent results of large-scale experiments indicate that we are far from knowing all factors that play a role protein translation. Kazuta and colleagues checked the influences of expressed *E. coli* ORFs genome-wide in an *in vitro* translation system. They identified

344 proteins to have a beneficial and 159 proteins to have a deleterious effect on translation efficiency. Thus, approx. 12% of all *E. coli* proteins might also be involved in translation *in vivo* (Kazuta et al., 2008). It turned out that many enzymes had a positive or negative effect on translation. Also many hypothetical proteins affected the translation system. Combination with protein complex data revealed that 60 of the 159 effecting proteins interact with minimum one core component of the ribosome (ribosomal proteins and translation factors) and thus are physically associated with the ribosome (Arifuzzaman et al., 2006; Butland et al., 2005; Kazuta et al., 2008). Other high-throughput interaction studies indicate additional interactions of unknown and unexpected proteins from various species with the translation core components (for references, see Tab. 7). Medium-scale experiments could also classify a large set of novel ribosome-associated proteins. For instance, 10 novel mitochondrial ribosomal proteins of yeast and nine of *E. coli* were identified (Gan et al., 2002; Jiang et al., 2007). Consequent characterization of few of such proteins could already determine their functions. For instance, *E. coli* protein YbeY was shown to function as general chaperone for ribosomal proteins under high temperature growth conditions (Rasouly et al., 2009). *E. coli* CgtA and YjeQ were shown to function as GTPases in late steps of ribosome biogenesis (Campbell and Brown, 2008; Jiang et al., 2006). The examples mentioned here underline the impact of proteome-wide experiments for unraveling protein functions step by step. Highly conserved and widely evolutionary distributed proteins of unknown function might be involved in protein translation but are often tough to analyze since phenotypes or other experimental indicators are not detectable.

Ribosome biogenesis (E. coli)

Ribosome biogenesis is complex and involves stepwise rRNA modification, ordered binding of ribosomal proteins and metal ions, and sequential conformational changes. Assembly *in vivo* takes about two minutes (Lindahl, 1975).

Biogenesis starts in parallel to transcription of the rRNA genes that are encoded by one transcriptional unit. Formation of hairpins and other secondary structures in the polygenic RNA leads to binding of ribosomal proteins. Cleavage by several RNases as well as chemical base modifications take place (Williamson, 2003). Endonuclease RNase III cleavage results in precursors of 23S, 16S, and 5S rRNA and some tRNAs. While trimming the 23S rRNA precursor is dependent on RNaseIII, RNaseT, and an unknown RNase, 16S rRNA termini are shortened by RNaseE and G and an unknown RNase (Kaczanowska and Ryden-Aulin, 2007). Production of 5S rRNA requires RNaseP and E (Bothwell et al., 1976; Roy et al., 1983).

Chemical modifications of base and sugar residues take place frequently in rRNAs (as well as in tRNAs). However, in 5S rRNA chemical modifications are unknown. Uridine can be converted to pseudouridine or carbonyl, methyl, amino, or thio groups can be added. The 16S rRNA is known to have 11 modified positions, of which 10 are methylations and one is a pseudouridine. 23S rRNA contains 25 known modifications. Here, 14 are methylations, nine are pseudouridines, one is a

methylated pseudouridine, and one is unknown. Most of the modifications in both the 16S and 23S rRNA are clustered in the decoding region and the peptidyl transferase center (Decatur and Fournier, 2002). Chemical modifications are needed for an increased binding efficiency of tRNAs to the 16S rRNA or protein L16 to the 50S particle, and for optimization of rRNA folding (Chelbi-Alix et al., 1981; Krzyzosiak et al., 1987). Modifications are mediated by maturation enzymes belonging to two classes (methyltransferases and pseudouridine synthases). About 27 enzymes are known to be responsible to modify a certain rRNA residue. Notably chemical modifications take also place in ribosomal proteins. Post-translational modifications have been reported for five and six ribosomal proteins of the small and large subunit of *E. coli*, respectively. Methylation, ethylation, addition of glutamic acid residues, and a partial modification with isoaspartate are known (Kaczanowska and Ryden-Aulin, 2007). For instance, L11 is polymethylated at nine various positions (Colson et al., 1979). S6 is modified by six glutamic acid residues (Kang et al., 1989). Although the role of such protein modifications in the context of ribosomal proteins is unclear, they may alter and optimize the protein specificities.

Moreover, various RNA helicases, RNA and protein chaperones, and ribosome-dependent GTPases are important for the proper ribosome maturation. They guarantee the proper folding of rRNAs at certain checkpoints during early, middle, and late phases of ribosome biogenesis (Kaczanowska and Ryden-Aulin, 2007).

From the genetic code to a protein

Essentially, three major steps occur during protein translation (described here for *E. coli* translation system):

Initiation is promoted by three initiation factors (IF1, IF2, IF3). It involves the binding of the Shine-Dalgarno (SD) sequence of mRNAs to the anti-SD sequence of 16S rRNA on the 30S subunit. Then, the initiation factors help to accommodate the start codon at the P-site followed by the assembly of 50S subunit with the initiator complex on the 30S subunit. Initiation is thought to be the rate-limiting step. Initiation assembly takes on the order of seconds while elongation occurs at a faster rate with the synthesis of ~20 amino acids per second (Rodnina et al., 2007).

During elongation the polypeptide chain is synthesized (Fig. 9). The initiator tRNA is loaded with formylmethionine and located to the P-site while in the A-site a corresponding aminoacylated tRNA is loaded corresponding to the second mRNA codon. The latter is brought as ternary complex with elongation factor Tu (EF-Tu)-GTP. After GTP hydrolysis, the elongation factor releases the aminoacyl end of the A site tRNA. This allows the tRNA to swing into the P-site. As a consequence, aminoacyl residues of both tRNAs are positioned at the peptidyl transferase center on the 50S subunit and peptide bond formation occurs. This is catalyzed by the 23S rRNA peptidyl transferase center which has a ribozyme activity. The deacylated initiator tRNA is moved for release from the P site to the E site. The translocation reaction is driven by GTP hydrolysis of EF-G. Repetitions of this step lead to a peptidyl

chain that exits the ribosome through a tunnel at the top of the 50S subunit (Kaczanowska and Ryden-Aulin, 2007).

Release factor (RF) RF-1 or RF-2 recognize the mRNAs stop codons as soon as they are located in the A-site. This triggers termination by the release of the peptide chain from peptidyl-tRNA located in the P-site. Afterwards, RF3 binds to RF1 or RF2 and releases them from the A-site.

The ribosomal particle still contains the empty tRNA in the P-site and the mRNA. Dissociation of the subunits is accomplished by ribosome recycling factor (RRF) with EFG. Subsequently, IF-3 removes the tRNA from the 30S subunit, allowing the mRNA to dissociate or to undergo new SD/anti-SD contacts (Kaczanowska and Ryden-Aulin, 2007).

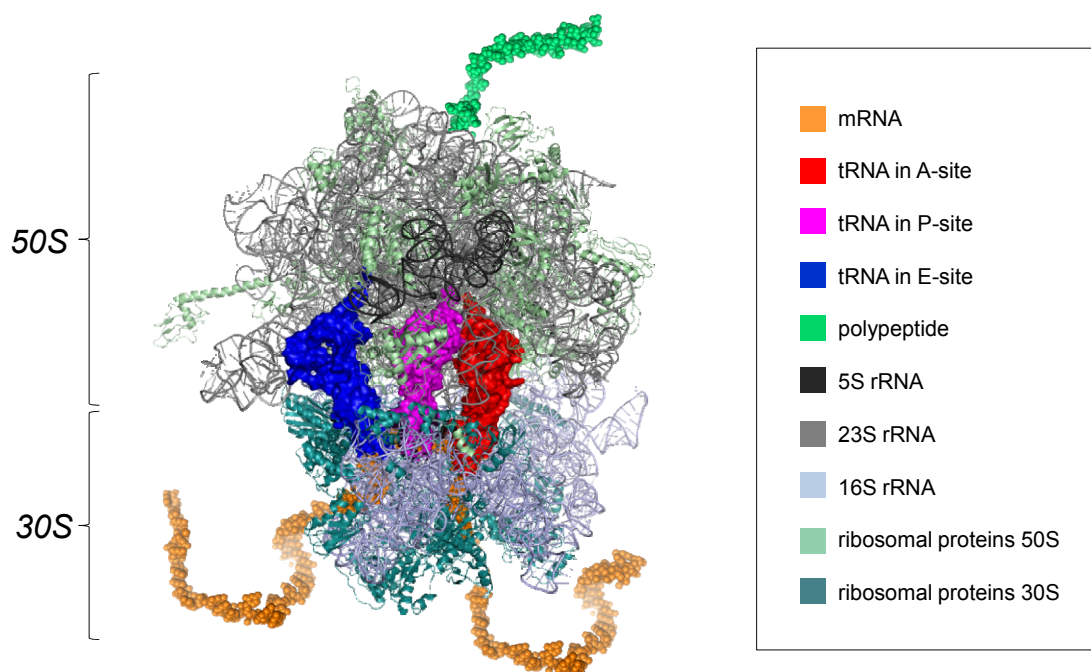


Fig. 9 A translating ribosome during peptidyl chain elongation

A ribosomal particle bound with mRNA and tRNAs. The mRNA chain is figuratively extended and a freshly synthesized polypeptide chain is indicated. Figure was drawn with Pymol after PDB entry 3i8f and 3i8g (Jenner et al., 2010).

Comparison of translational systems

Each organism has at least one complete ribosomal system for protein synthesis. Eubacteria and Archaea contain cytosolic ribosomes. Plants have cytosolic ribosomes and separate systems that act in their mitochondria and plastids. Fungi and animals have cytosolic and mitochondrial ribosome systems. Although protein translation and the ribosome are well-conserved there are some remarkable differences between the major clades.

Mitochondrial ribosomes and thus the mitochondria are thought to be derived from aerobic α -proteobacteria. In contrast, the chloroplast's origin including its translation system is believed to be of cyanobacterial origin (Aravind and Koonin, 2000; Gray, 1989; Gray, 1992). Cytosolic ribosomes of Eukaryota are more closely related to Archaea and thus are thought to be of archaeal origin. It was shown that archaeal rRNAs and ribosomal proteins are more closely related to eukaryotic ribosomal

rRNAs and proteins (Allers and Mevarech, 2005; Bult et al., 1996). However, Archaea frequently utilize Shine-Dalgarno sequences to identify translation start codons and use polycistronic mRNAs, similar to Eubacteria. Alternatively, Archaea can also initiate translation with leaderless mRNAs, similar to initiation of Eukaryotes (Benelli et al., 2003). Another criterion that indicates a closer relationship of Archaea and cytosolic eukaryotic translation is the set of initiation factors. Ten are known for Archaea while Eubacteria have only three. The archeal factors share a high homology with those from Eukaryotes. Similarly, archaea and eukaryote translation initiation makes use of methionine while bacteria use formyl-methionine (Allers and Mevarech, 2005; Londei, 2005). Moreover, in *E. coli* many cis- and trans-acting regulatory factors are known to control the speed of initiation positively or negatively. For instance, secondary structures in the mRNAs' 5'-untranslated region (5'-UTR) can alter during environmental changes. This phenomenon is also called "riboswitch". Differences in temperature can lead to conformational changes, making the mRNA more or less affine to the anti-SD sequence. Moreover, trans-acting factors like small molecules/metabolites, short regulatory RNAs, as well as protein ligands can regulate translation initiation through 5'-UTR/mRNA binding (Marzi et al., 2008). In general, bacteria modulate the accessibility of the 5'-UTR to control initiation.

In comparison, eukaryotic and archaeal translation initiation is controlled by a rather low number of trans-acting ligands. In eukaryotes, at least 30 initiation factors participate in translation (Sonenberg and Hinnebusch, 2009). In principle, initiation factors including eIF-E, A, and G, bind to a mRNA 5'-cap region and scan for triplets until they find a start codon. Proteins that bind to the 3'-poly-A tail can lead to a loop formation with the 5'-cap complex and can stimulate initiation. This occurs on the small ribosomal subunit and locates the initiator Met-tRNA to the P-site. Additional factors bound to the 5'-cap region can trigger the usage of alternative start codons. Other factors can bind to the 3'-UTR poly-A tail and make the mRNA inaccessible for translation also by loop formation of the mRNA-protein complex. Moreover, initiation regulation in eukaryotes includes inactivation of eIF-2 by phosphorylation as an example for post-translational modification of regulatory factors. Translational regulation is important during environmental changes, virus infection, and during the cell cycle, e.g., the negative regulation of translation in G₀ phase. Then eIF-2 is phosphorylated and thus inactivated. Moreover, miRNAs are known to be highly relevant for translational regulation (Sonenberg and Hinnebusch, 2009). In summary, regulation of translation initiation in bacterial and eukaryotic translation systems is different since these life forms had to face different issues regarding cell cycle control, development, and environmental changes which is more complex in eukaryotes.

1.3 Interactome screening with the Yeast Two-Hybrid (Y2H) System

Shortly after Stanley Fields and Ok-kyu Song invented the Yeast Two-Hybrid System in 1989 and demonstrated the system to work for a single protein interaction (Fields and Song, 1989) it was adapted for screens of random libraries. Today, matrix-based screens are used primarily for smaller and medium size clone collections in combination with automation and cloning techniques that allow for reliable and fast interaction screening. Matrix-based Y2H screens are an alternative to library-based screens. However, intermediary forms are also possible. Recent improvements of matrix screens (also called array screens) use various pooling strategies as well as novel vectors that increase its efficiency while decreasing false-negative rates and increasing reliability. Since the Y2H system was used in this work as elementary method I will introduce it briefly.

1.3.1 The Yeast Two-Hybrid principle

Like other variations of the Y2H system matrix-based assays are usually carried out in living yeast cells although in theory any other cell could be used. This is a crucial advantage since it represents an “*in vivo*” situation. The proteins of interest are provided as plasmid-encoded recombinant fusion proteins (Fig. 10). The bait protein is fused to a DNA-binding domain (DBD) of the yeast GAL4 transcription factor. The prey protein is tagged by the activation domain (AD) of GAL4. A physical contact of the bait and prey protein simulates the reconstitution of the GAL4 transcription factor. Once the bait protein binds to its promoter sequence by its DBD the interacting proteins recruit the basal yeast transcription machinery and thus activate the expression of a reporter gene.

For high-throughput screens a HIS3 auxotrophy marker is used routinely. It encodes the essential enzyme imidazoleglycerol-phosphate dehydratase which catalyses the sixth step of histidine biosynthesis. Hence, yeast growth on minimal medium that lacks histidine can be used to indicate an interacting protein pair. In contrast, non-interacting pairs cannot support growth on minimal medium since they cannot activate transcription of the reporter gene. This reporter system is very simple and easy to use because the presence of yeast colonies indicates a binary protein interaction. Many other reporter genes are conceivable as long as they can be activated by the interacting fusion proteins.

Before the binary tests are carried out, the bait and prey plasmids must be brought into the same yeast cell. This is done conveniently by mating. The bait and prey plasmids are transformed separately into haploid yeast cells of different mating types, a and α . Mating results in diploid yeast strains that carry the genetic material of both haploids including the bait and prey plasmids.

Alternative reporter genes are LEU2 and URA3. They allow selection on readout medium that lacks leucine or uracil. Auxotrophy markers are not the only ones that can be used. The ADE2 reporter system changes colony color from red to white on adenine starvation medium when diploids express interacting proteins. β -galactosidase (*lacZ*) or GFP (green fluorescent protein) can be used as

colorimetric or fluorescence reporters. Finally, transcription-independent two-hybrid systems have been developed. The Split-Ubiquitin System is based on the cleavage of one of the interacting fusion proteins by the proteasome (Johnsson and Varshavsky, 1994).

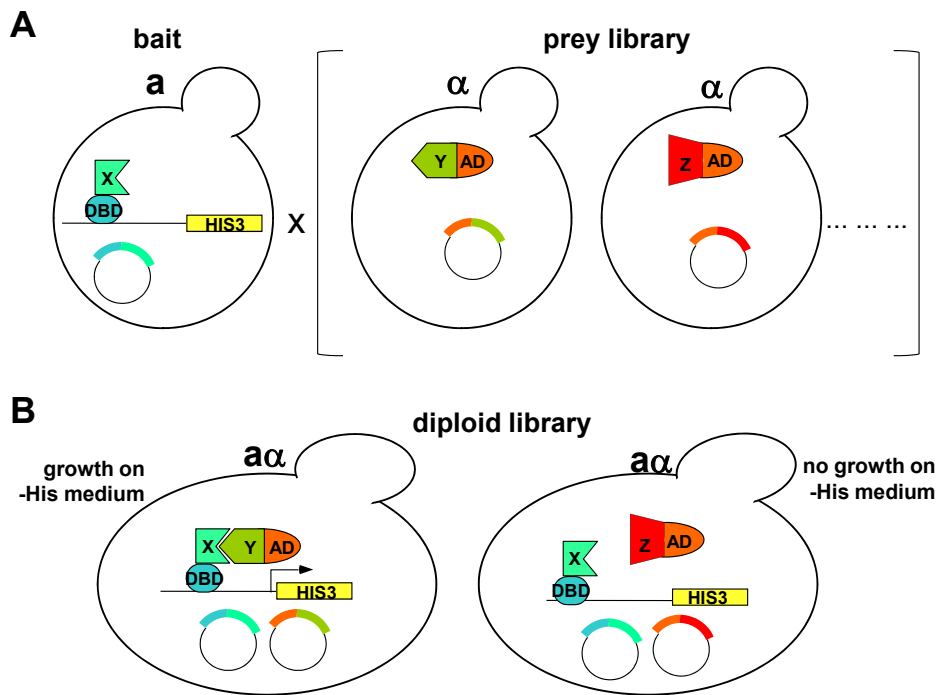


Fig. 10 The Y2H principle

(A) Haploid yeast cells of mating type *a* are transformed with a bait plasmid and those of mating type α with prey plasmids. A single bait strain is mated with a prey library. (B) Resulting diploids (*a*/ α) carry the genetic material of mated haploids. Interacting fusion proteins activate expression of the HIS3 reporter gene which assures survival on minimal medium that lacks histidine (left); diploids with non-interacting fusions cannot grow (right).

1.3.2 Applications

It has become clear that the ability to conveniently perform unbiased library screens is the most powerful application of the Y2H system. With whole-genome arrays such unbiased screens can be expanded to all proteins of an organism or any subset thereof. Arrays, like traditional Two-Hybrid screens, can also be adapted to answer many questions that involve protein-protein or protein-RNA interactions (Tab. 6).

Tab. 6 Applications and use of two-hybrid assays besides protein-protein interaction screens

Application	Reference
Identification of mutants that prevent or allow interactions	(Schwartz et al., 1998)
Screening for drugs that affect interactions	(Vidal and Endoh, 1999; Vidal and Legrain, 1999)
Identification of RNA-binding proteins	(SenGupta et al., 1996)
Semi-quantitative determination of binding affinities	(Estojak et al., 1995)
Map interacting domains	(Boxem et al., 2008; Rain et al., 2001; Vollert and Uetz, 2004)
Study protein folding	(Raquet et al., 2001)
Map interactions within protein complexes	(Cagney et al., 2001)

Recent large-scale projects have been successful in mapping systematically whole or partial proteomes of various higher and lower organisms (Tab. 7). In addition to bacteria and eukaryotic genomes, several viral proteomes have been mapped, e.g., bacteriophage T7 (Bartel et al., 1996) and several herpesviruses (Fossum et al., 2009; Uetz et al., 2006).

In combination with structural genomics, gene expression data, and metabolic profiling, the enormous amount of information in these networks help us to model complex biological phenomena in molecular detail or to link unknown proteins to certain pathways.

Tab. 7 Published proteome-scale and comprehensive Y2H projects

Species	Reference
<i>Saccharomyces cerevisiae</i>	(Ito et al., 2001; Uetz et al., 2000; Yu et al., 2008)
<i>Drosophila melanogaster</i>	(Giot et al., 2003)
<i>Caenorhabditis elegans</i>	(Li et al., 2004)
<i>Homo sapiens</i>	(Rual et al., 2005; Stelzl et al., 2005)
<i>Helicobacter pylori</i>	(Rain et al., 2001)
<i>Campylobacter jejuni</i>	(Parrish et al., 2007)
<i>Treponema pallidum</i>	(Titz et al., 2008)
T7 phage	(Bartel et al., 1996)
Herpesvirus EBV, KSHV, mCMV, VZV, HSV-1	(Fossum et al., 2009; Uetz et al., 2006)

1.3.3 Matrix-based Yeast Two-Hybrid Screens (one-on-one)

“Matrix-” or “array-based” means that preys are organized in a defined array format. For high-throughput purposes preys can be arranged in 384-format on a single test plate. This was first demonstrated on a global scale by Uetz and colleagues (Uetz et al., 2000). Each prey clone maps to an individual position. Preys may be organized as individual colonies. An arrangement as duplicate or quadruplicate copies helps to ensure reproducibility (Fig. 11).

The whole array of haploid preys is usually mated with a single bait of the opposite mating type. Thus, each potential interaction pair is tested one-on-one (Fig. 11). For high-throughput analyses the usage of a replication robot is recommended, typically with a 96- or 384-pin tool (Fig. 11B). It automates the procedure by reproducibly stamping up to hundreds of array position in a single step, e.g. to transfer diploids onto readout plates.

Why matrix-based screens?

Matrix-based screens are excellent to control experimental background signals. Background can be caused by self-activation of certain bait proteins. Self-activation is defined as a detectable bait-dependent reporter gene activation in the absence of any prey interaction partner. For instance, in yeast about 7.5% of 6,000 bait proteins exhibit self-activation and many of the strongly self-activating baits have been identified as transcription factors (Titz et al., 2006b). In matrix-based screens, interactions of self-activating baits can be identified even if such a background growth occurs. Therefore, prior to the Two-Hybrid analyses, the bait yeast strains have to be examined for their self-activation property.

Weak to intermediate-strength self-activator baits can be used in Two-Hybrid array screens because the corresponding bait-prey interactions confer stronger signals than the self-activation background. If the HIS3 reporter is used (not possible with other genetic reporters), the self-activation background can be suppressed by adding a bait-specific, minimal inhibitory concentration of 3-AT (3-Amino-1,2,4-triazole) to the readout medium, a competitive inhibitor of HIS3 (Hilton et al., 1965). 3-AT was originally applied as herbicide.

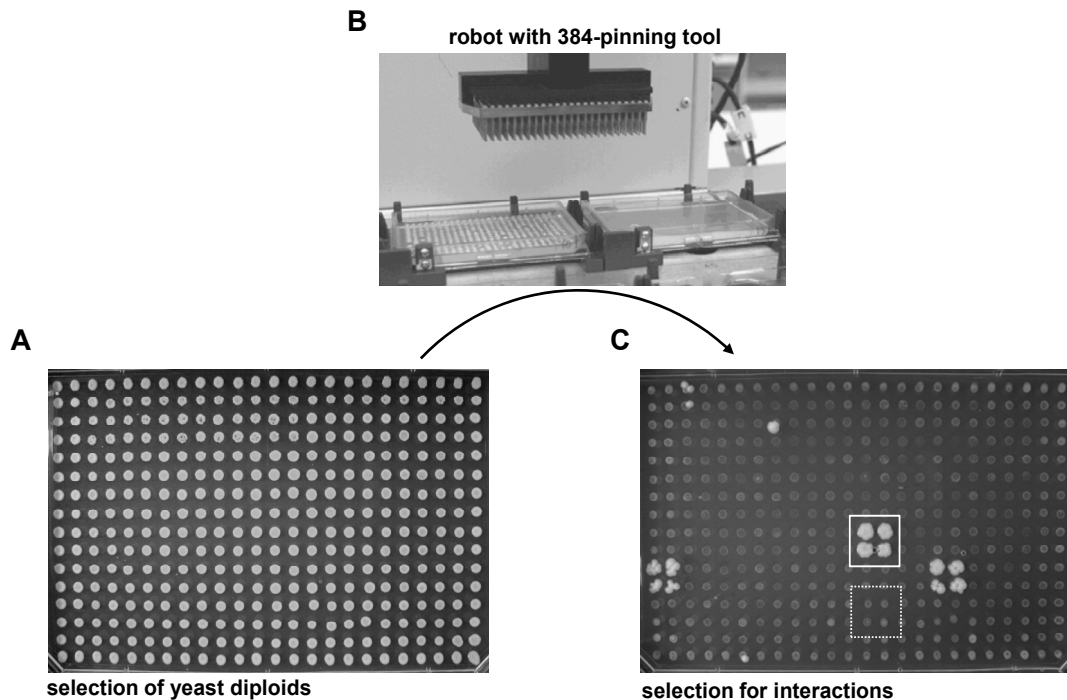


Fig. 11 A matrix-based screen

(A) A prey array mated with a single bait on diploid selective agar medium containing 96 individual preys. Single preys are replicated as quadruplicates to check interaction reproducibility. (B) 384-pinning tool of replication robot during pinning step of diploids onto readout medium. (C) Diploids on readout medium that lacks histidine. Diploids are usually grown on selective medium for one week at 30°C. Activation of the HIS3 reporter leads to growth on readout medium indicating a pairwise interaction (see quadruplicate position in white square). Non-interacting pairs do not support growth on minimal medium (e.g., quadruplicate in dashed square). Background signals like randomly growing cells can be tracked easily on the matrix and can be retested (e.g., top left).

In a matrix screen of a single bait the signal-to-noise ratio can be easily determined because all protein pairs are assayed under equal conditions. Furthermore, background of spontaneously appearing colonies caused by mutations or other random effects can be identified (e.g., see Fig. 11C). The redundancy of two or more test positions helps to winnow random colonies.

Another crucial advantage is that interacting preys can be simply identified by their positions. The matrix positions can be stored in a list or more comfortably in a database. Thus, identification of the interacting prey by sequencing is not required and time and costs can be minimized.

Finally, the matrix approach helps to distinguish strong from weak and spurious interactions since the size of growing yeast colonies is a semi-quantitative measure of binding affinity (Estojak et al., 1995).

1.3.4 (Mini-) pool screening strategies

The capacity of matrix-based screens is limited by the size of the clone set to be tested. For instance, a small proteome that encodes for 1,000 proteins requires at least $1,000^2$ individual pairwise tests in a comprehensive screen. For large genomes, such as the human, $23,000^2$ (over half a billion) one-on-one tests would be necessary to test all possible combinations. Genome-wide screens face three main issues: cost and efficiency (the number of assays and speed), specificity (detecting false-positives), and sensitivity (avoiding false-negatives).

Solutions to make large-scale matrix screens more efficient require pooling (Tab. 8) which may reduce the number of individual Y2H tests dramatically as well as the need for sequencing while keeping the advantages of matrix-based screens. Smart pooling and arrangements of prey as well as bait clones can help to speed up the screening procedure drastically, resulting in interaction detection with (almost) the same sensitivity and specificity as in one-on-one Yeast Two-Hybrid screens. Sophisticated pooling strategies which are partially based on complicated mathematical background operations have been developed (Tab. 8).

Tab. 8 Y2H pooling strategies

Name	Description	Reference
Library-prey pool screen	Screen 1 bait x whole ORFeome pool of preys	e.g., used in (Bartel et al., 1996)
Mini-pool screens	Matrix-based screen of 1 bait x n preys per array position (in case of Rual et al. 1 bait x 188 human prey clones).	e.g., used in (Rual et al., 2005)
Two-Phase Mating	Mating of 1 prey x bait array followed by mating of positive bait x prey array. Identification of interacting bait/prey possible just by matrix position.	(Zhong et al., 2003)
PI-Deconvolution	Screening of 2n baits in 2n pools against a redundant prey matrix; interacting bait is identified by a unique tag that correlates with preys positives.	(Jin et al., 2006)
Smart Pool Array	PI-Deconvolution strategy: instead of individual preys, well-designed prey pools are screened in an array format that allows built-in replication and prey-bait deconvolution.	(Jin et al., 2007)
Shifted Transversal Design	Achievement of an extra large experimental redundancy by a shifted transversal prey matrix design. Interacting preys are identified by their characteristic matrix pattern.	(Xin et al., 2009)

In matrix-based pooling screens several preys share a position. In a simple case a prey array that consists of 960 individual preys can be collapsed into a single 96-well plate with 10 clones in each position (Fig. 12). This minimizes the required mating operations with a single bait by 1/10. The disadvantage of this strategy is that interacting preys cannot be identified immediately as it is possible in one-on-one matrix screens. They must be identified by yeast colony PCR followed by sequencing or retesting of individual bait-prey pairs.

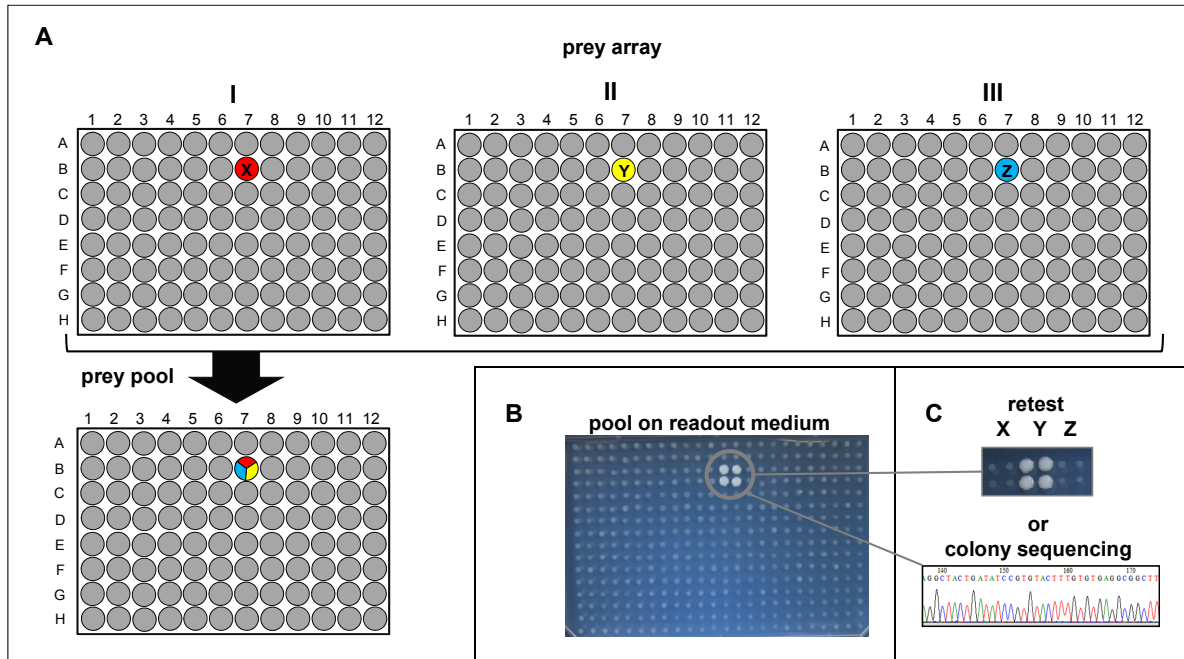


Fig. 12 Principle of mini-pool screens

(A) A full-matrix prey array that consists of three plates (I, II, III), each with 96 individual preys. Positions are collapsed into a single prey pool plate. Pooling results in mini-pools which consist of three different preys (e.g., X, Y, and Z). (B) Mini-pool on readout medium as quadruplicates. Pool B7 exhibits an interaction. Prey X, Y, and Z are potential interaction partners. (C) Identification of the interaction partner by a one-on-one retest assay. Prey Y is identified as the interaction partner whereas X and Z do not interact. Alternatively the interacting prey can be identified by a colony PCR followed by sequencing.

1.4 Motivation

This work was motivated by a major and a minor aim:

Major project: Functional proteomics of the lytic bacteriophages Cp-1 and Dp-1 of *S. pneumoniae*

The sequencing era of whole phage genomes reached a critical information limitation: 50 up to 90% of gene products cannot be annotated at all by homology predictions. To understand the biology of these phages, there is an urgent need of comprehensive experiments beyond sequencing. Since systematic, functional experiments of bacteriophages have been rare in contrast to bacteria and higher organisms, the motivation for me was to push the traditional phage biology into the proteomics era.

As a starting point, I systematically screened the proteomes of *Pneumococcus* phage Cp-1 and Dp-1 against themselves with the expectation to find novel functional protein links for unknown as well as annotated proteins. The mentioned phages are of general medical interest since they are among the few lytic viruses infecting *S. pneumoniae*, a major cause of human death caused by bacterial infections. Thus, they are promising candidates for phage therapy. Although these two phages are the best characterized lytic *Pneumococcus* phages, we are far from understanding their biology. But exactly this is important to create “Frankenstein’s phage”, a synthetic or modified phage that infects the host cell efficiently and might be applicable in phage therapy. Systematic approaches are necessary to find unexpected phenomena that are usually not found by “bottom up” approaches.

Moreover, nobody systematically screened any phage for protein-protein interactions with host proteins so far. Since viruses do not stand alone but reproduce in dependence of a host cell, their biology cannot be regarded in isolation. To get deeper and comprehensive insights into the phage-host relationship I carried out proteome-wide screens for virus-host interactions. Since Dp-1 and Cp-1 belong to completely different families I also addressed the question if interactions can be found that indicate similar or different manipulation strategies and thus if these two phages evolved analogous mechanisms.

Finally, there was the task to annotate the Dp-1 genome. Its genome has been sequenced. However, it has been only partially annotated and its comprehensively annotated genome was never published.

By combining results from genomic as well as proteomic data, I addressed the following questions:

- How can a phage hijack the host?
- Which pathways and which cellular levels are targeted?
- Evolution: can interactions be found that indicate analogous manipulation/infection strategies?
- Is the genomic organization of phages reflected by their protein interaction network?
- And thus, can protein interaction data be valuable to predict functions of orphan proteins?
- What can we learn about individual proteins and their functions?

Minor project: Functional analysis of the conserved hypothetical protein YbeB and its interaction with the ribosomal protein L14

I started to work on conserved hypothetical proteins during my master's thesis. At that time I screened systematically for protein-protein interactions of conserved hypothetical proteins of the Spirochaete *Treponema pallidum* assuming that interactions among conserved proteins might help to get new ideas about their physiological function. However, the identified interactions and thus the functional links were hard to interpret since nothing or not much was known about these putative proteins except that they are conserved. In addition, available low-, medium-, and high-throughput experimental data from other species that would have allowed for a comprehensive comparison were rare.

In the meantime several high-throughput studies have been published. These were very useful to find novel functional links on proteins I was working on: here I focused on the functional analysis of the conserved hypothetical protein YbeB from *E. coli* and its orthologous relatives – a widely evolutionary distributed group of “Tojap-like” proteins, present from bacteria to man. A recent study revealed co-migration of YbeB of *E. coli* with the large ribosomal subunit and proposed involvement of YbeB in protein translation although phenotypes of its gene deletion mutant are absent. To find novel hints about its mysterious function I reviewed the existing literature ranging from homologs of various species. Furthermore, I examined putative protein-protein interaction links of YbeB experimentally that were published from high-throughput experiments. The aim for me was to learn more about the function of YbeB because of its high evolutionary distribution. The general questions I wanted to answer are:

- Where does YbeB bind on the ribosome?
- Is this/these interaction(s) conserved and thus YbeB's function?
- Can this/these interaction(s) help to find deeper insights into YbeB function?
- Since there was no phenotype reported so far for *E. coli ybeB* gene knock-out strains: does the loss of YbeB in *E. coli* result in phenotypes that link it with the translational process *in vivo*?
- What is the molecular function of YbeB and its orthologs?
- Can we learn something about ourselves (or better, can we still find conserved principles in bacteria that are relevant to human biology)?

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals and enzymes

Tab. 9 Chemicals

Name	Source
3-AT (3-amino-1,2,4-triazole)	Sigma-Aldrich, Taufkirchen
Acetic acide, glacial	Roth, Karlsruhe
Acrylamide	Roth, Karlsruhe
ADP (adenosine diphosphate)	Sigma-Aldrich, Taufkirchen
Agar agar	Roth, Karlsruhe
Agarose	Peqlab, Erlangen
ATC (anhydrotetracycline)	Sigma-Aldrich, Taufkirchen
Ammonium chloride (NH ₄ Cl)	Roth, Karlsruhe
Ammonium molybdate tetrahydrate ((NH ₄) ₆ Mo ₇ O ₂₄ •4H ₂ O)	Roth, Karlsruhe
Ampicillin sodium salt	Roth, Karlsruhe
Amylose resin (E8021)	New England Biolabs, Frankfurt
APS (ammonium persulfate)	Roth, Karlsruhe
ATP (adenosine triphosphate)	Sigma-Aldrich, Taufkirchen
Bacto Yeast Extract	Roth, Karlsruhe
Boric acid (H ₃ BO ₃)	Roth, Karlsruhe
Bromophenolblue	Sigma-Aldrich, Taufkirchen
BSA (bovine serum albumin)	Promega, Mannheim
Calcium chloride dihydrate (CaCl ₂ •2H ₂ O)	Roth, Karlsruhe
Chloramphenicol	Sigma-Aldrich, Taufkirchen
Cobalt(II) chloride (CoCl ₂)	Roth, Karlsruhe
Coomassie Brilliant Blue R-250	Sigma-Aldrich, Taufkirchen
Copper(II) sulfate (CuSO ₄)	Roth, Karlsruhe
DEPC-Treated Water, UltraPure	Invitrogen, Karlsruhe
DMEM - Dulbecco's Modified Eagle Medium	Invitrogen, Karlsruhe
DMSO (dimethyl sulfoxide)	Roth, Karlsruhe
DNA ladder – 1kB	Invitrogen, Karlsruhe
dNTP mix (10 mM, each dNTP)	Sigma-Aldrich, Taufkirchen
dNTP mix (2.5 mM, each dNTP)	Takara Bio Inc., Potsdam
Donor calf serum (DCS)	Invitrogen, Karlsruhe
DRAQ5	Biostatus, Shepshed, UK
DTT (dithiothreitol)	Roth, Karlsruhe
EDTA (ethylenediaminetetraacetic acid)	Roth, Karlsruhe
Ethidium bromide	Roth, Karlsruhe
EtOH (ethanol)	Roth, Karlsruhe
FCS (fetal calf serum)	Invitrogen, Karlsruhe
Fungizone (500x)	Invitrogen, Karlsruhe
G418 sulfate	PAA Laboratories, Cölbe
Gel filtration standards	Biorad, München

MATERIALS AND METHODS

Name	Source
Gentamicin Reagent Solution (50 mg/ml), liquid	Invitrogen, Karlsruhe
Gentamycine	Roth, Karlsruhe
Glucose	Roth, Karlsruhe
Glutaraldehyde	Sigma-Aldrich, Taufkirchen
Glutathione Sepharose 4 Fast Flow	GE Healthcare, München
Glycerol	Roth, Karlsruhe
Glycin	Roth, Karlsruhe
HCl	Roth, Karlsruhe
Imidazole	Sigma-Aldrich, Taufkirchen
IPTG (isopropyl β -D-1-thiogalactopyranoside)	Sigma-Aldrich, Taufkirchen
Isopropanol	Roth, Karlsruhe
Kanamycin sulfate	Roth, Karlsruhe
L-(+)-arabinose	Sigma-Aldrich, Taufkirchen
Leupeptin	Sigma-Aldrich, Taufkirchen
Lithium acetate	Roth, Karlsruhe
Magnesium chloride (MgCl ₂)	Roth, Karlsruhe
Maltose	Sigma-Aldrich, Taufkirchen
Manganese(II) chloride (MnCl ₂)	Roth, Karlsruhe
Methanol	Roth, Karlsruhe
Milk powder	Saliter, Obergünzburg
MitoTracker® Green FM	Invitrogen, Karlsruhe
MOPS	Roth, Karlsruhe
Ni-NTA Superflow	Qiagen, Hilden
Non-essential amino acids (NEAA; 100x)	Invitrogen, Karlsruhe
NP-40 (Igepal CA-630)	Sigma-Aldrich, Taufkirchen
ONPG (2-nitrophenyl β -D-galactopyranoside)	Sigma-Aldrich, Taufkirchen
Optimem	Invitrogen, Karlsruhe
PBS 10 x solution	Invitrogen, Karlsruhe
PEG (polyethylenglykole)	Sigma-Aldrich, Taufkirchen
Penicillin/streptomycin (100x)	Invitrogen, Karlsruhe
peqGOLD prestained protein ladder	Peqlab, Erlangen
PMSF (phenylmethylsulfonyl fluoride)	Sigma-Aldrich, Taufkirchen
Potassium sulfate (K ₂ SO ₄)	Roth, Karlsruhe
PromoFectin	Promokine, Germany
Rifampicin	Sigma-Aldrich, Taufkirchen
Salmon sperm DNA	Sigma-Aldrich, Taufkirchen
Sarkosyl (N-lauroylsarcosine, sodium salt)	Sigma-Aldrich, Taufkirchen
SDS (sodium lauryl (dodecyl) sulfate)	Roth, Karlsruhe
Sodium chloride (NaCl)	Roth, Karlsruhe
Sodium hydroxide	Roth, Karlsruhe
Spectinomycin dihydrochloride pentahydrate	Sigma-Aldrich, Taufkirchen
TEMED (N,N,N',N'-tetramethylethylenediamine)	Sigma-Aldrich, Taufkirchen
Tetracycline hydrochloride	Sigma-Aldrich, Taufkirchen
Tricine	Roth, Karlsruhe
Tris-Base	Roth, Karlsruhe
Tris-HCl	Roth, Karlsruhe
Triton X-100	Sigma-Aldrich, Taufkirchen
Trypsin, 0.25% (1 x), liquid	Invitrogen, Karlsruhe

MATERIALS AND METHODS

Name	Source
Tryptone/peptone	Roth, Karlsruhe
Tween-20	Roth, Karlsruhe
UMP (uridine monophosphate)	Sigma-Aldrich, Taufkirchen
Uridine	Sigma-Aldrich, Taufkirchen
X-Gal (5-Bromo-4-chloro-3-indolyl β -D-galactopyranoside)	Sigma-Aldrich, Taufkirchen
Yeast Nitrogen Base	Roth, Karlsruhe
Zinc sulfate (ZnSO ₄)	Roth, Karlsruhe
β -Mercaptoethanol	Roth, Karlsruhe

Tab. 10 Enzymes and antibodies

Name	Source
DNase I, RNase-free	Fermentas, St. Leon-Rot
Exonuclease I	Fermentas, St. Leon-Rot
Gateway® BP Clonase® enzyme mix	Invitrogen, Karlsruhe
Gateway® LR Clonase® enzyme mix	Invitrogen, Karlsruhe
Goat α -GST polyclonal antibody	Rockland, Gilbertsville, USA
Goat α -mouse/HRP polyclonal antibody	Dako, Hamburg
Goat α -rabbit/HRP polyclonal antibody	Dako, Hamburg
Lysozyme, from chicken egg white	Sigma-Aldrich, Taufkirchen
Mouse α -MBP monoclonal antibody	New England Biolabs, Frankfurt
PrimeSTAR HS DNA Polymerase	Takara Bio Inc., Potsdam
Rabbit α -goat/HRP polyclonal antibody	Dako, Hamburg
Rabbit α -HA antibody polyclonal antibody	Santa Cruz Biotechnology, CA, USA
Rabbit α -His ₆ polyclonal antibody	Santa Cruz Biotechnology, CA, USA
Restriction enzymes, various	Promega, Mannheim; Fermentas, St. Leon-Rot
RevertAid™ H Minus Reverse Transcriptase	Fermentas, St. Leon-Rot
RiboLock™ RNase Inhibitor	Fermentas, St. Leon-Rot
SAP (Shrimp Alkaline Phosphatase)	Fermentas, St. Leon-Rot
T4 DNA-ligase	Fermentas, St. Leon-Rot
Taq polymerase	Promega, Mannheim
VentR® DNA Polymerase	New England Biolabs, Frankfurt
Zymolase - Yeast Lytic Enzyme	Zymo Resaerch, Orange, CA

2.1.2 Instruments, consumable materials, and kits

Tab. 11 Instruments

Name	Source
AKTA FPLC™ System	GE Healthcare, München
Agarose gel electrophoresis chamber	Peqlab, Erlangen
Bacteria/yeast incubator	Heraeus, Stuttgart
Bacteria/yeast shaker	Infors, Bottmingen, Switzerland
Biomek 2000 laboratory work station (robot)	Beckman Coulter, USA
Biorad Photometer	Biorad, München
Bioruptor	Diagenode, Liège
Centrifuge for corex tubes, Avanti J-20	Beckman Coulter, USA
Centrifuge, bench top C5403	Eppendorf, Hamburg
Centrifuge, Labofuge 400R	Heraeus, Stuttgart
Centrifuge, micro (Biofuge Pico)	Heraeus, Stuttgart
Centrifuge, stand alone J2-HS	Beckmann, Stuttgart
Centrifuge, table top 5810R (swing out centrifuge)	Eppendorf, Hamburg
Developing machine	Kodak, New Haven, USA
Eagle eye	Stratagene, Heidelberg
ELx808 96-well plate reader	Biotek Instruments, Friedrichshall
FLUOStar OPTIMA fluorescence reader	BMG Labtech, Offenburg
NanoDrop	Peqlab, Erlangen
PAGE electrophoresis chamber	Amersham, Freiburg
PCR thermocycler PTC-200	MJ Research, Waltham, USA
pH-Meter Calimatic 766	Knick, Egelsbach
Semidry blotter	Amersham, Freiburg
Sonifier Cell Disruptor B15	Branson, Boston
Spectrophotometer	Biorad, München
Thermomixer	Eppendorf, Hamburg
Vortex Genie2	Bender und Hobein, Karlsruhe
Zeiss LSM 510 Meta confocal laser scanning microscope	Zeiss, Germany

Tab. 12 Kits and reagents

Name	Source
Bradford reagent	Bio-Rad, UK
PCR clean up innuPREP DOUBLEpure Kit	Analytic Jena bio solutions, Jena
Pierce® ECL Western blotting substrate	Thermo Scientific, Schwerte
Plasmid Maxi Kit	Qiagen, Hilden
QuickLyse Miniprep Kit	Qiagen, Hilden
RNeasy Mini Kit	Qiagen, Hilden

Tab. 13 Consumable materials

Name	Source
96-well microtiter plates, flat bottom	Sarstedt, Nümbrecht
96-well microtiter plates, round bottom	Sarstedt, Nümbrecht
Bottle top sterile filter	Nalge Nunc International, USA
Cellulose MN 300 DEAE for TLC, 20 x 20 cm	Macherey-Nagel, Düren
Conical (falcon) tubes, 50 ml and 15 ml	GBO, Frickenhausen
COREX 8445 tubes, 30 ml	Corning Glass Works, Corning, N. Y.
Dialysis membranes Visking, MWG cutoff 14 kDa	Roth, Karlsruhe
Disposable photometer cuvettes	Brand, Wertheim
Eight-well chamber slides	Ibidi, Martinsried
Eppendorf microcentrifuge tubes, 2 ml and 1.5 ml	Eppendorf, Hamburg
MultiScreenHTS 96-well plates	Millipore, Billerica
OmniTray (single well) plates	Nalge Nunc International, USA
PCR tubes, 8-stripes and 96-format	Eppendorf, Hamburg
Petri dishes, 5 and 10 cm	Greiner, Nürtingen
PVDF membrane for low MWG proteins	Millipore, Billerica
Sterile syringe filter	Schleicher & Schuell, Dassel
Super RX films	Fujifilm Global, Düsseldorf
Superose 12 10/300 column	GE Healthcare, München
Whatman paper	Bender und Hobein, Karlsruhe

2.1.3 Primers

Tab. 14 Common primers

Application: application the oligos were used for: (cPCR) control PCR, (seq) sequencing. **Template:** corresponding PCR template.

Name	Sequence (in 5'→3' direction)	Application	Template
GFPluc_for	GGGCGGAAAGATCGCCGTG	cPCR	pCR3.1-N-eGFPLuc
MBP_for	AGACGCGCAGACTCCCGGT	cPCR	pCR3.1-N-MBP
BGH_rev	CTAGAAGGCACAGTCGAGGCTG	cPCR	pCR3.1-N-eGFPLuc/MBP
GWY_B1_for	GGGGACAAGTTTGTACAAAAAAGCAGGCT	GW cloning	1 st round PCR product
GWY_B2_rev	GGGGACCACTTTGTACAAGAAAGCTGGGT	GW cloning	in Gateway® cloning
M13_for	CGTTGTAAAACGACGGCCAG	cPCR, seq	pDONR221
M13_rev	GCCAGGAAACAGCTATGACC	cPCR, seq	
pDONR207_for	TTAACGCTAGCATGGATCTC	cPCR, seq	pDONR207
pDONR207_rev	CATCAGAGATTTTGAGACAC	cPCR, seq	
pGAD_for	TTTAATACGACTCACTATAGGGCG	cPCR	pGADT7g
pGAD_rev	AGATGGTGCACGATGCACAG	cPCR	
pGBK_for	GTAATACGACTCACTATAGGGCG	cPCR	pGBKT7g
pGBKT7_rev	TTTTCGTTTAAAACCTAAGAGTC	cPCR	
pDEST32_for	GTCAAAGACAGTTGACTGTATCG	cPCR	pDEST32
pDEST32_rev	ACATTTTATGTTAGCTGGTGGAC	cPCR	
pDEST22_for	TATAACGCGTTTGAATCACT	cPCR, seq	pDEST22
pDEST22_rev	AGCCGACAACCTTGATTGGAGAC	cPCR, seq	
pETG_for	GGATCGAGATCTCGATCCCG	cPCR	pETG vectors
pETG_rev	GGGCTTTGTTAGCAGCCGG	cPCR	
pBAD24_for	AGCGGGACCAAAGCCATGAC	cPCR	pBAD24HA
pBAD24_rev	CGGCGCGTACGGCGTTTCAC	cPCR	

Tab. 15 Gene-specific primers

Oligos for ORF cloning and other applications. Gene-specific sequences of oligos used for high-throughput Gateway® cloning are not given. Description of the primer design and 12attB terminal sequences are given under (2.2.1.1).

Name	Sequence (in 5'→3' direction)
b0637_H1+197up	CTCCGCAGCCAGACTTTTTCC
b0637_H2+106down	GCTCGAAGGGCATATCTTTTCGG
lacZ_for	CATGGTACCCACCATGATTACGGATTCACTGGCC
lacZ_rev	GTAAGCTTTTTTTTGACACCAGACCAACTGGTAATG
lacZ_for@1518	CCCGGCTGTGCCGAAATGGT
lacZ_rev@2139	CTGCTGCCAGGCGCTGATGT
phoA_for@465	GGAAATGGCAAAGCCGCAGG
phoA_rev@1294	GTGAATCCTCTTCGGAGTTCCC
C7orf30_for	AAAAGCTGGCTAGCATGGGGCCGGGCGG
C7orf30_rev	GCCGCGCCGCTCACAAAGTTCTGAGCAATGG-3
L14mt_for	CTGGCTAGCATGGCTTTCTTTACTGGGCTC
L14mt_rev	GCCGCGCCGCTCACAAAGTTCTGAGCAATGG
K114_rev	CTGGTGCCAGAGAGATAATTGCCATGAA
TRIF_for	GCAGCAGCAGCAGGGCCGGTAACTCGTGA
TRIF_rev	TGCTGCTGCTGCACCGATAGGCTGCTCGC
b3310a1_for	12attB1-ATGATCCAAGAACAGACTATGCTGAACG
b3310a2_rev	12attB2-GAGTACTTCTGGTGCCAGAGAG

2.1.4 Plasmids

Tab. 16 Plasmids

Plasmid maps can be found in the supplementary information (Fig. 49).

Name	Source	Description	Application
pDEST32	Invitrogen, Karlsruhe	Gateway® compatible, N-terminal Gal4-DBD tag, bacterial gent ^R selection marker, yeast LEU2 selection marker	Y2H
pDEST22		Gateway® compatible, N-terminal Gal4-AD tag, bacterial amp ^R selection marker, yeast TRP1 selection marker	
pGBKT7g	Clontech, USA (Uetz et al., 2006)	Gateway® compatible, N-terminal Gal4-DBD tag, kan ^R selection marker, yeast TRP1 selection marker	
pGADT7g		Gateway® compatible, N-terminal Gal4-AD tag, bacterial amp ^R selection marker, yeast LEU2 selection marker	
pETG-30A	EMBL, Heidelberg	Gateway® compatible, N-terminal GST tag, IPTG inducible, amp ^R selection marker	BL21(DE3) protein expression, pull downs, protein purification
pETG-40K		Gateway® compatible, N-terminal MBP tag, IPTG inducible, kan ^R selection marker	
pNusA	Santhera pharmaceuticals, Liestal, Switzerland	Gateway® compatible, N-terminal NusA-His ₆ -tag, ATC inducible, amp ^R selection marker	
pHGWA	(Busso et al., 2005)	Gateway® compatible, N-terminal His ₆ -tag, IPTG inducible, amp ^R selection marker	
pGEX-4T-1	Amersham, Freiburg	IPTG inducible, amp ^R selection marker, N-terminal GST-tag	
pBAD24HA	(Guzman et al., 1995) (Titz et al., 2007)	C-terminal HA tag, L-arabinose inducible P _{BAD} promoter, amp ^R selection marker, low copy	BW25113 protein expression
pBAD-GFP	(Albano et al., 1998)	L-arabinose inducible P _{BAD} promoter, induction of GFP orf, amp ^R selection marker, low copy	
pCA24N pCA24N-<i>ybeB</i>	(Kitagawa et al., 2005)	N-terminal His ₆ -tag, empty vector or with <i>ybeB</i> ORF, IPTG inducible pT5/lac promoter, cat ^R , high copy	
pCP20	(Cherepanov and Wackernagel, 1995)	Encodes flipase (FLP), flip-out of FRT flanked regions by site-specific recombination, heat sensitive ORI, bacterial selection marker amp ^R and cam ^R .	BW25113 mutant flip-out
pDONR221	Invitrogen, Karlsruhe	Gateway® donor vector for library construction	GW cloning
pDONR207		Gateway® donor vector for library construction	
pcDNA3.1-HA-mCherry	(Diefenbacher et al., 2008)	pcDNA3.1 backbone, C-terminal mCherry tag	protein expression in HeLa cells, localization study
pcDNA3.1(+)-HA-VN	(Roder et al., 2010)	pcDNA3.1 backbone, C-terminal tag of N-terminal Venus fluorescent protein fragment	protein expression HeLain cells, BiFC
pcDNA3.1(+)-HA-VC		pcDNA3.1 backbone, C-terminal tag of C-terminal Venus fluorescent protein fragment	
pECFP-mem	Clontech, USA	Encodes enhanced CFP, labeling of cell membranes	protein expression HeLa cells
pCR3.1-N-eGFPLuc pCR3.1-N-MBP	(Vizoso Pinto et al., 2009)	Both with amp ^R and kan ^R selection marker; GW compatible; N-terminal tags, respectively (see name)	LuMPIS

2.1.5 Strains

Tab. 17 Bacterial strains

Strain	Source	Genotype/description	Application
<i>E. coli</i> DB3.1	Invitrogen, Karlsruhe	F ⁻ gyrA462 endA1 glnV44 Δ(sr1-recA) mcrB mrr hsdS20(r _B ⁻ , m _B ⁻) ara14 galK2 lacY1 proA2 rpsL20(Sm ^r) xyl5 Δleu mt1	Propagation GW pDESTand pDONR
<i>E. coli</i> BL21(DE3)	Stratagene	F ⁻ ompT gal dem lon hsdS _B (r _B ⁻ m _B ⁻) λ(DE3 (lacI lacUV5-T7 gene 1 ind1 sam7 nin5))	Protein expression
<i>E. coli</i> TOP10	Invitrogen, Karlsruhe	F ⁻ mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str ^R) endA1 λ	Propagation various plasmids
<i>E. coli</i> BW25113	(Baba et al., 2006)	lacI ⁺ , rrmB _{T14} , ΔlacZ _{WJ16} , hsdR514, ΔaraBAD _{AH33} , Δrha, BAD _{LD78}	Phenotyping
<i>E. coli</i> BW25113 gene KO mutants		According to <i>E. coli</i> BW25113	

Tab. 18 Yeast strains and human cell lines

Strain	Source	Genotype	Application
AH109 (yeast)	(James et al., 1996)	MAT _a , trp1-901, leu2-3,112, ura3-52, his3-200, Δgal4, Δgal80, LYS2: GAL1 _{UAS} GAL1 _{TATA} -HIS3, GAL2 _{UAS} -GAL2 _{TATA} -ADE2, URA3: MEL1 _{UAS} -MEL1 _{TATA} -lacZ	Y2H
Y187 (yeast)	(Harper et al., 1993)	MAT _α , ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, Δgal4, met-, Δgal80, URA3: GAL1 _{USA} -GAL1 _{TATA} -lacZ	
HeLa (human)	ECACC No. 930210013	Human cervical epithelial adenocarcinoma cells	Protein localization, BiFC
HEK293TT	ECACC No. 85120602	Human embryonic kidney cells	LuMPIS

Tab. 19 Basic PCR templates for gene-specific amplification

Strain	Source	Template sample
<i>E. coli</i> DB3.1 <i>E. coli</i> BW25113 Yeast AH109	This work	Heat inactivated cells (one colony boiled for 10 min at 98°C in 100 μl H ₂ O dist.)
<i>Treponema pallidum</i> subsp. <i>pallidum</i>	Kindly provided by Prof. Timothy Palzkill, Baylor College of Medicine, Houston, USA	Genomic DNA extract
<i>Streptococcus pneumoniae</i> TIGR4	Kindly Provided by Prof. Daniel Nelson, UMBI, MD, USA	Genomic DNA extract
<i>Synechocystis</i> PCC 6803	Kindly Provided by Prof. Tilman Lamparter, KIT, Karlsruhe, Germany	Heat inactivated cells
<i>Homo sapiens</i> cDNA	Kindly Provided by Dr. Olivier Kassel, KIT, Karlsruhe, Germany	cDNA from human HeLa cells
<i>Zea mays</i> cDNA	Kindly Provided by PD Frank Hochholdinger, University Tübingen, Germany	cDNA from 4 d old seedlings, coleoptile nodes
Cp-1 phage Dp-1 phage	Kindly Provided by Prof. Sylvain Moineau, Département de Biochimie et de Microbiologie, Université Laval, Québec	Genomic DNA extract

2.2 Methods

2.2.1 DNA/RNA methods and assays

2.2.1.1 Primer design

High-throughput primer design

For high-throughput oligonucleotide design the Express Primer Tool was used (Yoon et al., 2002) (<http://tools.bio.anl.gov/bioJAVA/jsp/ExpressPrimerTool/>). Gene-specific primer pairs were designed using the following parameters: optimal $T_m=55^\circ\text{C}$, maximum T_m difference= 3°C , T_m range= 3°C , minimum allowable primer length (16 nucleotides), maximum allowable primer length (30 nucleotides). A fasta formatted text file containing the ORF nucleotide sequences of interest was uploaded. Forward primers were all designed with the native ORF start codon. For Gateway® cloning constant terminal regions were added to the gene-specific sequence to attach a 12 bp region of attB attachment sites: forward oligo constant region (12attB1) site (5'-aa aaa gca ggc tta-3'), reverse oligo constant region (12attB2) site (5'-a gaa agc tgg gtg tta-3'). Reverse oligos all contained a stop codon (TAA). This method was used for cloning the Cp-1 and Dp-1 ORFeome libraries. Primers for fragment/domain constructs of ORFs (domains, truncations, etc.) were designed in-frame. Primer pairs that failed a successful design using these parameters were designed as described below.

Standard primer design

Primers were designed with OligoCalc (Kibbe, 2007). Ideally a GC content of 50% was chosen. Gene-specific oligos were chosen from 15 to 33 nucleotides in length with a maximal T_m difference of 1°C per oligo pair. T_m varied from 50 to 65°C depending on the application. Oligos were checked for self-complementation (<http://www.basic.northwestern.edu/biotools/oligocalc.html>).

2.2.1.2 Common protocols for Polymerase Chain Reaction (PCR)

Taq DNA polymerase PCR

Taq DNA polymerase (Promega) from *Thermus aquaticus* was used for control PCR reactions to determine correct insert sizes of cloning products. Per reaction the following setup was used in a total volume of 25 μl :

- 13.75 μl H₂O dist.
- 1 μl dNTP mix (10 mM, each dNTP (Sigma-Aldrich))
- 2 μl forward primer (10 pmol/ μl)
- 2 μl reverse primer (10 pmol/ μl)
- 5 μl 5 x Taq buffer (Promega)
- 1 μl template (plasmid DNA ~1:10 in H₂O dist. or heat-inactivated cells for colony PCR¹)
- 0.25 μl Taq polymerase (Promega, 5 U/ μl)

¹ For colony PCRs single *E. coli* colonies were picked with a pipet tip and transferred to PCR tubes loaded with 100 μl H₂O dist. Samples were boiled for 8 min at 98°C in a thermocycler.

The following cycler protocol was used (annealing temperature was calculated as $T_m - 5^\circ\text{C}$ per oligo pair; extension time was calculated for Taq polymerase with a processivity of 1.7 kb/min and adapted to the expected PCR product size):

Step	Temperature/ $^\circ\text{C}$	Time/s	Cycles
Initial denaturation	95	30	1 x
Denaturation	95	30	30 x
Annealing	variable	30	
Extension	72	variable	
Final extension	72	30	1 x
Storage	8	∞	1 x

PrimeSTAR DNA polymerase PCR

DNA insert amplification for cloning application was done with PrimeSTAR® HS DNA Polymerase (Takara). This enzyme contains a 3'→5' proofreading activity (error rate is estimated with approx. 50 mismatches per one million bases) and has a very high primer efficiency resulting in very specific product amplification. If not specified anywhere else in this section, the DNA templates were used for ORF amplification as given in Tab. 19. The following protocol was used per reaction in a total volume of 50 μl :

- 30.5 μl H₂O dist.
- 4 μl dNTP mix (2.5 mM, each dNTP, Takara)
- 2 μl forward primer (10 pmol/ μl)
- 2 μl reverse primer (10 pmol/ μl)
- 10 μl 5 x PrimeSTAR buffer (Takara)
- 1 μl DNA template (plasmid 1:10 to 1:100 (~1 ng) diluted in H₂O dist., heat inactivated cells, cDNA, or PCR products (Tab. 19))
- 0.5 μl PrimeSTAR polymerase (Takara, 2.5 U/ μl)

The following cycler protocol was used (extension time was calculated for PrimeSTAR polymerase with a processivity of 1.0 kb/min and adapted to the expected PCR product length; given the polymerase's high primer efficiency, a permanent annealing temperature of 55 $^\circ\text{C}$ was used):

Step	Temperature/ $^\circ\text{C}$	Time/s	Cycles
Initial denaturation	94	20	1 x
Denaturation	98	10	30 x (dependent on application)
Annealing	55	15	
Extension	72	variable	
Final extension	72	20	1 x
Storage	8	∞	1 x

2.2.1.3 Gateway® cloning

The Gateway® system (Invitrogen) provides an universal technology to clone DNA sequences for functional analyses and expression in multiple systems. It is based on the site-specific recombination system of bacteriophage λ (Landy, 1989). The λ recombinase is employed to recombine attachment (*att*) sites from different vectors, which leads to the transfer of a DNA insert from one to the other vector. The reaction is conservative (no sequence is added or lost). The advantages are the ligase-independent PCR product cloning, the directed insertion of ORF sequences, the feasibility of C- and N-terminal fusions, parallel transfer into several destination vectors, its high efficiency, and the availability of a large number of vectors and clones for this system. However, the costs for the cloning reaction mixes are high.

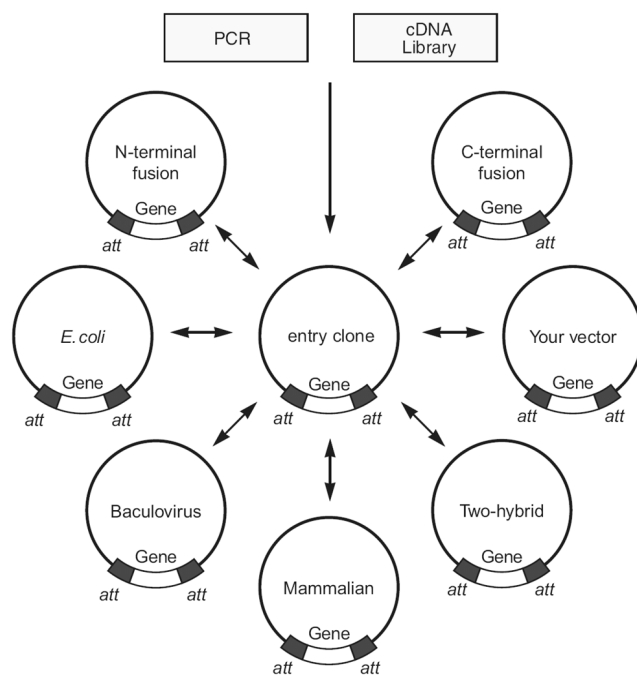


Fig. 13 Gateway® technology

PCR products or cDNA libraries can be recombined into donor vectors. This results in an entry vector (library). The entry vectors can be flexibly recombined with any Gateway® compatible destination vector of interest. The resulting expression plasmids can be applied flexibly. Figure taken from Gateway® manual (Invitrogen).

Gateway® PCR

The first step to achieve a Gateway® clone is the amplification of a PCR product of the ORF of interest that is flanked by *att*B1 and *att*B2 attachment sites. Because of cost issues the PCR was performed in two steps.

Oligos were used that contain 12 bp *att*B1 and *att*B2 sites plus the gene-specific ORF priming site in frame:

- 12*att*B1 forward oligo: 5'-AA AAA GCA GGC TTA-ORF 3'
- 12*att*B2 reverse oligo: 5'-A GAA AGC TGG GTG *TTA*-ORF 3'

Note that the 12*att*B1-site was completed by the 3'-terminal "A" for in-frame cloning resulting in a leucine codon. To the 12*att*B2-site a terminal stop codon was added (see primer sequence in italic).

Primer design was done as described in section (2.2.1.1). In a first-round PCR the ORF of interest was amplified and thereby the 12*att*B sites were attached to the PCR product termini. PCR reactions were

done with PrimeSTAR DNA polymerase (Takara). Because of cost reasons only ½ volumes of all components were used in a total volume of 25 µl as described in (2.2.1.2). Only 8 PCR cycles were used to avoid mutagenic effects.

In a second-round PCR the full attB1 and attB2 sites were attached to the first-round PCR product. PCR reactions were done with PrimeSTAR polymerase (2.2.1.2). 30 PCR cycles were made and the following oligos were used:

- GWY_B1_for 5'-g ggg aca agt ttg tac aAA AAA GCA GGC T-3'
- GWY_B2_rev 5'-ggg gac cac ttt gta caA GAA AGC TGG GT-3'

The priming regions with the corresponding 12attB sequences are highlighted in capital letters.

Afterwards the specificity and correctness of PCR products were analyzed by agarose gel electrophoresis. If no by-products occurred, the PCR product was purified with innuPREP DOUBLEpure Kit (Analytic Jena). If by-products appeared, the whole PCR sample was separated by agarose gel electrophoresis and DNA bands with the correct size were excised and cleaned up again with innuPREP DOUBLEpure Kit.

Gateway® BP reaction

The purified PCR products from the previous steps were used in a BP reaction to obtain an entry clone (Fig. 14). The following protocol was used for the BP reactions:

- 1 µl BP clonase II enzyme mix (Invitrogen)
- 1 µl donor vector DNA (pDONR207 or pDONR221, 150 ng/µl)
- 3 µl purified PCR product DNA (ideally 150 ng/µl)

The samples were incubated for 5 to 16 h at 25°C. Reactions were transformed into chemocompetent *E. coli* TOP10 by heat shock and selected on the appropriate LBA selective medium (for pDONR207 LBA containing 50 µg/ml gentamycine and for pDONR221 LBA with 50 µg/ml kanamycin). Cells were selected o/n at 37°C. Since the *ccdB* product is toxic in TOP10 (blocks gyrase) the non-recombined donor vectors are deselected automatically. To verify the entry clone, a colony PCR was performed and the product size was determined by agarose gel electrophoresis. From positive clones the plasmids were isolated by miniprep plasmid isolation and verified by forward and reverse sequencing (Qiagen). The entry plasmid specific sequencing oligos are given in Tab. 14.



Fig. 14 BP reaction

Recombination of attB-flanked PCR product into a donor vector results in an entry vector and a by-product. Figure taken from Gateway® manual (Invitrogen).

Gateway® LR reaction

The purified entry plasmids were used in a LR reaction to obtain the expression vectors (Fig. 15). The following protocol was used for LR reactions:

- 1 µl LR clonase II enzyme mix (Invitrogen)
- 1 µl of a Gateway compatible destination plasmid(s) DNA (150 ng/µl or 75 ng/µl when two are combined)
- 3 µl purified entry plasmid DNA (ideally 150 ng/µl)

The samples were incubated for 5 to 16 h at 25°C. Reactions were transformed into chemocompetent *E. coli* TOP10 by heat shock and selected on the appropriate LBA selective medium (plasmid-specific selection conditions are described in Tab. 20)

pENTR221 plasmids contain a kan^R resistance and are compatible for recombination with all destination vectors that do not contain this selection marker. pENTR207 contains a gentamycin marker. Here, recombination is compatible with all GW destination vectors that have been used except pDEST32. Reactions were carried out with mixtures of destination vectors to save costs. For instance, pENTR207 plasmids were combined with pGBKT7g and pGADT7g, pENTR221 with pDEST32 and pDEST22, simultaneously. After plasmid transformation into TOP10 aliquots were plated onto appropriate LB selective medium (Tab. 20).

The pool of non-recombined destination vectors is automatically deselected in TOP10 because they still contain *ccdB*. Due to combination of non-overlapping antibiotic selection markers of entry and destination plasmids the pool of non-recombined entry vectors is deselected since selection occurs on medium adequate for the expression vector.

To verify the expression clones, a colony PCR was performed and the product size was determined by agarose gel electrophoresis. From positive clones the plasmids were isolated by a plasmid miniprep. The vector specific oligos are given in Tab. 14.



Fig. 15 LR reaction

Recombination of *attL*-flanked entry plasmid with a destination vector that contains *attR* sites results in an expression vector and a by-product. Figure taken from Gateway manual (Invitrogen).

2.2.1.4 Classical cloning by restriction/ligation

Classical cloning by restriction/ligation was done to obtain the construct pBAD-lacZ-HA and constructs for localization and BiFC studies. Common protocols are given here:

First a PCR was performed with PrimeSTAR polymerase (Takara) according to (2.2.1.2) from corresponding templates (see below). PCR product size was determined by agarose gel electrophoresis and afterwards cleaned up with innuPREP DOUBLEpure Kit (Analytic Jena). Then the PCR product and target vectors were digested with restriction enzymes (see below). Vectors were alternatively

dephosphorylated by adding 1 μ l SAP (Fermentas, 1 U/ μ l) to the digestion reaction. Afterwards PCR products were cleaned up again with innuPREP DOUBLEpure Kit. Digested vectors were separated by agarose gel electrophoresis and then cleaned up with innuPREP DOUBLEpure Kit (Analytic Jena). Then the PCR products were ligated with the linearized vectors in a total volume of 20 μ l:

- 1 μ l T4 DNA ligase (Fermentas, 5 U/ μ l)
- 2 μ l 10 x T4 DNA ligase buffer (Fermentas)
- 300 ng digested PCR product DNA (ideally, 3 x molar excess)
- 100 ng linearized vector DNA
- Fill up to 20 μ l with H₂O dist.

Ligation reaction occurred in a thermocycler for 1 h at 22 °C. Then the reaction was terminated by incubation for 10 min at 65 °C. 5 μ l of the sample were transformed into chemocompetent *E. coli* TOP10 or BW25113. Cells were plated on selective LBA medium and incubated o/n at 37°C. Positive clones were verified by different methods (see below).

Cloning of pBAD-lacZ-HA

lacZ ORF was PCR amplified from genomic DNA of *E. coli* DB3.1 (heat inactivated cells) using the primers *lacZ_for/lacZ_rev* (2.1.3) with Taq polymerase (Promega). Forward primer contained a KpnI restriction and reverse primer a HindIII restriction site. pBAD24HA (Guzman et al., 1995; Titz et al., 2007) and *lacZ* PCR product were co-digested with FastDigest® KpnI and HindIII (Fermentas). Linearized vector was treated additionally with 1 μ l SAP (Fermentas) for 1 h at 37°C. Restriction products were separated by agarose gel electrophoresis and purified with innuPREP DOUBLEpure Kit (Analytic Jena). Linearized vector and PCR product were ligated with T4 DNA ligase (Fermentas) (see above) and then transformed into TSS chemocompetent *E. coli* BW25113. Cells were plated and positives identified by Blue-White Screening on LB agar medium containing 100 μ g/ml ampicillin, 0.2% (w/v) L-arabinose (Sigma) for *lacZ* induction, and 50 μ g/ml X-Gal (Carl Roth). Co-digestion reactions were done as follows:

- 0.5 μ l HindIII (Fermentas, FastDigest®)
- 0.5 μ l KpnI (Fermentas, FastDigest®)
- 2 μ l 10 x FastDigest® Buffer (Fermentas)
- A corresponding volume of ~1 μ g purified *lacZ* PCR product or pBAD24HA DNA
- Fill up with H₂O dist. to 20 μ l

Samples were incubated for 1 h at 37°C. After selection, one positive (blue) colony was used for plasmid isolation. The vector here is named pBAD24-lacZ-HA.

Cloning of localization and BiFC constructs

Human C7orf30 and L14_{mt} full-length open reading frames were amplified with PrimeSTAR® HS DNA Polymerase (Takara Bio Inc). As templates the corresponding pENTR207 entry plasmids were used. C7orf30 was amplified with primer C7orf30_for and C7orf30_rev and L14_{mt} with L14mt_for and L14mt_rev (2.1.3). Forward oligos contained NheI and reverse oligos KpnI restriction sites. Reverse oligos were designed without a stop codon because of C-terminal fusions. The PCR products

were purified with innuPREP DOUBLEpure Kit (Analytic Jena). pcDNA3.1-HA-mCherry (Diefenbacher et al., 2008), pcDNA3.1(+)-HA-VN, pcDNA3.1(+)-HA-VC (Roder et al., 2010) and the PCR products were digested with NheI/NotI with the MULTI-CORE™ buffer system (Promega) and then separated by agarose gel electrophoresis. Hereby the N-terminal HA tag sequence was removed from the source vectors. DNA bands were excised and purified with innuPREP DOUBLEpure Kit. L14_{mt} and C7orf30 PCR fragments were ligated with T4 DNA ligase (Fermentas) into all mentioned linearized plasmids and then transformed into chemocompetent *E. coli* TOP10. Selection followed on LB agar medium with 100 µg/ml ampicillin at 37°C. From single colonies DNA was isolated with a Plasmid Maxi kit (Qiagen). Correctness of clones was verified by NheI/NotI digestion followed by insert size determination by agarose gel electrophoresis. Plasmids were verified by sequencing.

The following co-digestion protocol was used for cloning and for plasmid verification:

- 0.5 µl NheI (Promega, 10 U/µl)
- 0.5 µl KpnI (Promega, 10 U/µl)
- 2 µl 10 x MULTI-CORE™ buffer (Promega)
- Corresponding volumes of 1 µg PCR product or plasmid DNA
- Fill up with H₂O dist. to 20 µl

Co-digestion samples were incubated for 1 h at 37°C and then 5 µl were checked on a agarose gel.

2.2.1.5 Directed mutagenesis of L14 by fusion PCRs

As PCR template the entry vector pENTR207 containing the wild-type sequence of *E. coli* L14 (b3310) was used. Codons for residue K114 and T97-R98-I99-F100 were substituted by alanine codon(s) GCA using mutagenic primers. Therefore, for mutant K(114)A the reverse oligo K114_rev was used and for mutant T97-F100 forward oligo (TRIF_for) and reverse oligo (TRIF_rev) (2.1.3). Mutations were inserted by a PCR reaction (PrimeSTAR® HS DNA Polymerase (Takara Bio Inc)) according to (2.2.1.2). Single steps are illustrated in Fig. 16.

For mutant construct K114A the reverse oligo K114_rev was combined with the wild-type forward oligo b3310a1_for. An initial PCR was made with these oligos (10 cycles) and then attB sites were attached as described in the Gateway cloning protocol by a first and second round PCR (2.2.1.3).

For construction of mutant T97-F100 a fusion PCR was performed. First, the 5'- and 3'-terminal regions of b3310 (L14) were amplified separately by using TRIF_for and TRIF_rev oligos in combination with b3310a1_for and b3310a2_rev, respectively. 30 cycles were performed. Size of PCR products was checked by agarose gel electrophoresis. Next, 0.5 µl of the two reactions were mixed, fused, and PCR amplified for 10 cycles with the terminal Gateway oligos in one reaction followed by PCR amplification with the attB attachment oligos as described in the Gateway cloning protocol by a second round PCR. Entry clones (pDONR207) were obtained by BP reactions and plasmids were verified by sequencing.

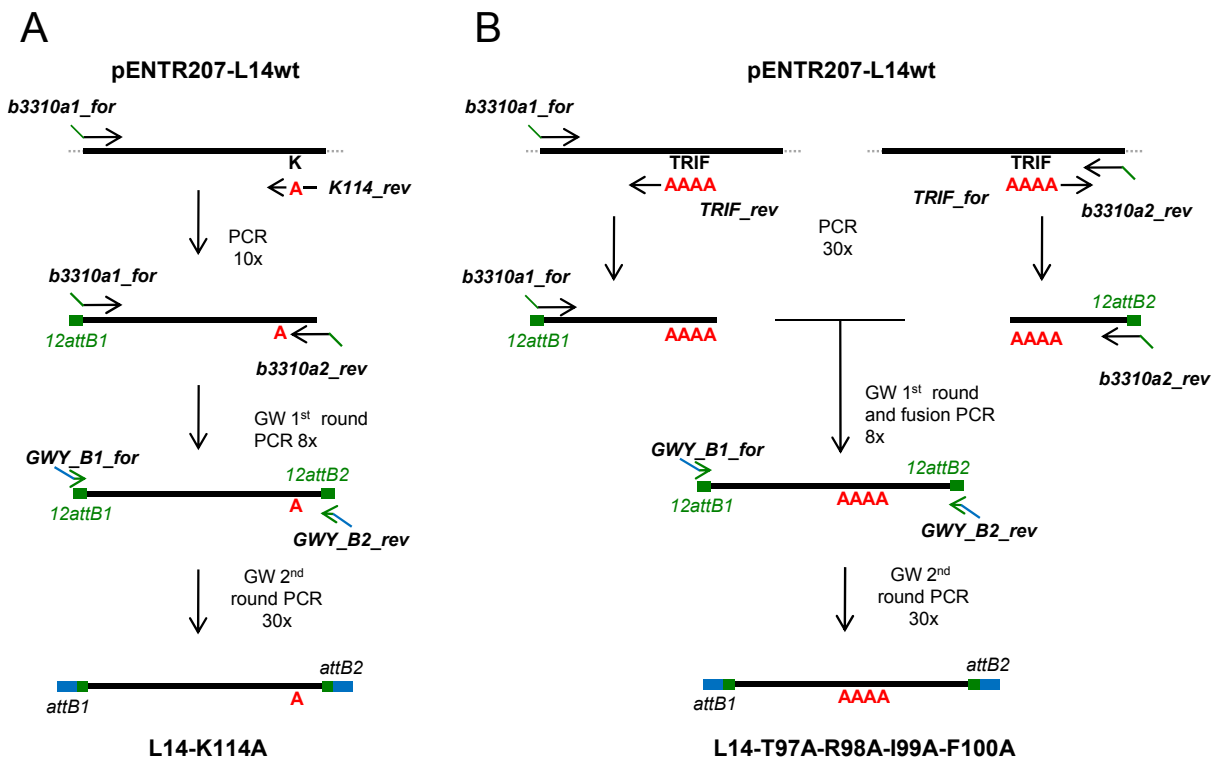


Fig. 16 Illustration of mutagenic fusion PCRs of *E. coli* L14 mutant constructs

DNA is shown as a single strand and codons by the amino acid code. (pENTR207-L14wt) primary PCR template of L14 (b3310) wild-type DNA sequence. Final PCR products contained the flanking attB termini for the BP reaction. **(A)** PCR strategy to mutate K114 codon to an alanine codon. **(B)** PCR strategy to mutate the codon stretch T97- F100 to four continuous alanine codons. The number of PCR cycles for each step is indicated.

2.2.1.6 PCR product clean up

PCR products and linearized plasmids were purified with innuPREP DOUBLEpure Kit (Analytic Jena). Samples were purified by a direct clean up protocol. Alternatively, samples were separated by agarose gel electrophoresis and then DNA bands were excised from the gel and purified according to the manufacturer's instructions.

2.2.1.7 Plasmid isolation and purification

Large quantities of purified plasmids were achieved with a Plasmid Maxi Kit (Qiagen). Low quantity purification of plasmid DNA with designation for sequencing was done with QuickLyse Miniprep Kit (Qiagen). Plasmid isolation was done according to the manufacturer's protocol. For standard low quantity plasmid isolation, an isopropanol precipitation method was applied. Buffer P1, P2, and P3 was used from Qiagen Plasmid Maxi Kit. If not indicated, the protocol step was done at room temperature:

- Pick isolated single bacterial colonies from a LBA plate and inoculate in ~2 ml of LB liquid medium with appropriate antibiotic.
- Incubate o/n with shaking at 37°C.
- Pellet 1 ml of bacterial sample at 13,000 rpm for 1 min in micro centrifuge (Biofuge, Heraeus).

- Discard supernatant and resuspend pellet in 200 μ l buffer P1.
- Add 200 μ l P2, mix by inverting the tube, and incubate for lysis 3 min.
- Add 200 μ l P3, mix by inverting the tube.
- Centrifuge for 10 min at 13,000 rpm in micro centrifuge (Biofuge, Heraeus).
- Transfer 500 μ l to 1.5 ml micro tube.
- Add an equal volume isopropanol and 1/10 volume of a 3 M pH 5.5 sodium acetate solution.
- Vortex for 1 min.
- Centrifuge at 4°C and 13,000 rpm for 20 min in bench top centrifuge (C5403, Eppendorf).
- Discard supernatant and add 500 μ l -20°C cold 80% ethanol.
- Centrifuge at 4°C and 13,000 rpm for 5 min in bench top centrifuge (C5403, Eppendorf).
- Discard supernatant and allow plasmid pellet to dry at room temperature for 1 h or longer.
- Resuspend DNA pellet in 50 μ l autoclaved H₂O dist.

2.2.1.8 Agarose gel electrophoresis

The required amount of agarose (final concentration between 1 to 2 % (w/v)) was dissolved in 1 x TAE buffer and boiled in a microwave until the agarose dissolved. Ethidium bromide was added at a final concentration of 0.3 μ g/ml. The molten gel was poured into a horizontal gel chamber (Peqlab). Combs with the appropriate number and size of the teeth were used to make the loading slots. After the gel had solidified, it was immersed in 1 x TAE buffer, the samples were loaded onto the gel in loading buffer. Gels were run at 50-100 V for ~1 h. Afterwards DNA was visualized by transillumination with 302 nm ultraviolet radiation.

TAE buffer: 40 mM Tris-Acetate, 1 mM EDTA pH 8.0

5 x DNA loading buffer: 50% (v/v) glycerol, 0.2% (w/v) SDS, 0.05% (w/v) bromophenolblue in 1 x TAE buffer

2.2.1.9 RNA isolation

Isolation of total RNA from *E. coli* was done with the RNeasy Mini Kit (Qiagen). This method was used for the analysis of *lacZ* reporter mRNA levels at various time points post induction of *lacZ*. Therefore *E. coli* o/n cultures were grown at 37°C (wild-type BW25113 and flip-out mutant *ΔybeB* harboring plasmid pBAD24-*lacZ*-HA) in 10 ml LB medium with 100 µg/ml ampicillin. Pre-cultures (50 ml LB + ampicillin) were inoculated 1:25 with o/n cultures and shaken for 80 min into logarithmic phase at 37°C. OD600 was determined by photo-spectrometry and cultures were diluted once more to a final OD600=0.2 LB medium (100 µg/ml ampicillin). 2 x ONPG (4 mM) L-arabinose (0.00.4% (w/v)) stock medium was prepared. 2 x stock medium was prepared with 1 x LB medium containing 100 µg/ml ampicillin. To 10 ml cultures (OD600=0.2) 10 ml of 2 x ONPG/inductor stock solution was added. Then cultures were shaken at 37°C in Erlenmeyer flasks. In intervals of 1 h OD600 was determined and a volume saved corresponding to a density of OD600=0.2. Cells were pelleted by centrifugation for 1 min at room temperature with a micro centrifuge (Biofuge, Heraeus) and the supernatant was discarded. Total mRNA from each cell pellet was isolated as follows:

- Prepare TE-lyso buffer (components, see end of the protocol).
- Add 10 µl β-mercaptoethanol to 1 ml buffer RLT (Qiagen). Prepare an appropriate volume.
- Resuspend pellets in 100 µl TE-lyso buffer and incubate at room temperature for 5 min.
- Add 350 µl buffer RLT (prepared with β-mercaptoethanol) (Qiagen). Vortex vigorously.
- Add 250 µl 100% P.A. EtOH. Mix briefly by vortexing.
- Apply whole sample to an RNeasy mini column.
- Centrifuge at room temperature at 13,000 rpm for 15 s with micro centrifuge. Discard filtrate.
- Add 700 µl buffer RW1 (Qiagen) to column and repeat previous step.
- Transfer column to new 2 ml micro tube. Pipet 500 µl buffer RPE into column.
- Repeat previous centrifugation step and discard filtrate.
- Repeat the previous wash step.
- For removal of buffer traces centrifuge another 2 min at 13,000 rpm.
- For elution add 50 µl DEPC H₂O to the column. Centrifuge 1 min at 13,000 rpm.
- Store total RNA samples at -80°C.

TE-lyso buffer(for bacterial lysis): 10 mM Tris-HCl, 1 mM EDTA pH 8.0, 400 µg/ml lysozyme

2.2.1.10 Determination of DNA and RNA concentration

Determination of nucleic acid concentrations (DNA, RNA) from purified total RNA samples, PCR products, plasmids, etc. was done with a NanoDrop (Pepqlab).

2.2.1.11 Reverse transcription of mRNA, gene-specific

The total RNA samples from (2.2.1.9) were applied to reverse transcriptase reactions. Thereby mRNA of *lacZ* reporter mRNA and house-keeping mRNA *phoA* were transcribed with gene-specific primers in a first strand synthesis reaction by the enzyme Reverse Transcriptase into a single DNA strand (reverse complement of mRNA). Therefore, the total RNA concentration of each sample was determined with a Nanodrop. A corresponding volume of 1 µg RNA per sample was transferred into a PCR tube and filled up to a final volume of 10 µl with DEPC water. To remove DNA traces the samples were treated with DNaseI according to the protocol:

- Prepare an appropriate volume of 2 x DNaseI mastermix (volume per sample):
 - 1 µl DNaseI (1 U/µl, Fermentas)
 - 2 µl 10 x DNaseI buffer (Fermentas)
 - 1 µl RNase inhibitor (40 U/µl, RiboLock™ RNase Inhibitor, Fermentas)
 - 6 µl DEPC H₂O
- Transfer 10 µl of 2 x DNaseI mastermix to 10 µl samples and mix.
- Incubate samples for 30 min at 37°C.
- Inactivate DNaseI by adding 2 µl EDTA solution (25 mM, Fermentas); incubate 10 min at 65°C.

Then the RNA samples were applied to the first DNA strand synthesis reaction:

- Transfer 5 µl of DNaseI treated sample into PCR tube.
- Add 4 µl DEPC water and 1 µl of the corresponding gene-specific reverse oligo (10 pmol/µl).
- Incubate 10 min at 70°C and prepare an appropriate volume of 2 x mastermix (volume per sample):
 - 4 µl 5 x reaction buffer (Fermentas)
 - 2 µl dNTP mix (10 mM each dNTP, Sigma)
 - 1 µl RT (200 U/µl, RevertAid™ H Minus Reverse Transcriptase, Fermentas)
 - 0.5 µl RNase inhibitor (40 U/µl, RiboLock™ RNase Inhibitor)
 - 2.5 µl DEPC water
- Add 10 µl of 2 x mastermix to each sample and mix.
- Incubate in a thermocycler using the steps:
 - 60 min at 42°C (synthesis)
 - 10 min at 72°C (inactivation)
- Store first strand DNA samples at -20°C.

Simultaneously, samples were prepared for each RNA isolate according to this protocol but without addition of reverse transcriptase. These samples were used as negative controls in the following PCR reaction to verify the complete removal of DNA traces. 1 µl of first strand samples and negative controls were applied to a standard PCR with Taq polymerase (2.2.1.2) with gene-specific primer pairs for *lacZ* and *phoA*. Primers *lacZ_for@1518* and *lacZ_rev@2139* were used for *lacZ* amplification. *PhoA_for@465* and *phoA_rev@1294* for *phoA* amplification (Tab. 15).

2.2.2 Protein methods and assays

2.2.2.1 Protein expression

The following protocol was used for protein expression in *E. coli* BL21(DE3):

- Inoculate 5 ml selective LB liquid medium with *E. coli* BL21(DE3) strain that was transformed with a certain protein expression plasmid. Shake pre-culture o/n at 37°C.
- Prepare a main culture: inoculate 10 ml (for low-quantity protein expression as done for pull down assays (pETG-30A, pETG-40K, pNusA, pGEX4T-1)) or 500 ml (for high-quantity protein expression as done for protein purification (pHGWA)) selective LB liquid medium with o/n culture 1:100. Shake for 1 h at 37°C.
- Induction: add an appropriate volume of inductor (for IPTG use usually 0.1 mM, for ATC 0.2 µg/ml (f.c.)). In parallel shake a culture w/o inductor. This sample is needed as a negative control for SDS PAGE.
- Shake culture at 37°C for 3 h.
- Transfer culture depending on its volume into a 15 ml conical tube or a 500 ml centrifuge beaker.
- Pellet cells by centrifugation at 4°C for 10 min (4,000 rpm with centrifuge 5810R (Eppendorf) or 6,000 rpm with J2-HS (Beckman, rotor JA-10)). Discard supernatant and store pellet at -80°C.

2.2.2.2 Purification of His₆-tagged proteins

Protein extraction

- Thaw cell pellet on ice.
- Determine weight of cell pellet.
- Chill extraction buffer on ice.
- Per 1 g cell pellet add 5 ml extraction buffer and resuspend pellet by pipetting.
- Transfer suspension to a 50 ml conical tube and incubate for lysis for 30 min on ice.
- Sonicate sample on ice until the lysate becomes clear (Cell Disruptor B15, Branson).
- Add NP-40 (20% (v/v) stock solution) to finally 0.5% and mix by inversion.
- Save 5 µl lysate for SDS PAGE analysis (total fraction).
- Transfer lysate to a 30 ml COREX tube. Centrifuge for 30 min at 4°C and 9,000 rpm (Avanti J-20).
- Decant supernatant into a new 50 ml conical tube.
- Save 5 µl lysate for SDS PAGE analysis (soluble fraction) and apply lysate to the next step.

Extraction buffer: 50 mM Tris-HCl pH8.0, 300 mM NaCl, 50 µg/ml lysozyme, 100 µM PMSF.
Prepare fresh.

His₆-protein isolation (batch protocol)

All centrifugation steps here can be done at room temperature with 500 rpm for 5 min (table top centrifuge 5810R, Eppendorf).

- Add imidazole from a 1 M stock solution to finally 10 mM to the soluble fraction of the protein lysate (this is done to avoid unspecific protein binding to the beads).
- Per lysate use up to 1 ml of a 50% slurry Ni-NTA beads (Ni-NTA Superflow, Qiagen). Volume of slurry depends on the protein band intensity determined by SDS PAGE. If fat protein bands are obvious, use 1 ml. If protein is weakly expressed, use less beads.
- Transfer beads to a 15 ml conical tube and equilibrate beads by washing them twice with 14 ml wash buffer (centrifuge and discard supernatant twice).
- Protein binding: apply protein lysate to beads and roll tubes for 30 min at room temperature.

- Centrifuge beads and discard supernatant by pipetting or decanting. Optionally, save 5 μ l supernatant for SDS PAGE analysis to check binding efficiency.
- Wash beads with 14 ml wash buffer, centrifuge, and discard supernatant. Repeat wash step in total 4 times.
- Elution: add 500 μ l elution buffer and roll the tube at room temperature for 10 min.
- Centrifuge.
- Transfer eluate to 1.5 ml tube and store on ice.
- Repeat the latter three steps twice.
- Save from all eluates 5 μ l for SDS PAGE analysis to examine quality of purified protein sample.

Wash buffer: 50 mM Tris-HCl pH8.0, 300 mM NaCl, 20 mM imidazole.

Elution buffer: 50 mM Tris-HCl pH8.0, 300 mM NaCl, 250 mM imidazole.

Dialysis

To remove imidazole from samples the eluate fractions were dialyzed twice against 2 liter 1 x PBS with 1% v/v glycerol in dialysis membranes (14 kDa MWG cutoff, Visking, Roth). Dialysis was done o/n at 4°C. 100 μ l aliquots were transferred to 1.5 ml tubes, frozen in liquid nitrogen and stored at -80°C.

2.2.2.3 Bradford assay

The protein amount was determined by Bradford assay. This is a colorimetric method and is based on an absorbance shift in the dye coomassie when the previously red form coomassie reagent changes and stabilizes into coomassie blue by the binding of protein. A straight calibration line was determined with BSA (Promega). Therefore 1 μ l of BSA solutions containing 0, 1, 10, 100, and 1,000 ng/ μ l BSA were mixed with 1 ml Bradford reagent (Bio-Rad) and extinction was measured at 595 nm with a Biorad photospectrometer. Determination of protein concentrations from samples was done by diluting 1 μ l sample in 1 ml reagent followed by extinction measurement at 595 nm. Protein amount was calculated from the BSA calibration line equation.

2.2.2.4 Polyacrylamide gel electrophoresis (PAGE)

Tris-glycine SDS-PAGE

Components for mini gels are mixed in the following order (5 ml for one stacking, 10 ml for two separating gels):

Component	Stacking	Separation gel			
	5%	8%	10%	12 %	18%
H ₂ O dd.	6.8 ml	4.6 ml	4 ml	3.3 ml	2.2 ml
30% acrylamide mix	1.7 ml	2.7 ml	3.3 ml	4.0 ml	5.1 ml
1.0 M Tris-HCl pH 6.8	1.25 ml	-	-	-	-
1.5 M Tris-HCl pH 8.8	-	2.5 ml	2.5 ml	2.5 ml	2.5 ml
10% (w/v) SDS	0.1 ml	0.1 ml	0.1 ml	0.1 ml	0.1 ml
20% (w/v) APS	0.05 ml	0.05 ml	0.05 ml	0.05 ml	0.05 ml
TEMED	0.01 ml	0.006 ml	0.004 ml	0.004 ml	0.004 ml

- Pour separating gel (add reagents as indicated above) into gel chamber (so that it fills up 2/3 of the chamber) and cover with isopropanol and let solidify.
- Remove isopropanol and pour stacking gel. Immediately insert the combs and let solidify.
- Insert gel into gel chamber and fill up lower and upper reservoir with 1 x running buffer.
- Add to 7.5 μ l protein sample 2.5 μ l 4 x Laemmli sample buffer, mix, and boil in thermocycler for 5 min at 95°C. Cool down to room temperature and load with pipet into sample pockets.
- Separate by electrophoresis at ~120 V until the dye front leaves the gel.

Solutions required (prepare all solutions with H₂O dd.):

Separating gel buffer (4x)

- 18.17 g tris base
- 4 ml 10% (w/v) SDS
- Adjust pH to 8.8 with HCl
- Add H₂O dd. to 100 ml
- Autoclave

Stacking gel buffer (4x)

- 6.06 g tris base
- 4 ml 10% (w/v) SDS
- Adjust pH to 6.8 with HCl
- Add H₂O dd. to 100 ml
- Autoclave

20% APS solution

- 20% (w/v) APS
- Store at -20°C

4 x SDS sample buffer (Laemmli buffer)

- 250 mM Tris pH 6.8
- 12.5% (w/v) SDS
- 40% (v/v) glycerol
- Few crystals bromphenolblue
- 20% (v/v) β -mercaptoethanol
- Store at -20°C

SDS-PAGE running buffer

- 30 g tris base
- 144 g glycine
- 100 ml 10% (w/v) SDS
- Add H₂O dd. to 1 liter

Tricine-PAGE (gradient gel)

Components are mixed in the order shown below (prepare a total volume of 3 ml each for 5% and 15% component):

Component	5% gel	15% gel
30% acrylamide mix	0.45 ml	1.35 ml
1 x gel buffer	1 ml	1 ml
glycerol	0.31 ml	0.31 ml
H ₂ O dd.	1.223 ml	0.324 ml
20% (w/v) APS	0.015 ml	0.015 ml
TEMED	0.0024 ml	0.0012 ml

- Place a comb between gel plates.
- Fill each gel mix into corresponding chambers of a gradient mixer (self-made) and transfer mix between glass plates under stirring. Let gel solidify.
- Protein samples can be prepared with Laemmli sample buffer as described above.
- Place gel in gel chamber. Fill upper reservoir with 1 x cathode buffer, lower reservoir with 1 x anode buffer.
- Separate samples in electric field with ~120 V until the dye front leaves the gel.

Solutions required (prepare all with H₂O dd.):

5 x cathode buffer

- 121.1 g tris base
- 179.2 g tricine
- Dissolve in 800 ml H₂O dd.
- Adjust to pH 8.25 with HCl
- Fill up to 1 liter with in H₂O dd.
- Store at room temperature

1 x gel buffer

- 1.5 g SDS
- 181.65 g tris base
- Dissolve in 300 ml H₂O dd.
- Fill up to 500 ml H₂O dd.
- Adjust to pH 8.45 with HCl
- Store at 4°C

10 x anode buffer

- Dissolve 242.2 g tris base in 800 ml H₂O.
- Adjust to pH 8.9 with HCl
- Fill up to 1 liter with H₂O dd.
- Store at 4°C

2.2.2.5 Coomassie Blue staining of protein gels

- Transfer polyacrylamide gel into a plastic tray, cover with staining solution, and shake for 45 min.
- For destaining, the gel is boiled in water in a microwave oven. Change water a couple of times.

Staining Solution: 50% (v/v) methanol, 10% acetic acid (v/v),
0.2% (w/v) Coomassie Brilliant Blue R250

2.2.2.6 Western Blot (semi-dry)

- Cut blot paper (8 sheets of Whaman paper per gel) to the size of the gel and equilibrate in transfer buffer.
- Cut PVDF blotting membrane to the size of gel and activate in methanol for 2 min. Then equilibrate in transfer buffer.
- Build up blot (Semi-dry blotter, Amersham): place in this order: 4 x filter paper, membrane, gel, 4 x filter paper. If necessary, roll out air bubbles with a conical tube.
- Run 1 mini gel at 110 mA for 1 h.
- Discard filter paper and briefly wash blot in 1 x PBS to remove SDS.

Transfer buffer: 3.03 g tris base, 14.27 g glycerol, 200 ml methanol, add H₂O dd. to 1 liter.

2.2.2.7 Protein immunodetection and ECL (enhanced chemiluminescence)

- Wash PVDF membrane twice for 10 min in 1 x PBS.
- Blocking: shake membrane for 45 min in ~30 ml blocking buffer.
- Transfer membrane onto a plastic foil and seal by heat-sealing.
- Apply to 6 ml (per membrane) primary antibody solution and invert at room temperature for 1 h.
- Wash membrane three times in ~30 ml wash buffer for 10 min.
- Transfer membrane onto a plastic foil and seal by heat-sealing.
- Apply to 6 ml (per membrane) secondary antibody solution (secondary antibody must be conjugated with HRP for ECL detection) and invert at room temperature for 1 h.
- Wash membrane three times in ~30 ml wash buffer for 10 min.
- ECL: per membrane mix 500 µl ECL solution I with 500 µl ECL solution II (ECL (Pierce® ECL Western blotting substrate, Thermo Scientific)).
- Transfer membrane into a cassette on a thin, transparent plastic foil.
- Wipe off wash buffer from membrane using a tissue.
- Pipet ECL mix onto the membrane, disperse allover over membrane, and incubate for 1 min in the dark.
- Wipe off remaining ECL mix from membrane using a tissue.
- Cover membrane with a thin, transparent plastic foil.
- Expose membrane in a darkroom for a defined period with a piece of film (Super RX films were used, Fuji) and develop film (Developing machine, Kodak).

Solutions required:

Wash buffer

- 1 x PBS
- 0.2 % (v/v) Tween-20

Blocking buffer

- 1 x PBS
- 5% (w/v) milk powder

Primary and secondary antibody solution

- Blocking buffer
- 0.2 % (v/v) Tween-20
- Primary or secondary antibody (see below)
- Use per membrane/mini gel a total volume of 6 ml

2.2.2.8 Membrane stripping

- Wash PVDF membrane after ECL detection three times in ~30 ml PBS to remove ECL substrate.
- Shake for 30 min at room temperature in ~20 ml stripping solution A.
- Shake for 30 min at room temperature in ~20 ml stripping solution B.
- Repeat the latter two steps twice.
- Wash membrane three times in ~30 ml PBS and apply to a second round of protein immunodetection.

Solutions required:

Stripping solution A

- 50% (v/v) EtOH
- 10% (v/v) sodium acetate

Stripping solution B

- 50 mM Tris-HCl pH8.0
- 10% (w/v) SDS
- 10 mM DTT

2.2.2.9 Protein cross-linking with glutaraldehyde

Chemical cross-linking of proteins followed by SDS-PAGE analysis can be used to get information about the condition and organization of components of interacting proteins and protein complexes. Determination of the complex size on a protein gel gives information, e.g., which quaternary structure a protein complex consists of. Here, chemical cross-linking of proteins with glutaraldehyde was used. Glutaraldehyde cross-links primary amino residues, in the case of proteins lysine residues and the N-terminal amino group of the initiator methionine residue:



Glutaraldehyde has a short spacer arm length of 5Å. For glutaraldehyde cross-linking the purified proteins have to be dissolved in PBS since buffers like TBS contain Tris that consists of amine residues that interfere with the chemical reaction. Thus amino-free phosphate buffer conditions are suitable.

The following protocol was used for chemical protein cross-linking of His₆-tagged Udk (SP1208) from *S. pneumoniae* and Cp-1 Cpl1 (lysozyme):

- Thaw a protein aliquot on ice.
- Centrifuge at 4°C and 13,000 rpm (bench top centrifuge C5403, Eppendorf).
- Determine protein concentration by Bradford assay.
- Transfer 25 or 50 µg of protein to a PCR tube (different proteins solely or as mixed samples).
- Fill up with PBS to 97.5 µl and mix.
- Pre-incubate ½ h at room temperature.
- Add to each sample (but not to loading control sample) 2.5 µl 25% glutaraldehyde stock solution (0.1% f.c.) and mix.
- Incubate for 4 min at room temperature.
- Terminate reaction by adding 10 µl of 1 M Tris-HCl pH 8.0 and mix.
- Add a corresponding volume of 4 x Laemmli sample buffer.
- Boil samples for 5 min at 95°C and let them cool down at room temperature.
- Load 50 µl on a 5% to 15% tricine-PAGE gradient gel and separate by electrophoresis.
- Stain proteins by Coomassie Blue staining.

2.2.2.10 Uridine kinase (Udk) enzyme assay and thin layer chromatography (TLC)

Udk enzyme assay

This protocol describes a competitive Udk (SP1208) enzyme assay to qualitatively determine the influence of the Cp-1 phage protein interaction partner Cpl1 (lysozyme) on Udk activity. The assay was carried out under the optimal enzyme conditions that were determined for the human uridine-cytidine kinase-ortholog (Ahmed and Welch, 1979).

- Thaw a protein aliquot on ice.
- Centrifuge at 4°C and 13,000 rpm (bench top centrifuge C5403, Eppendorf) to remove putative precipitates.
- Determine protein concentration by Bradford assay.
- Transfer per sample 20 µl 5 x sample buffer to 1.5 ml microcentrifuge tube.
- Transfer a corresponding volume of H₂O dd. that is needed for a total sample volume of 100 µl.
- Transfer corresponding volumes of purified proteins at last (negative control w/o proteins). (For detailed information of protein concentrations that have been used, refer to results section (3.1.5)).
- Mix and pre-incubate samples at 37°C for 30 min.
- Add per sample 4 µl of substrate solution and mix.
- Incubate at 37°C and transfer after 5, 10, and 30 min 30 µl per sample into a microcentrifuge tube.
- Incubate sample aliquots immediately for reaction inactivation at 95°C for 10 min and store on ice.
- Precipitate denatured protein by centrifugation for 10 min at 4°C and 13,000 rpm.
- Transfer 10 µl per sample to PCR tubes and apply to thin layer chromatography.

Solutions required (prepare in H₂O dd.):

5 x reaction buffer

- 1 M Tris HCl pH 7.5
- 7.5 mM MgCl₂
- 50 mM DTT, add always fresh

Substrate solution

- 500 mM ATP
- 500 mM uridine

Thin layer chromatography (TLC)

To identify the reaction educts (ATP and uridine) and products (ADP and UMP) a reference experiment has to be done in advance by separating nucleotide samples by TLC solely (load per nucleotide/nucleoside a corresponding volume of 200 µmol) and proceed as described below. Calculate retention factor $R_f = a/b$ (a= running distance of nucleoside or nucleoside; b= running distance mobile phase). R_f can be used to identify the nucleotides and nucleosides in the assay:

- Define start line and label it on DEAE plate with a pencil on the bottom (~2.5 cm above end).
- Load all samples (10 µl/sample) as 1.5 cm long lines with a glass capillary on DEAE cellulose plate (Cellulose MN 300 DEAE, 20 x 20 cm plate, Macherey-Nagel, Düren).
- Let plate dry on air for 1 h.
- Use 0.03 N HCl as mobile phase. Fill HCl in chromatography chamber until the bottom is covered ~1 cm.
- Transfer TLC plate to chamber and close with lid.
- Stop TLC when running front of the mobile phase is close to the plate top. Label running front for determination of R_f .
- Dry plate o/n on air.
- Visualize nucleotides and nucleosides under UV light at 254 nm. Label spots with pencil and calculate R_f .

2.2.2.11 Protein gel filtration

Purified His₆-tagged Udk from *S. pneumoniae* and Cp-1 lysozyme (Cp11) were applied to a protein gel filtration assay by FPLC (fast protein liquid chromatography). Therefore, a Superose 12 10/300 column (GE Healthcare) was used. The AKTA FPLC™ System (GE Healthcare) contained all the pumps, conductivity, pH, and absorbance instrumentation in-line. Protein absorbance was measured at 280 nm. As buffer system 1 x PBS pH 7.4 was used. The column was run at 0.8 ml/min and 0.8 ml fractions were collected. 100 µg Udk and Cp11 were loaded solely or combined with 100 µg of each protein. For calibration the following gel filtration standards (Biorad) were used and run under identical conditions: thyroglobulin (670 kDa), gamma-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B12 (1.35 kDa).

2.2.2.12 Pull downs

This protocol describes the pull down assays for verification and identification of protein-protein interactions among *E. coli* YbeB-L14, human C7orf30-L14mt, and *Zea mays* Iojap-RPL14 test pairs:

Entry plasmids were recombined with pNusA (Santhera Pharmaceuticals, Liestal, Switzerland), pETG-40A, or pETG-30A (EMBL, Heidelberg, Germany) by Gateway® LR reactions and transformed or co-transformed into chemocompetent *E. coli* BL21(DE3). Specification of tested construct pairs is described in the results section. A 20 ml main culture (selective LB) was inoculated at a ratio 1:100 with an o/n grown pre-culture. The main culture was shaken for 1 h at 37°C and then induced with 100 µM IPTG (isopropyl β-D-1-thiogalactopyranoside) for pETG plasmids or 0.2 µg/ml ATC (anhydrotetracycline, Sigma) for pNusA expression plasmids. The cultures were shaken at 37°C for 3 h. Cells were harvested by centrifugation at 4°C and supernatant was discarded. Cell pellets were treated with 500 µl lysis buffer (50 mM Tris-HCL pH 8.0, 100 mM NaCl, 50 µg/ml chicken egg white lysozyme (Sigma), 50 µM PMSF (Sigma)) and incubated for 30 min on ice. Then a corresponding volume of a 10%/10% (v/v) sarcosyl/Triton-X 100 solution was added to a final concentration of 0.1%/0.1%. Samples were sonicated for 8 min (Bioruptor, Diagenode) and then centrifuged for 15 min > 9,000 g at 4°C. Supernatant was saved and 10 µl samples were analyzed for protein expression on a 12% SDS PAGE gel stained with Coomassie Blue R-250 (Sigma). Lysates were stored at -80°C. For pull down experiments total protein amounts of the soluble fractions were determined by Bradford assay (Bio-Rad). For co-expressed proteins 50 µg of soluble protein was applied to beads and 50 µg of soluble protein was saved as input control. For separately expressed proteins 25 µg of each soluble protein fraction were mixed and then applied to the beads. GST fusions were purified in combination with MBP preys via 20 µl glutathione beads (GE Healthcare), and MBP baits in combination with NusA-tagged preys via 20 µl amylose beads (New England Biolabs). Beads were first equilibrated twice with 1 ml wash buffer (50 mM Tris-HCL pH 8.0, 100 mM NaCl, 0.1%/0.1% (v/v) sarcosyl/Triton-X 100). Binding occurred at room temperature for 30 min. Beads were washed four times in 1 ml wash

buffer and precipitated by centrifugation (500 rpm, micro centrifuge). Buffer was discarded and beads were boiled for 5 min in 50 μ l 1 x Laemmli buffer. 10 μ l of output samples and a corresponding volume of input samples were loaded and separated on a 12 % SDS PAGE gel. Proteins were transferred onto a PVDF membrane (Amersham) by a semi-dry Western blot. Membranes were blocked for 1 h with blocking buffer (5% (w/v) low-fat milk powder in PBS pH 7.4) and then applied to immunodetection of recombinant proteins. GST fusions were probed with polyclonal goat α -GST (1:10,000) (Rockland), MBP fusions with monoclonal mouse α -MBP (1:20,000) (New England Biolabs), and NusA fusions containing an additional His₆ tag with polyclonal rabbit α -His6 (1:5,000) (Santa Cruz Biotechnology) immunoglobulins in 6 ml blocking buffer with 0.2% (v/v) Tween-20. Incubations were done for 1 h at room temperature. Then membranes were washed three times in washing buffer (PBS pH 7.4 with 0.2% (v/v) Tween-20) and applied to secondary antibody binding. Polyclonal rabbit α -goat/HRP (1:5,000), goat α -rabbit/HRP (1:2,500) or goat α -mouse/HRP (1:10,000) (Dako) conjugates were used as secondary antibodies (6 ml/membrane) and incubated at room temperature for 1 h. Finally membranes were washed three times in washing buffer and applied to ECL (Pierce® ECL Western blotting substrate, Thermo Scientific). Super RX films were used (Fuji).

2.2.3 Bacterial methods and assays

2.2.3.1 Bacterial media

LB (lysogeny broth) medium

Per 1 liter medium use:

- 0.5% (w/v) Bacto Yeast Extract
- 1% (w/v) Tryptone/Peptone
- 1% (w/v) NaCl

The medium was autoclaved and optionally antibiotics were added to the medium after it was cooled down to 50°C. The same recipe was used for the LB agar medium (LBA). In addition 2% (w/v) agar-agar was added before autoclaving.

MOPS minimal medium

MOPS medium is a minimal medium that contains all essential micronutrients necessary for *E. coli* growth (Neidhardt et al., 1974). Due to the application various carbon sources can be added to it. MOPS medium was prepared as follows:

Final 1 x MOPS medium contains:

- 10 x MOPS mixture 100 ml
 - 0.132 M K_2HPO_4 10 ml
 - H_2O dd. 880 ml
- Mix ingredients above and adjust pH to 7.2 with 10 M NaOH.
 - Filter sterilize.
 - Before use add 10 ml 100 x carbon source (e.g., 20% (w/v) sterile glucose solution)

Preparation of 10 x MOPS mixture:

- Add the following components to ~300 ml H_2O dd. in a 1 liter beaker glass with a stir bar:
 - MOPS 83.72 g
 - Tricine 7.17 g
- Add 10 M KOH to a final pH of 7.4 (10 to 20 ml).
- Bring total volume to 440 ml.
- Make a fresh 0.01 M $FeSO_4 \cdot 7H_2O$ solution (10 ml) and add it to the MOPS/Tricine solution.
- Add the following solutions to the MOPS/tricine/ $FeSO_4$ solution (see below how to make each of these). Mix following solutions:

▪ 1.9 M NH_4Cl 50 ml	▪ 0.276 M K_2SO_4 10 ml
▪ 0.02 M $CaCl_2 \cdot 2H_2O$ 0.25 ml	▪ 2.5 M $MgCl_2$ 2.1 ml
▪ Micronutrient stock 0.2 ml	▪ 5 M NaCl 100 ml
▪ Autoclaved H_2O dd. 387 ml	
- Filter sterilize.
- Aliquot into sterile 50 ml conical tubes and freeze at $-20^\circ C$.

Stock solutions used in 10 x MOPS mixture (store at room temperature):

- | | |
|---------------------|-------------------------------|
| ▪ NH_4Cl 1.9 M | ▪ $MgCl_2$ 2.5 M |
| ▪ K_2SO_4 0.276 M | ▪ $CaCl_2 \cdot 2H_2O$ 0.02 M |
| ▪ NaCl 5 M | |

Micronutrient stock (for 50 ml)

- Mix everything together in 40 ml autoclaved H_2O dd., bring up total volume to 50 ml. Store at room temperature.

▪ ammonium molybdate 0.009 g	▪ $CuSO_4$ 0.006 g
▪ boric acid 0.062 g	▪ $ZnSO_4$ 0.007 g
▪ $CoCl_2$ 0.018 g	▪ $MnCl_2$ 0.040 g

Potassium phosphate K_2HPO_4 solution:

- K_2HPO_4 0.132 M

2.2.3.2 Preparation of chemocompetent *E. coli* cells

CaCl₂-method

Inoculate 150 ml LB medium with 1 ml bacterial o/n culture. Incubate at 37°C to OD₆₀₀=0.35 and cool bacterial culture on ice for 30 min. Centrifuge for 10 min at 6,000 rpm (Beckmann J2-HS) at 4°C and discard supernatant. Wash cells with 30 ml ice-cold 100 mM CaCl₂, centrifuge, resuspend in 30 ml ice-cold 100 mM CaCl₂, and incubate for 1 h on ice. Centrifuge again and resuspend in 3 ml ice-cold 100 mM CaCl₂ 10% (v/v) glycerol solution. Freeze 100 aliquots in liquid nitrogen and store at -80°C. Proceed with transformation (2.2.3.3).

TSS-method

Inoculate a bacterial strain from an o/n culture 1:100 in 50 ml LB medium (with antibiotic(s) for transformed strains for co-transformation) and shake at 37°C to OD₆₀₀ = 0.3-0.4. Transfer culture to 50 ml conical tube and incubate for 1 h on ice. Centrifuge culture (4,000 rpm for 20 min at 4°C (table top centrifuge 5810R, Eppendorf)). Resuspend in 1/10 volume of ice-cold TSS buffer (transformation and storage solution). Freeze 100 µl aliquots in liquid nitrogen and store at -80°C. Proceed with transformation (2.2.3.3).

TSS buffer: 5 ml DMSO, 10 g PEG 6000 (Sigma), 1 M MgCl₂, 90 ml LB medium, sterile filter.

2.2.3.3 Transformation of chemocompetent cells by heat shock

- Remove a tube of TSS- or CaCl₂-competent cells from -80 °C freezer.
- Thaw the cells on ice.
- Add the plasmids (usually ~100 ng), swirl to mix and incubate on ice for 20 min.
- Heat shock: transfer the tubes into a 42°C water bath or heat block for 60 s.
- Rapidly transfer the tubes to ice for 1 to 2 min.
- Add 900 µl LB medium (without antibiotic) and shake the cultures at 37°C for 1 h.
- Centrifuge the cells at 6,000 rpm for 2 min in micro centrifuge.
- Discard 800 µl supernatant and resuspend by pipetting pellet in the remaining 200 µl.
- Transfer the suspension onto a LBA plate with appropriate antibiotic.
- Add a couple of autoclaved glass beads and spread suspension by shaking the plates.
- Remove beads and incubate plates o/n at 37°C.

2.2.3.4 Selection conditions for plasmid-transformed *E. coli* strains

For selection of plasmid-transformed *E. coli* strains antibiotics were added to liquid or agar LB-medium as given in the following table. Propagation of Gateway® compatible destination and donor plasmids was done in media with the corresponding bacterial selection marker plus 34 µg/ml chloramphenicol (positive selection of the Gateway® cassette) in *ccdB* resistant *E. coli* strain DB3.1. All other plasmids were propagated in *E. coli* TOP10. General selection temperature was 37°C except for pCP20 that contains a heat sensitive *ori*. This plasmid was propagated at 30°C. Co-selection, e.g.,

for protein co-expression occurred through selecting towards corresponding antibiotics simultaneously.

Tab. 20 Plasmid selection conditions in *E. coli*

Gateway® compatible destination and donor plasmids are labeled by an asterisk.

Plasmid	Antibiotic	C (µg/ml)	Plasmid	Antibiotic	C (µg/ml)
pDEST32*	Gentamycine	50	pBAD24HA	Ampicillin	100
pDEST22*	Ampicillin	100	pBAD-GFP	Ampicillin	100
pGBKT7g*	Kanamycin	50	pCP20	Chloramphenicol	170
pGADT7g*	Ampicillin	100	pDONR221*	Kanamycin	50
pETG-30A*	Ampicillin	100	pDONR207*	Gentamycine	50
pETG-40K*	Kanamycin	50	pcDNA3.1-HA-mCherry	Ampicillin	100
pNusA*	Ampicillin	100	pcDNA3.1(+)-HA-VN	Ampicillin	100
pHGWA*	Ampicillin	100	pcDNA3.1(+)-HA-VC	Ampicillin	100
pGEX-4T-1	Ampicillin	100	pECFP-mem	Kanamycin	50
pCR3.1-N-MBP*	Ampicillin	100	pCR3.1-N-eGFPLuc*	Ampicillin	100
pCA24N	Chloramphenicol	30			

2.2.3.5 Flip-out of kanamycin cassette from *E. coli* gene deletion mutant

To remove the kanamycin cassette from a *E. coli* $\Delta ybeB$ gene deletion mutant (KEIO collection, (Baba et al., 2006)) plasmid pCP20 (Cherepanov and Wackernagel, 1995) was transformed into $\Delta ybeB$ TSS competent cells by a heat shock method. Because of thermo-sensitive replication of the plasmid the culture was incubated after heat shock for 1 h at 30°C. Then cells were plated on LBA plates containing chloramphenicol and incubated o/n at 30°C. Colonies were picked and the success of flip-out was checked by a genotyping colony PCR using an oligo pair that flanks *ybeB* ORF (primer b0637_H1+197up and b0637_H2+106down). A positive colony was streaked on a non-selective LBA plate. For pCP20 elimination the plate was incubated o/n at 43°C. To check successful removal of pCP20 and kanamycin marker the strain was finally streaked on LBA medium containing ampicillin, kanamycin, or chloramphenicol and incubated o/n at 37°C. No colony growth was observable.

2.2.3.6 Determination of OD600

Standard measurements of OD600 (optical density, light absorption at 600 nm) of bacterial cultures were determined with a Biorad photospectrometer with 1 cm disposable cuvettes. 96-well OD600 was measured with ELx808 96-well plate reader (Biotek Instruments) at 595 nm. 150 µl samples were loaded per well. For standardization OD600 was recalculated with the formula:

$$OD600 = 4.3062 \cdot OD595 - 0.0119$$

For determination of the standardization formula reference measurements were done with the Biorad photometer (1 cm light path (*OD600*) and ELx808 (150 µl/well (*OD595*)) with serial dilutions of a bacterial suspension. The formula was calculated with MS Excel.

2.2.3.7 *AybeB* phenotyping growth curve assays

75 µl per well of 2 x concentrated antibiotic stock solutions in 1 x liquid LB medium were transferred into a 96-well plate (flat bottom). Bacterial strains were grown o/n in LB medium and then OD600 was determined from 1:10 dilutions by photo-spectrometry. Cultures were diluted with LB medium to OD600=0.02 and then 75 µl were added to the 2 x antibiotic stock solutions. Incubation followed at 37°C. In intervals of 1 h sample plates were measured at 595 nm (ELx808, Biotek Instruments, Friedrichshall) and OD600 was recalculated as described (2.2.3.6). Final concentrations of chemical components have been used as given in the following table:

Tab. 21 Chemical components for phenotyping assays

Component	Final concentrations
Chloramphenicol	0, 1, 2, 4 µg/ml
Gentamycine	0, 1, 2, 4 µg/ml
Rifampicin	0, 2, 4, 8 µg/ml
Spectinomycin	0, 1, 2, 4 µg/ml
Tetracycline	0, 1, 2, 4 µg/ml

2.2.3.8 *AybeB* phenotyping agar plate assays

OD600 was determined from *AybeB* and wild-type BW25113 1:10 diluted o/n cultures grown in LB liquid medium. Cultures were serially diluted in LB liquid medium to final OD600 of 0.1, 0.01, 0.001, and 0.0001. Then, 5 µl per dilution was plated on non-selective LB agar test plates. Plates were incubated at 42°C (high temperature condition, standard LB agar medium) or 37°C (high salt condition; LB agar medium with 2, 4, and 8 % (w/v) NaCl) o/n.

2.2.3.9 Competitive fitness assay of *ybeB* gene deletion mutant

Minimal MOPS medium was prepared as described in (2.2.3.1). Sterile filtered glucose was added from a 20% (w/v) stock solution to finally 0.2% (medium named here as MOPS+G). *AybeB* and wild-type BW25113 o/n cultures were grown at 37°C in MOPS+G, diluted 1:10, and OD600 was determined. Cultures were diluted with MOPS+G to a final OD600=0.1 in 50 ml. These samples were used for the reference growth curve. The cultures were shaken in Erlenmeyer flasks at room temperature and the OD600 was monitored hourly. Furthermore cultures were diluted to finally OD600=5e-8 (both strains mixed in the same sample) in 50 ml MOPS+G. Sample cultures were prepared three times independently and then shaken for 72 h at room temperature in Erlenmeyer flasks. From each of the three samples, approx. 500 cells were plated on non-selective LB agar plates. Bacteria were grown o/n at 37°C. These samples were applied to a genotyping colony PCR to distinguish between wild-type and *AybeB* colonies. 3 x 93 colonies were randomly picked from each sample plate and boiled in 50 µl H₂O dist. for 10 min at 95°C. Then a genotyping colony PCR was

performed using the primer pair that has been used for *ΔybeB* mutant verification (2.2.3.5). PCR products were analyzed by agarose gel electrophoresis on a 2% agarose gel.

2.2.3.10 ONPG reporter assay of *ybeB* gene deletion mutant

TSS-competent *ΔybeB* and wild-type BW25113 were transformed with pBAD24-lacZ-HA and selected on LB agar medium with 100 μg/ml ampicillin. Both were grown as o/n cultures in LB liquid medium with 100 μg/ml ampicillin. Pre-cultures (50 ml LB + ampicillin) were inoculated 1:25 with o/n cultures and shaken for 80 min into logarithmic phase. OD600 was determined by photo-spectrometry and cultures were diluted once more to an OD600=0.2. 2 x ONPG (4 mM) L-arabinose (0, 0.004 or 0.4 % (w/v)) stock solutions were prepared with 1 x LB medium + 100μg/ml ampicillin. To 10 ml cultures (OD600=0.2) 10 ml of 2 x ONPG/inductor stock solutions were added. Cultures were shaken at 37°C in Erlenmeyer flasks. Every hour 1 ml suspension was saved. Cells were pelleted by centrifugation. Supernatant was saved for ONPG turnover measurement. 200 μl were loaded per sample on a 96-well plate (flat bottom) and extinction was measured at 415 nm (ELx808, Biotek Instruments, Friedrichshall). Cell pellets were washed in 1 ml PBS. The supernatant was discarded and pellets stored at -80°C. These samples were used for detection of the β-galactosidase level (see below). In parallel to ONPG containing samples a reference growth curve (OD600 determined by photo-spectrometry) was measured from *ΔybeB* and WT, both transformed with pBAD24-lacZ-HA. Here, 10 ml OD600=0.2 cultures from above were diluted at a ratio of 1:1 with corresponding 2 x L-arabinose stock solutions but without ONPG since ONPG treated samples interfered with OD600 measurement.

2.2.3.11 Rescue assay

Chemocompetent BW25113 WT and *ΔybeB* cells, harboring pBAD24-lacZ-HA, were transformed with pCA24N empty vector or pCA24N containing the *ybeB* ORF (Kitagawa et al., 2005) and selected on LB agar medium (+100 μg/ml ampicillin and 30 μg/ml chloramphenicol). Expression of YbeB from pCA24N was confirmed by protein expression analysis followed by Coomassie protein gel staining. Although pC24N contains *lacI^q*, maximal protein expression was constitutively observable in BW25113 w/o inductor (IPTG). An ONPG assay was repeated under equal conditions (2.2.3.10) and the *lacZ* reporter was induced with 0.2% (w/v) L-arabinose. The assay was carried out in LB medium containing 100 μg/ml ampicillin and 30 μg/ml chloramphenicol.

2.2.3.12 Detection of β-galactosidase reporter protein levels in *ΔybeB* and wild-type cells

Cell pellets harvested in intervals of 1 h from the ONPG assay (see above) were treated with 200 lysis buffer according to the pull down protocol (2.2.2.12). The soluble protein fraction was saved and the

total protein amount determined by Bradford assay (2.2.2.3). Protein concentration was adjusted with lysis buffer to 400 ng/ μ l per sample. 4 x Laemmli buffer was added and samples were boiled for 5 min at 95°C. 4 μ g total soluble protein per sample was loaded on an 8% SDS PAGE gel. Proteins were separated, Western blotted, and detected as described for the pull down experiments. Recombinant β -galactosidase was detected with polyclonal rabbit α -HA antibody (1:10,000) (Santa Santa Cruz Biotechnology) and goat α -rabbit/HRP antibody (1:5,000) (Dako). After ECL the PVDF membrane was stripped (2.2.2.8). For detection of a loading control the membrane was probed with polyclonal goat α -GST immunoglobulin (1:10,000) (Rockland) and secondary antibody as described in the pull down protocol.

2.2.3.13 GFP fluorescence reporter assay of *ybeB* gene deletion mutant

pBAD-GFP was used as reporter construct (kindly provided by Prof. Govind Rao, UMBC) (Albano et al., 1998). Plasmid was transformed into TSS-competent *Δ ybeB* and wild-type BW25113 and selected on LB agar medium with 100 μ g/ml ampicillin. These strains were diluted from o/n cultures 1:50 in 50 ml LB liquid medium with 100 μ g/ml ampicillin. Cultures were shaken in Erlenmeyer flasks at 37°C for 2 h. Then OD600 was determined and finally the cultures diluted to OD600=0.5. 15 ml cultures were induced with finally 0, 0.002, and 0.2% (w/v) L-arabinose and shaken at 37°C in Erlenmeyer flasks. In $\frac{1}{2}$ h steps 200 μ l were transferred into a 96-well plate (flat bottom) and GFP fluorescence was determined with FLUOStar OPTIMA fluorescence reader (BMG Labtech) (excitation filter 355 nm, emission filter 520 nm, gain 1,264). For fluorescence microscopy the assay was repeated as described for induction condition with 0.002% (w/v) L-arabinose. *Δ ybeB* and wild-type cultures were inoculated 1 h deferred to guarantee equal conditions for microscopy. 1.5 h post induction 200 μ l suspension was mixed with 200 μ l 40°C LB agar medium (containing 0.002% (w/v) L-arabinose and 100 μ g/ml ampicillin) for cell immobilization. 30 μ l were quickly loaded on a microscopy slide and covered with a cover slide. Samples were applied to fluorescence microscopy (2.2.4.4).

2.2.4 Cell culture and cell culture related assays

2.2.4.1 Cultivation

Cells (HeLa, HEK293TT) were cultured at 6% CO₂, 95% humidity, and 37°C in an incubator. Manipulation of cells was performed under a sterile hood. Media, buffers, and glassware were sterilized before work by autoclaving. Cells were grown in 10% DCS-DMEM medium.

Passaging cells

After removing the old tissue-culture media trypsin-containing solution (2 ml, 0.25% trypsin) was added to the cells and transferred to an incubator for 5 min. Thereafter, 10 ml fresh medium was

applied to the detached cells to stop trypsin-dependent digest and mixture was resuspended to separate cells. This mixture was administered to new Petri dishes in defined dilutions.

Seeding cells

After removing the old tissue-culture media, trypsin-containing solution (2 ml, 0.25% trypsin) was added to the cells and transferred to an incubator for 5 min. Thereafter, 10 ml fresh medium was applied to the detached cells to stop trypsin-dependent digestion and mixture was resuspended to separate cells. Cells were collected within a conical tube (15 ml), centrifuged (1,500 rpm, 5 min), and the supernatant was removed. The cell pellet was resuspended in 10 ml fresh medium. To count cells, 10 μ l were transferred to a Neubauer counting chamber. Cells were counted by using a brightfield-microscope.

Freezing and thawing cells

For freezing, logarithmically growing cells were trypsinized as described above and collected by centrifugation. Cells were resuspended in freezing medium (DMEM, 40% FCS, 10% DMSO) and transferred into cryo tubes. After incubation on ice for 30 min, cells were slowly frozen at -80°C and then transferred to liquid nitrogen. For re-propagation cells were thawed quickly at 37°C and transferred to fresh medium. The next day medium was replaced with fresh cell culture medium.

2.2.4.2 Cell transfection

Cells were seeded and grown in 250 μ l DMEM per well containing 10% Donor calf serum (DCS) (Invitrogen) in eight-well chamber slides (Ibidi) and incubated at 37°C . 16 h later cells were transfected with 100 ng DNA of single plasmid or 50 ng of plasmid for co-transfection using PromoFectin (Promokine). Plasmid DNA was diluted in 10 μ l culture medium. For each well 0.5 μ l PromoFectin solution was diluted in 10 μ l culture medium. Then, these two solutions were mixed, incubated for 15 min at room temperature, pipetted into the wells, and homogenized by swirling. Cells were incubated for 24 h at 37°C .

2.2.4.3 Cell preparation for fluorescence microscopy

24 h post transfection MitoTracker® Green FM for mitochondria staining (Invitrogen) (100 nM f.c.) was added to the growing cells for 30 min. Afterwards, the cells were washed 3 x with DMEM containing 10% DCS. Finally, DRAQ5 (Biostatus) was diluted 1:2,000 into the growing medium. Cells were examined 30 min post DRAQ5 administration by confocal microscopy.

2.2.4.4 Fluorescence microscopy

Fluorescence samples were analyzed using a Zeiss LSM 510 Meta confocal laser scanning microscope. Images were acquired using a 63x/1.40 plan apochromat oil objective. Tab. 22 gives the used fluorophore-specific excitation and emission filters:

Tab. 22 Fluorophores used in confocal microscopy

Fluorophore	Excitation laser/nm	Emission filter/nm
eGFP ²	488	510-530
mCherry	543	560-610
BiFC Venus (complements)	488	510-530
eCFP-membrane	458	460-475
DRAQ5	633	660-690 with META detector set
MitoTracker® Green FM	488	510-530

2.2.4.5 Co-localization and BiFC (bi-molecular fluorescence complementation) assay

For co-localization studies, HeLa cells were seeded and grown in 250 µl DMEM containing 10% Donor calf serum (DCS) (Invitrogen) in eight-well chamber slides (Ibidi, Martinsried Germany). 12 to 16 h later cells were transfected with 100 ng pECFP-mem and 100 ng C7orf30 fused to mCherry or 100 ng L14_{mt} fused to mCherry using Promofectin (Promokine, Germany) according to the manufacturer's manual. 24 h afterwards, Mitotracker (100 nM f.c.) was added for 30 min. Thereafter, cells were washed 3 x with DMEM containing 10% DCS and finally DRAQ5 (finally diluted 1:2000) was added. Cells were examined 30 min post DRAQ5 administration by fluorescence microscopy.

For BiFC (Hu et al., 2002), HeLa cells were seeded and grown in 250 µl DMEM containing 10% Donor calf serum (Invitrogen) in eight-well chamber slides (Ibidi, Martinsried, Germany). 12 to 16 h later cells were transfected with 100 ng pECFP-Mem (Clontech) and 50 ng of C7orf30 fused to the N-terminal part of Venus (VN) together with L14_{mt} fused to the C-terminal part of Venus (VC), C7orf30-VC together with L14_{mt}-VN, C7orf30-VC together with C7orf30-VN or L14_{mt}-VN together with L14_{mt}-VC using Promofectin (Promokine, Germany). 24 h post transfection, DRAQ5 was added (finally diluted 1:2000) and cells were examined 30 min post DRAQ5 administration by fluorescence microscopy.

² Enhanced fluorescence activity and shorter chromophore cyclization time of 95 min compared to 4 h of wt GFP (Albano et al., 1998; Cramer et al., 1996).

2.2.4.6 LuMPIS

LuMPIS (Luminescence-based MBP pull-down Interaction screening system) is a LUMIER-derived pulldown assay (Luminescence-based mammalian interactome mapping). It makes use of vectors with N-terminally MBP-tagged baits to capture N-terminally eGFP-luciferase-tagged preys which can be co-purified in a pulled down assay via amylose beads. Proteins are expressed in human embryonic kidney cells (HEK) and raw protein extracts are used for the assay. The system was used to verify phage-host PPIs, detected in the Y2H screens. It was especially developed for expression of low GC content ORFs (Vizoso Pinto et al., 2009). The pulled down prey is detected by measurement of the luciferase activity.

ORFs of PPI candidates were transferred from the corresponding Gateway® entry plasmids via LR reactions into the expression plasmids: phage ORFs were inserted into pCR3.1-N-eGFP_{Luc} and used as preys. *S. pneumoniae* ORFs were transferred into pCR3.1-N-MBP and used as baits. Plasmids were isolated by isopropanol precipitation (2.2.1.7) and applied for cell transfection. Each PPI was measured as quadruplicates and compared to a quadruplicate negative control. The empty bait plasmid (MBP w/o ORF) was used in combination with the GFP_{Luc} preys to determine prey binding to MBP.

The following protocol was used:

- Chill all buffers needed and solutions on ice (see below).
- *Day 1*: Seed 0.5×10^6 HEK293TT per well of a 12-well plate in the morning. $\frac{1}{2}$ h before transfection, check that the cells are adherent and change the medium by aspirating the old medium and add 500 μ l OPTIMEM per well. Transfect cells with 500 ng/total DNA per well (250 ng MBP-tagged-bait and 250 ng eGFP-Luc-tagged-prey plasmid) (PromoFectin) in the afternoon (after at least 6 h).
- *Day 2*: Change medium (DMEM added with 10% FCS; 1% penicillin/streptomycin (1:100), 1% non essential amino acids (1:100); 1% gentamicine (stock 10 mg/ml), Fungizone (1:500), G418 (100 μ g/ml f.c.).
- *Day 3*: Prepare MultiScreenHTS 96-well plates (Millipore): Dispense 100 μ l slurry (amylose beads, NEB) per well in HTS plate. Equilibrate the beads in LuMPIS buffer (wash 4 x in 200 μ l using a vacuum device).
- Wash the transfected cells (in 12 well format) with 500 μ l ice cold PBS.
- Resuspend the cells in lysis buffer supplemented with 0.05 % Tween-20 using 500 μ l lysis buffer.
- Lyse the cells by sonication on ice. Use six bursts at 75 W with a 10 s cooling period between each burst (Bandelin Sonorex RK 100).
- Centrifuge the lysate at 10,000 g for 10 min at 4°C to pellet cellular debris and DNA. Save the supernatant.
- Take 100 μ l and dilute it 1:10 with LuMPIS buffer.
- Apply 150 μ l lysate into each well, mix thoroughly by pipetting up and down (approx. ten times).
- Apply vacuum to wash the column; add 200 μ l wash buffer (mix thoroughly each time by pipetting up and down) and wash by applying to vacuum. Repeat the washing steps 4 x.
- Resuspend the beads in 150 μ l elution buffer.
- Pipet 50 μ l of each suspension into the wells of a white 96-well plate. Measure luciferase activity in Optima FLUOstar Luminometer (BMG LABTech) of the lysates (1:10) and of the beads suspensions according to the Promega protocol, using 50 μ l reagent.

The luciferase luminescence intensity ratio (LIR) is calculated as follows:

$$\text{LIR} = \frac{\text{Eluate}_{\text{bait+prey}} / \text{Lysate}_{\text{bait+prey}}}{\text{Eluate}_{\text{control+prey}} / \text{Lysate}_{\text{control+prey}}}$$

Hereby (_{control+prey}) represents the luminescence intensity of a negative control sample determined for individual eGFP-luciferase tagged preys. Here, the unspecific binding of a prey was tested against the MBP tag only. (_{bait+prey}) represents the sample with the eGFP-luciferase tagged preys and MBP-tagged baits.

LuMPIS Buffer (column buffer for amylose beads, wash buffer)

- 20 mM Tris-HCl
- 200 mM NaCl
- 1 mM EDTA

Elution buffer

- 10 mM maltose in H₂O dist.

Lysis Buffer (50 ml)

- | | |
|--|---------------------|
| ▪ 20 µl Leupeptin (Sigma, Stock aliquots: 25 mg/ml; -20°C) | (5 µg/ml f.c.) |
| ▪ 25 µl DNase I (Sigma, Stock aliquots: 10 mg/ml; -20°C) | (5 µg/ml f.c.) |
| ▪ 0.125 g BSA | (2.5 mg/ml f.c.) |
| ▪ 25 µl Tween-20 | (0.05 % (v/v) f.c.) |
| ▪ Prepare in column buffer | |

2.2.5 Yeast methods and Yeast Two-Hybrid assays

2.2.5.1 Yeast media

YEPD medium

- 10 g yeast extract
- 20 g peptone
- 20 g glucose
- (16 g agar agar for solid medium, only)
- add H₂O to 1 liter
- Autoclave
- Cool down to approx. 60°C and add 4 ml of 1% adenine solution (1% in 0.1 M NaOH)

5 x medium concentrate

- 8.5 g yeast nitrogen base
- 25 g ammonium sulfate
- 100 g glucose
- 7 g dropout mix (see below)
- add H₂O to 1 liter and sterile filter; store at 4°C

Selective plates

- autoclave 16 g agar in 800 ml water
- cool medium to 60-70°C
- add 200 ml 5 x medium concentrate
- add components as follows:

Plate type

add

- | | |
|--|------------------------------------|
| ▪ -Trp (w/o Trp) | 8.3 ml Leu and 8.3 ml His solution |
| ▪ -Leu (w/o Leu) | 8.3 ml Trp and 8.3 ml His solution |
| ▪ -Leu-Trp (w/o Trp and Leu) | 8.3 ml His solution |
| ▪ -Leu-Trp-His (w/o Trp, Leu, and His) | nothing or 3-AT, only |

For -Leu-Trp-His, 3-AT can be added from a 0.5 M stock solution. Approx. 20 ml of the medium is then poured into OmniTray plates (Nunc).

Dropout mix (-His, -Leu, -Trp)

- | | | |
|-----------------------|---------------------|------------------|
| ▪ 1 g methionine | ▪ 5 g glutamic acid | ▪ 1 g adenine |
| ▪ 1 g arginine | ▪ 5 g aspartic acid | ▪ 1 g uracil |
| ▪ 2.5 g phenylalanine | ▪ 7.5 g valine | ▪ 20 g serine |
| ▪ 3 g lysine | ▪ 10 g threonine | ▪ 4 g isoleucine |
| ▪ 3 g tyrosine | | |

Mix all components well and store under dry, sterile conditions.

Amino acid solutions

- *Histidine (His)*: dissolve 4 g of histidine in 1 liter water and sterile filter.
- *Leucine (Leu)*: dissolve 7.2 g of leucine in 1 liter water and sterile filter.
- *Tryptophan (Trp)*: dissolve 4.8 g of tryptophan in 1 liter water and sterile filter.

2.2.5.2 Yeast transformation

This protocol is suitable for 100 yeast transformations and may be scaled up or down as needed. Selection of the transformed yeast cells requires leucine or tryptophan-free media (“-Leu” or “-Trp”, depending on the selective marker on the plasmid).

- Inoculate 50 ml YEPD liquid medium with one colony of haploid AH109 or Y187 yeast strain in a 250 ml flask and grow o/n with shaking at 30 °C (min. 15 h, max. 24 h).
- Spin down cells in 50 ml conical tube (2,000 rpm, 5 min at room temperature, table top centrifuge 5810R), pour off supernatant and dissolve the pellet by adding 2 ml LiOAc (0.1 M) and transfer resuspended yeast to two 1.5 ml microfuge tubes. Spin out yeast (2,000 rpm, room temperature, micro centrifuge) and resuspend in a total volume of 1.8 ml LiOAc (0.1 M).
- Prepare CT110 solution.
- Add all the competent yeast cells prepared above and mix vigorously by hand or by vortexing for 1 min. Pipet immediately 245 µl into each of 96-wells of a 96-well plate.
- Add ~100 ng of plasmid and positive control (e.g., empty vector) and negative control (only CT110). Seal the 96-well plate with plastic or aluminum tape and vortex for 4 min.
- Incubate at 42°C for 30 min.
- Spin the 96-well plate for 10 min at 2,000 rpm (Labofuge 400R); discard the supernatant and aspirate with 8 channel wand or by tapping on cotton napkin for a couple of times. Add 150 µl of sterile water to all 96-wells, resuspend and plate cells on selective agar plates (e.g., standard Petri dishes) with -Leu for pGADT7g or -Trp for pGBKT7g selection.
- Incubate the plates at 30°C for 3 days.

Carrier DNA (salmon sperm DNA): Dissolve 7.75 mg/ml salmon sperm DNA (Sigma D1626) in H₂O and store at -20°C following a 15 min 120°C autoclave cycle.

96 PEG solution (100 ml): Mix 45.6 g PEG (Sigma P3640), 6.1 ml of 2 M LiOAc (lithium acetate), 1.14 ml of 1 M Tris pH 7.5, and 232 µl 0.5 M EDTA; make up to 100 ml with sterile water and autoclave. Store at room temperature.

CT110: Mix 20.73 ml 96PEG, 0.58 ml boiled salmon sperm DNA (boil at 95°C for 5 min) and 2.62 ml DMSO. Add DMSO last and mix quickly after adding by shaking vigorously for 30 s.

2.2.5.3 Setup of a screen array

Usually preys rather than baits are arranged on an array because the former do not result in self-activation of transcription. However, screening against a bait array is possible as well but must be controlled by screening a prey strain carrying an empty prey vector against the bait matrix. The minimal-inhibitory concentration of 3-AT for each bait strain is determined by a self-activation test and bait strains with equal minimal-inhibitory concentration are re-arrayed on the same test array plates. For creation of prey and bait arrays, the layout of the array is defined first. Each prey and bait construct of the proteome is given a specific position of a particular 96-well plate, e.g., position A03 of prey plate #1 contains the *orf3*-construct. The wells of these 96-well-plates are loaded with 100 µl YEPD medium. Several colonies from a specific prey and bait transformation are combined and

manually transferred into the well at the previously defined position. These 96-well plates carrying the prey or bait strains are incubated o/n at 30°C and replicated onto Omnitray plates with solid media – at least one selective Omnitray plate (-Leu) and one YEPD plate. 50 µl 50% (v/v) glycerol is added to the liquid culture plate; the plate is sealed and transferred to -80°C for long term storage. The solid prey and bait plates with prey or bait strains in the 96-format can be quadruplicated to the 384-format using a robotic procedure (Biomek2000 laboratory robot). For increased throughput duplicates rather than quadruplicates from two 96-formatted plates can be combined on one 384-formatted plate. This prey or bait array is usually stored on selective plates. Additionally, “Working” copies for Y2H screens on solid YEPD medium have to be made.

2.2.5.4 Bait and prey yeast selection conditions

The following table summarizes the selection conditions for the used vector systems. For instance, pDEST Y2H-expression vectors (Invitrogen) contain swapped yeast auxotrophy selection markers in contrast to the pGBKT7g/pGADT7g system. Thus, the pDEST and GBK/GAD system is incompatible for Y2H analysis (e.g., pDEST32 baits cannot be mated with pGADT7g preys since both contain a LEU2 marker). In this work, bait vectors were all transformed into AH109 yeast strain, prey vectors into Y187.

Tab. 23 Plasmid selection conditions in yeast

Plasmid	Auxotrophy marker	Selective medium	1n yeast strain used
pDEST32 (bait plasmid)	LEU2	-Leu	AH109
pDEST22 (prey plasmid)	TRP1	-Trp	Y187
pGBKT7g (bait plasmid)	TRP1	-Trp	AH109
pGADT7g (prey plasmid)	LEU2	-Leu	Y187

2.2.5.5 Bait self-activation test

The aim of this test is to measure the background reporter activity (here: HIS3) of bait proteins in absence of an interacting prey protein. This measurement is used for choosing the selection conditions used during the interaction screen and can be achieved by mating individual bait strains with a single prey strain that carries the empty prey plasmid. 96 individual bait activation tests can be carried out on one plate simultaneously. This test must be done prior to the screen to determine the appropriate concentration of 3-AT that has to be added to suppress self-activation of a certain bait:

- Load a 96-well plate (round bottom) with ~200 µl YEPD liquid medium.
- Inoculate plate with baits by replicating the 96-format bait plate from solid medium into the destination plate by using a sterile 96-pinning tool (Biomek2000).
- Inoculate the yeast strain Y187 which carries the empty prey vector in 30-50 ml YEPD liquid medium.
- Grow yeast for ~18 h at 30°C (it is not necessary to shake the 96-well plate whereas shaking of the prey strain in a flask is recommended).

- Pellet yeast in the 96-well plate by centrifugation for 10 min at 2,000 rpm (Labofuge 400R). Discard the supernatant and aspirate with 8 channel wand or by tapping on cotton napkin for couple of times.
- Use 96-replication tool (Biomek2000) to pin baits from 96-well source plate onto a YEPD single-well agar plate as quadruplicates.
- Pour the yeast strain with the empty prey vector into a single-well plate.
- Use 384-replication tool (Biomek2000) to pin yeast onto the YEPD single-well agar plate that harbors the baits already.
- Mating occurs at 30°C for min. 1 to max. 2 days.
- Replicate from mating plate on -Leu-Trp-agar single well plates to select diploids.
- Incubate for 2 to 3 days at 30°C.
- Pin diploids on -Leu-Trp-His agar medium in single-well plates with different concentrations of 3-AT (e.g., 0, 1, 2, 4, 8, ..., 128 mM).
- Select yeast for 7 days at 30°C.
- Determine minimal-inhibitory concentration of 3-AT which is needed for each bait to suppress self-activation growth for use in the interaction screen.

2.2.5.6 Matrix-based Yeast Two-Hybrid screens – one-on-one tests

Preparations

- *Sterilization steps*: sterilize the replication tool by dipping the pins into 20% bleach for 20 s, sterile H₂O for 1 s, 95% ethanol for 20 s, and sterile H₂O again for 1 s. Repeat this sterilization after each transfer step.
- *Consideration of automatization*: all replication steps can be automated by using a replication robot (Biomek2000). This ensures experimental reproducibility and increases the test throughput.
- *Prepare prey array for screening*: use the sterile replicator to transfer the yeast prey array (e.g., 384 format) from selective plates to single-well plates containing solid YEPD medium and grow the array o/n in a 30°C incubator (max. 24 h). Ideally, the template prey array should be kept on selective plates.
- *Prepare bait liquid culture (DBD fusion-expressing yeast strain)*: inoculate 20 to 30 ml of liquid YEPD medium in a 50 ml Erlenmeyer flask with a bait strain from plates with selective medium and grow o/n in 30°C shaker for 18 to 22 h.

Mating procedure

- Add adenine from a 1% (w/v) adenine stock solution (prepare in 0.1 M NaOH) to a final concentration of 0.004% into the bait liquid culture.
- Pour the o/n liquid bait culture into a sterile Omnitray plate. Dip the sterilized pins of the pin-replicator into the bait liquid culture and place directly onto a fresh single-well (Omnitray) plate containing solid YEPD medium. Repeat with the required number of plates.
- Pick up the prey array yeast colonies with sterilized pins and transfer them directly onto the baits pinned onto the YEPD plate, so that each of the 384 bait spots receives different prey yeast strains.
- Incubate 1 to 2 days at 30°C to allow mating.

Selection of diploids

- For the selection and enrichment of diploids, transfer the colonies from YEPD mating plates to single-well plates containing -Leu-Trp medium using the sterilized pinning tool. Grow for min. 2 to max. 3 days at 30°C until the colonies are >1 mm in diameter.

Interaction selection

- Transfer the colonies from -Leu-Trp plates to single-well plates containing solid -Leu-Trp-His agar, using the sterilized pinning tool. If the baits are self-activating, they have to be transferred to -Leu-Trp-His + the specific concentration of 3-AT which was determined in the self-activation assay. Incubate at 30°C for 7 days.
- Afterwards, score the interactions by looking for growing colonies that are significantly above background by size and that appear reproducibly as duplicates (or quadruplicates).

2.2.5.7 Yeast Two-Hybrid mini-pool screens

This protocol describes pooling of a prey array that contains 960 different clones resulting in one pool plate with 10 individual prey strains per mini pools. The protocol can be adjusted to create larger pools than 10. For this the corresponding volumes have to be adjusted.

- Load 10 96-well plates with ~200 µl liquid YEPD medium per well.
- Inoculate these plates with preys from the prey array grown on corresponding selective agar medium with a sterilized 96-pinning tool.
- Grow yeast for 18 to 22 h at 30°C.
- Resuspend yeast in all wells by pipetting with a 12-channel pipette.
- Transfer from each source plate 20 µl yeast suspension into the destination pooling plate. Pool all preys from source positions A1, A2, ..., H12.
- Once the transfer is done mix the wells of the pooling plate by pipetting.
- Spin the pooling plate for 10 min at 2,000 rpm (Labofuge 400R).
- Discard the supernatant and aspirate with 8-channel wand or by tapping on cotton napkin a couple of times.
- Pin yeast with a sterile 96-pinning tool (Biomek2000) from the prey pool plate onto YEPD solid medium as quadruplicates.

Mating of a single baits and selection steps follow the previous screening protocol. The prey pooling plate can be used directly to be mated against 10 different baits. Then the pools are depleted. After selection of -Leu-Trp-His medium interacting preys have to be identified by a yeast colony PCR followed by sequencing reaction (see below).

2.2.5.8 Yeast colony PCR

During prey-pooling screens the interacting prey cannot be identified immediately by its matrix position as in one-on-one screens. A possibility to identify positives is to amplify the prey ORF from the plasmid by a yeast colony PCR followed by sequencing of the prey PCR product:

- Pick diploid yeast from -Leu-Trp-His selective plates with a sterile pipet tip and transfer into a well with 100 µl H₂O dist. (96-PCR plate (named as “yeast plate”)). The amount of transferred yeast should be low since too many cells lead to a decreased PCR efficiency.
- Cell lysis:
 - Pipet 1 µl of Zymolase (5 U/µl, Zymo Research) per reaction into a new 96-PCR plate (“lysis plate”).
 - Transfer 7 µl of yeast suspension from “yeast plate” into “lysis plate” and mix.
 - Incubate “lysis plate” in a thermocycler for 30 min at 30°C.

- Use 1 μ l of lysed cells as PCR template and perform a standard colony PCR as described in (2.2.1.2) with Taq DNA polymerase (Promega). For the PCR of pDEST22 preys use primer pair pDEST22_for/pDEST22_rev.
- Check 5 μ l of PCR product by agarose gel electrophoresis. For proper sequencing no by-products should be observable.
- Prepare PCR samples:
 - Transfer 8 μ l of PCR samples into a new 96-PCR plate.
 - Prepare a corresponding volume of SAP mastermix: per sample use 0.5 μ l 10 x SAP buffer (Fermentas), 8.95 μ l H₂O dist., 0.5 μ l SAP (1 U/ μ l; Fermentas); 0.05 μ l Exonuclease I (20 U/ μ l, Fermentas).
 - Transfer to 8 μ l PCR sample 10 μ l of SAP mastermix and mix.
 - Incubate in a thermocycler for 1 h at 37°C and then for inactivation for 15 min at 72°C.
 - Add to each well 3 μ l of 3.2 pmol/ μ l pDEST22_for oligo for forwards sequencing reactions.

2.2.5.9 Yeast Two-Hybrid retests

Testing for reproducibility of interactions greatly increases the reliability of the interaction data. This protocol is used for specifically retesting interaction pairs detected in an one-on-one or pooling screen.

- Re-array bait and prey strains or a positively tested prey pool of each interaction pair in a 96-well plates. Use an individual 96-well plate for the baits as well as for the preys. For each retested interaction fill one well of the bait plate and one corresponding well of the prey plate with ~200 μ l YEPD.
- For each retested interaction inoculate the bait strain in a well of the 96-well bait plate and the prey strain at the corresponding position of the 96-well prey plate. E.g., bait “X” is transferred at positions B1, B2, and B3 of the bait plate. The preys to be tested are arrayed into B1 (prey “Y”), B2 (prey “Z”), and the prey strain that carries the empty prey vector into B3 of the prey plate. The B3 test position is the control that helps to verify the background/self-activation.
- Incubate the plates O/N at 30°C.
- Spin the bait and prey plates for 10 min at 2,000 rpm (Labofuge 400R).
- Discard the supernatant and aspirate with 8 channel wand or by tapping on cotton napkin a couple of times.
- Pin baits and preys with a sterile 96-pinning tool separately on corresponding selective medium as quadruplicates (-Trp or -Leu medium).
- Allow baits and preys to grow at 30°C for min. 2 to max. 3 days.
- Mating: First transfer baits with a sterile 384-pinning tool on YEPD mating plates. Then transfer preys onto baits.

The rest of the procedure can be done according to the screening protocol (2.2.5.6). For interaction retesting diploids are pinned on -Leu-Trp-His selective media plates with different concentration of 3-AT (e.g., as used in the bait self-activation pretest). The control test position has to be considered concerning bait self-activation background signals. Reproducible interactions should show up on different concentrations of 3-AT in comparison to the activation control test position.

2.2.6 Bioinformatics and statistics

Multiple alignments

Multiple sequence alignments were built with protein amino acid sequences using ClustalW2 web application tool by using default setting (Larkin et al., 2007).

(<http://www.ebi.ac.uk/Tools/clustalw2/index.html>)

Web logos

Logos of consensus sequences or the degree of amino acid conservation of multiple aligned protein sequences were built with WebLogo V2.8.2 web application tool (Crooks et al., 2004). This was done to improve the presentation of multiple sequence alignments since WebLogo highlight also less conserved amino acid residues. Therefore, multiple sequence alignments which have been built with ClustalW were uploaded using default parameters.

(<http://weblogo.berkeley.edu/logo.cgi>)

ConSurf

To visualize the amino acid residue conservation degree of *E. coli* L14 projected on its 3D-structure, ConSurf Surfer V3.0 web application tool was used (Landau et al., 2005). The ClustalW multiple sequence alignment of L14 homologs (Fig. 35) was uploaded for calculation the ConSurf conservation scores using default parameters. Conservation scores were projected onto PDB entry 2AW4, chain K (L14) and then fitted into the 50S subunit 3D-structure (2AW4) (Schuwirth et al., 2005). 3D-structures were visualized with PyMol (<http://pymol.org>).

Protein docking

For the protein docking simulations, the individual protein structures were collected as follows. The L14 protein was taken from 2AWB PDB entry, chain K. Because a crystal structure of *E. coli* YbeB is not available I-TASSER server was used (Roy et al., 2010) to build a computation model of YbeB. With its default parameters I-TASSER produced a single model that used both available PDB YbeB templates 2ID1_A and 2O5A_A. The server estimated the accuracy of the model as 0.90 ± 0.06 (TM-score) and 1.6 ± 1.4 Å (RMSD).

An unconstrained rigid body docking was performed for individual L14 and YbeB structures with GRAMM-X (Tovchigrechko and Vakser, 2006). Then coordinates of L14 protein were used to superimpose 100 top scored docking models onto the entire 70S unit (PDB entries 2AWB and 2AW7 taken together). Based on that, each model was evaluated for the presence of the backbone clashes between the predicted YbeB position and the rest of 50S subunit. A backbone clash was defined as having a distance between any two protein or nucleotide backbone atoms less than 2 Å. This definition was chosen in order to tolerate some degree of unknown conformational re-arrangement of the 50S components that were not used in docking. Each model was analyzed towards the presence at the predicted L14-YbeB interface of the critical residues identified by the directed mutagenesis. The

definition of the interface residue was to have at least one non-hydrogen atom within 4.6 Å distance from any non-hydrogen atom of the docking partner. The post-docking analysis and graphical output were performed with PyMol (<http://pymol.org>).

Local protein and nucleotide sequence blast

Identification of homologous protein sequences was done by using blastp function of BioEdit 7.0.9.0 (Hall, 1999). A local protein sequence database was created against which a set of query protein sequences was blasted. Sequenced PCR products from yeast colony PCRs (2.2.5.8) were blasted (blastn function) against a local nucleotide database containing all ORF nucleotide sequences of *S. pneumoniae* TIGR4.

Data storage

Data about clones and arrays positions were stored in FileMaker databases, a relational database (FileMaker Pro 8.5). It was used for scoring interactions detected in the Y2H screens and for relational calculations.

Sequence-based domain predictions

Domain information was collected from Pfam and InterPro database (Finn et al., 2010; Hunter et al., 2009). Pfam provides for prokaryotic protein sequences the most comprehensive dataset while InterPro integrates domain information from various primary databases. Pfam uses Hidden Markov Models (HMM) that detect conserved sequence signatures.

Annotation of the Dp-1 genome

The nucleotide sequence of Dp-1 genome was analyzed for the presence of open reading frames (ORFs) of a size longer than 30 codons using Genemark.hmm (Borodovsky et al., 2003; Lukashin and Borodovsky, 1998), Easygene 1.2 server (Larsen and Krogh, 2003) and NCBI ORF-finder. Identified ORFs were checked manually for the presence of a Shine-Dalgarno sequence (ribosome binding sites, RBS) located in the 25 nucleotides sequence upstream of putative ORFs. If alternative start codons were identified in ORF-finder the closest start-codon to the RBS was selected as the translation start site. Genes were searched against genomic databases using BLAST-N (Altschul et al., 1990). Protein sequences were blasted against public protein databases (non-redundant protein sequences accessible from NCBI BLAST interface, ACLAME (Leplae et al., 2010) and Uniprot Knowledgebase (Apweiler et al., 2004)) by using BLAST-P, Psi-BLAST (Altschul et al., 1990; Altschul and Koonin, 1998) and FASTA (Pearson, 2004)). Protein sequences were checked for presence of conserved domains using CDD (Marchler-Bauer and Bryant, 2004), InterProScan (Mulder and Apweiler, 2007), SMART HMM (Letunic et al., 2009), and Pfam sequence search (Finn et al., 2010). Promoters were identified by searching manually for sigma70 -35 box consensus sequences and -10 (TATAA(T)) box sequences upstream of putative ORFs. The presence of putative Rho-independent terminators was investigated using mFOLD (Zuker, 2003) to identify secondary RNA stem and loop structures and to search for

adenine- and thymine-rich regions upstream and downstream of these identified structures (Lesnik et al., 2001). The Dp-1 genome was searched for the presence of putative tRNA encoding sequences using tRNAscan (Schattner et al., 2005) and for a DnaA-dependent origin of replication by using Ori-Finder web-based application (Gao and Zhang, 2008).

Dp-1 genome context vs. detected protein-protein interactions

The number of interactions between various gene classes (predicted functional classes, predicted expression timing) was counted for the real intra-viral PPI network of Dp-1 and for randomized versions of this network. Randomized versions of the network were generated using a network rewiring algorithm, which conserves the node in- and out-degrees of each protein (Maslov and Sneppen, 2002). Overrepresentation of a link compared to 10,000 randomized networks was assessed by calculating a Z-score

$$Z = \frac{n - \langle n_{rand} \rangle}{\sigma_{rand}}$$

with the number of linking interactions n , its average in 1,000 randomized networks $\langle n_{rand} \rangle$, and its standard deviation σ_{rand} . A permutation p-value was calculated based on the fraction of randomized networks with a number of linking interactions higher (lower) or equal to the number of linking interactions in the actual network. A Bonferroni correction was used to control for multiple testing.

3 Results

3.1 Functional proteomics of the lytic bacteriophages Cp-1 and Dp-1

3.1.1 Functional annotation of the Dp-1 genome

To learn more about the fundamental biology of Dp-1 I annotated its genome. Work on this chapter was done in collaboration with Mourad Sabri (Département de biochimie et de microbiologie, Université Laval, Québec). A small section of the Dp-1 genome sequence was already published by Sheehan and colleagues (Sheehan et al., 1997). This sequence spans 4,735 bp of the lysis cassette. That investigation was done to identify and characterize the Dp-1 lytic enzyme Pal. In the meantime the whole Dp-1 genome sequence was determined. Sequencing was done in 2000 by The Medicines Company, Ville St. Laurent, Quebec, Canada and patented (Pelletier et al., 2000). However, the sequence was never published, submitted to GenBank, or fully annotated. Since Dp-1 is a lytic siphovirus with potential for medical applications, it is of general interest to provide a comprehensively annotated genome. Moreover, knowledge about the positions of all ORFs and the protein composition is required for further characterization and understanding of its biology, e.g., for my intention to analyze the Dp-1 interactome.

General characteristics of the Dp-1 genome and open reading frame prediction

The dsDNA genome of bacteriophage Dp-1 is 56,506 bp in length. The average GC content of Dp-1 nucleotide sequence is 40.3%, which is close to the 40% GC content of the *S. pneumoniae* R6 host genome. The full genome sequence/Genbank file is attached in the e-supplement (electronic supplementary material provided on the CD).

In total 72 putative protein-encoding ORFs were found that are longer than 30 codons and have a detectable Shine-Dalgarno sequence with one exception (see below) which is required for a successful non-leaderless translation initiation. Systematical numbers were given to ORFs in the order they occur on the DNA molecule (from 1 to 72). Correspondingly, putative gene products are named “gp”, e.g., gp1 (Tab. 24).

63 ORFs are located on the upper strand while a module of nine ORFs (*orf27-35*) is located on the opposite strand. The codon number (stop codons excluded) varies from the smallest ORF, *orf20* with 47 triplets, to the largest ORF, *orf54* with 1,230 codons (Fig. 17).

With the exception of *orf46*, a ribosome binding site (RBS) was found for each of the putative ORFs with 20 of these RBSs presenting the well conserved six bp Shine-Dalgarno consensus sequence AGGAGG. The average distance between the putative ORF and end of the RBS is 8 nt. The majority of RBSs end 7 bp upstream of the start codon of the respective ORFs. Out of 72 putative genes, 13 have an alternative start codon closer to the predicted RBS than the canonical ATG start codon (GTG

RESULTS

(4), TTG (6), ATT (3)). TAA is the most used stop codon (TAA (49%), TAG (31%), TGA (20%)). Overall 42% of Dp-1 ORFs are overlapping. The coding capacity of the Dp-1 genome is approx. 94%, which is similar to the genome of *Pneumococcus* bacteriophage Cp-1 (93%) (Martin et al., 1996b). All ORF coordinates are given in Tab. 24 and RBS sequences can be found in the e-supplement.

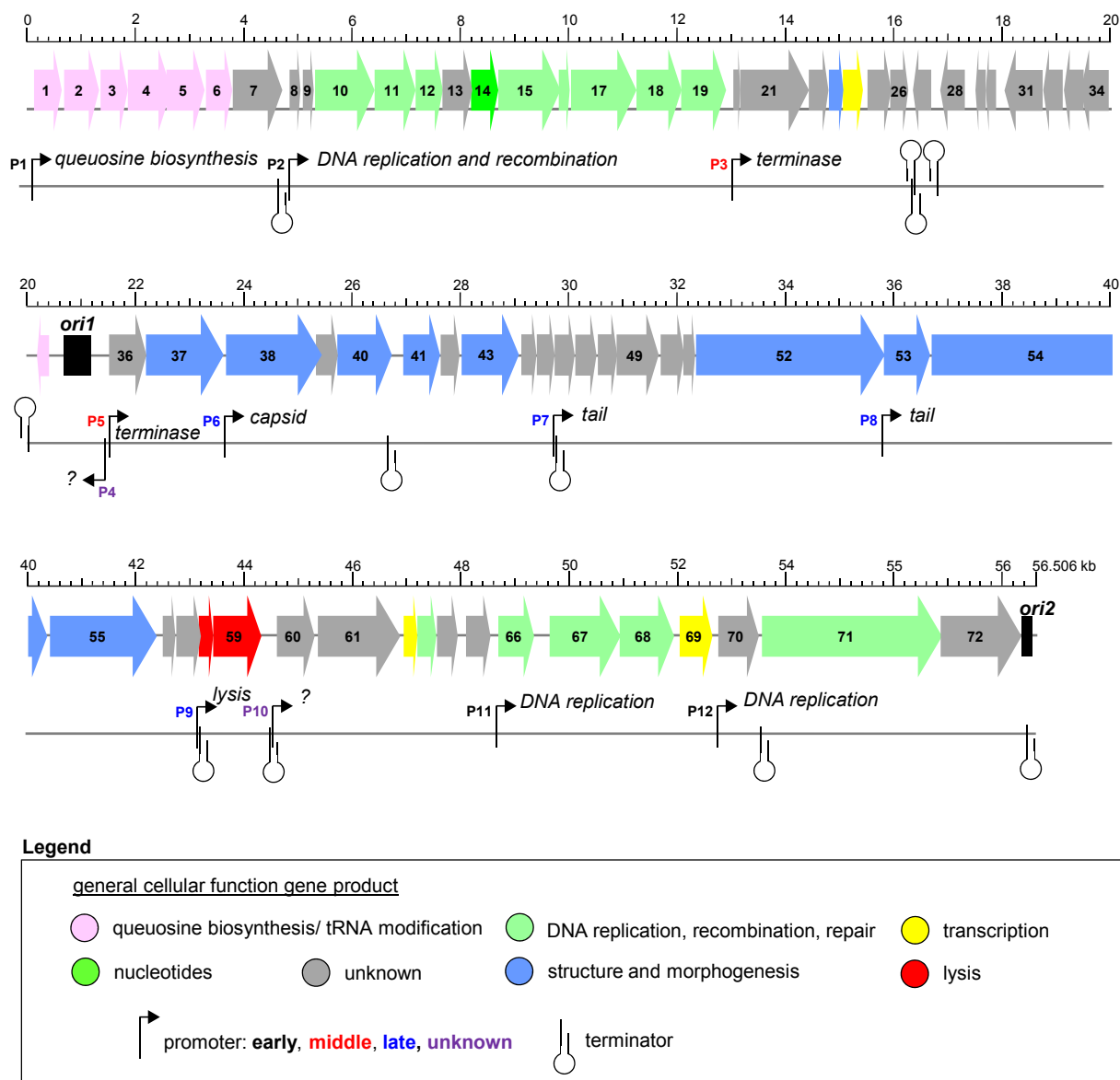


Fig. 17 Dp-1 genome map

Numbers in the ORF symbols correspond to the systematic ORF numbers. Homology-based functions and further details for certain ORFs can be looked up in Tab. 24. An assumed general function for each gene product is indicated by a color code (see legend). The predicted transcriptional start and termination sites are projected below the ORF map. Information was combined from Tab. 24, Tab. 25, and Tab. 26. Putative *oris* (replication origins) are indicated (see below).

Homology-based prediction of protein functions

The translated nucleotide sequences of the 72 predicted ORFs were compared to protein sequences from various databases (2.2.6). This was done to identify orthologous proteins and domain signatures as functional indicators for putative Dp-1 gene products. Only for three putative proteins no similarity matches at all were found (gp8, 9, and 20). These putative proteins are orphans and attributed to Dp-1 only. For all others homology matches were detected. Significant matches (defined by an e-value $\leq 1e^{-8}$ and a query alignment coverage of $\geq 50\%$) were found for 39 proteins. These proteins represent well-conserved Dp-1 homologs (Fig. 18A). However, only 16 of them were useful for adoption of a reliable functional annotation of the corresponding Dp-1 orthologs since most of the relatives had no functional annotation and are mostly described as “hypothetical proteins”. In addition, the Dp-1 sequences were blasted against several databases to predict homology-based domains, motifs, and protein families (Pfam, InterproScan, NCBI conserved domain search (CDS) (2.2.6)) and any available information that could be helpful to get ideas about the proteins’ functions was considered. Conserved sequence signatures could be detected for 49 Dp-1 proteins. The sequence features were compared with the blastp best-best hits. If a significant feature/domain correlated with the annotation of a non-significant blastp ortholog this additional information was considered to annotate a protein. Also, if sequence feature hits were detectable by independent methods (e.g., hits by Pfam domain and Smart domain prediction) the results were taken into account for functional annotation. Thus, 26 Dp-1 proteins could be functionally annotated in addition. 21 of them were annotated as “putatives”, “like”, or defined by the detected domain signature since they differed from canonical homologs, e.g., absence of a detectable second co-occurring domain that would have been necessary for a reliable prediction (explained for gp16, below) or appearance of a significant domain signature that is clearly present, but in the context of the query sequence allows no detailed functional annotation (explained for gp39 and gp44 (see below)).

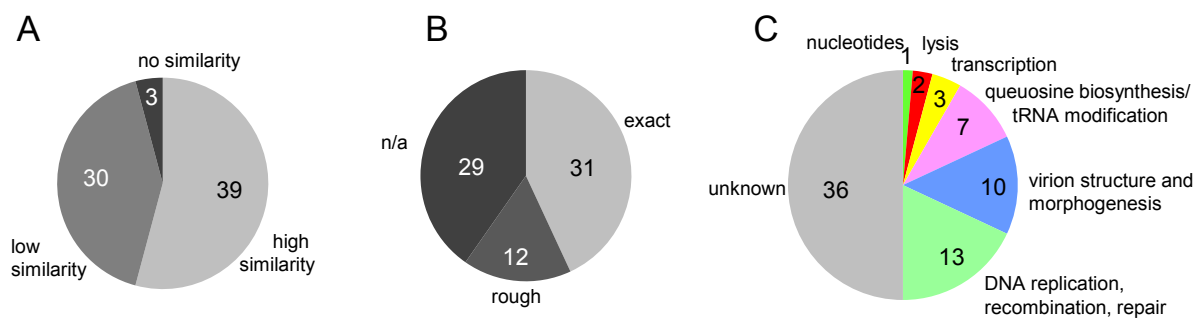


Fig. 18 Homology-based Dp-1 protein annotation

Absolute numbers of Dp-1 proteins are given in the diagrams. **(A)** Blastp best-best hit conservation degree: (no similarity) no sequence homologs at all were detected, (high similarity) best-best hit with e-value $< 1e^{-8}$ and a query sequence alignment coverage of $> 50\%$, (low similarity) all other best-best hits. **(B)** Information content of protein functions: (exact) exact function is predicted, (rough) only rough estimation of function is possible, (n/a) no functional information at all. **(C)** Simplified categorization of cellular functions of Dp-1 gene products.

Exact functions (e.g., an exact enzymatic function) of Dp-1 proteins could be predicted for 31 proteins, rough functional assumptions for 12 further proteins (e.g., “helicase function” or “membrane protein” based on domain or other feature predictions), and no functional assignments could be made for 29 proteins (Fig. 18B). Finally, each classified protein function was simplified and grouped to a general cellular function or pathway (Fig. 18C).

Extended documentation of homology-based function predictions is provided in the e-supplement.

Annotations that need some further explanations

Although for 31 proteins an exact, reliable annotation was found, some need a brief explanation.

Gp16 was annotated as putative DNA ligase since domain DNA ligase signatures were detected (NAD-dependent DNA ligase adenylation domain (Pfam PF01653) and NAD-dependent DNA ligase (COG0272)). A usually co-occurring BRCT domain that mediates the DNA contact is absent in gp16. Moreover, blastp hits of lower significance than the best-best hit indicate that gp16 might be a DNA ligase.

Gp23 sequence shows 10 blastp hits with protein sequences from *Streptococcus* phages all annotated as large terminase subunits. Although these hits are not significant (e-values >2) they can hardly occur by chance. The aligned region of gp23 spans a central part from residue 18 to 58 and fits to all subjected sequences approximately between position 140 to 180. Sequence similarities within these regions are high between 31% and 34%. However, this conserved region of gp23 does not match with the large terminase subunit of Dp-1 gp37. Gp23 was annotated as “terminase-like” protein since a sequence link to phage terminases is present.

Gp39 has a detectable zinc finger domain (Pfam domain zf-DNA_Pol, e=0.00011, position 17-84). Thus, it was annotated as “Zinc finger domain protein, putative” (Fig. 19). It could be involved in DNA or RNA binding. Interestingly, the fourth blastp hit aligns with 51% identical residues to an N-terminal region of transcriptional regulator NrdR from *Peptostreptococcus anaerobius* 653-L (ZP_06424961.1, e=0.088). The zinc finger cysteine residues seem to be well-conserved. Moreover, a Pfam signature (e= 0.00046) aligns from position 9 to 143 (Terminase_GpA) and nests the zinc finger domain. The signature is annotated as “Phage terminase large subunit (GpA) family”.

Gp44 contains at its C-terminal sequence region (position 73-111) a Rho termination factor N-terminal domain signature (Pfam: Rho_N, e=0.00021) (Fig. 19) which is presumably a DNA- or RNA-binding domain (Skordalakes and Berger, 2003). It matches within the same sequence region with the blastp best-best hit which is annotated as transcription termination factor Rho from *Chromohalobacter salexigens* DSM 3043. Since only position 73-111 aligned here, the domain annotation was more suitable. In addition, the N-terminus (position 1-25) hits a prokaryotic membrane lipoprotein lipid attachment site profile (Prosite PS51257). Because of this signature constellation the protein was annotated as a “putative Rho-like domain lipoprotein”.

RESULTS

Gp49 shows a high sequence similarity from residue 4-186 to hypothetical protein CLJ_B2521 from *Clostridium botulinum* Ba4 str. 657 (Fig. 19). Notably, its C-terminus is homologous to a wide range of internalins from *Listeria monocytogenes*, including blastp hits with internalin A, F, G with e-values $< e^{-4}$. This region contains the Pfam domain Flg_new (e= 0.00011) from position 221-258. Gp49 was annotated according to the domain and blastp hits as “*Listeria*-*Bacteroides* repeat domain protein”.

Gp55-59 sequences were already published (Sheehan et al., 1997). Annotation of the experimentally characterized lytic components gp58 (holin) and gp59 (Pal, N-acetylmuramoyl-L-alanine amidase) were adopted. Gp55, gp56, and gp57 best-best database hits (namely their own) were ignored and the nearest sequence homologs noted in Tab. 24.

Gp61's best blast hit is a non-specific serine/threonine protein kinase from *Eubacterium eligens* ATCC 27750 (2e-93). However, the protein kinase annotation is probably incorrect due to inferred automated annotation. This is indicated by multiple significant blastp hits that follow and that are annotated as Superfamily II DNA/RNA helicases belonging to the SNF2 family. Thus, the best-best hit was ignored and the second blast hit used in Tab. 24. Domain signatures (InterproScan) that indicate that gp61 is a RNA/DNA helicase are the N-terminal DEXDc domain with ATPase function (SMART domain, 1.2e-5) and the C-terminal helicase signature Helicase_C (Pfam, 4.62e-07). The protein was annotated as “Superfamily II DNA/RNA helicases, SNF2 family”. Speculations about its substrate specificity (if it unwinds RNA or DNA) are currently not possible.

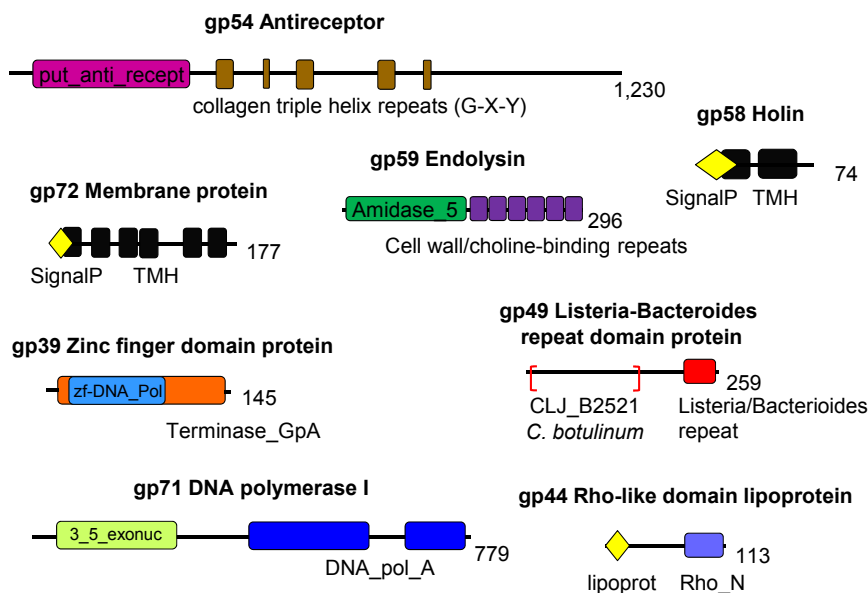


Fig. 19 Sequence features of some Dp-1 proteins

Sequence features are indicated, e.g., signal peptides (SigP), transmembrane helices (TMH), domain signatures, and repeat motifs. The protein length is given to the right of each illustration. In the illustration of gp49, the brackets indicate the blastp alignment region with hypothetical protein CLJ_B2521 of *C. botulinum*. Further explanations can be found in the text.

RESULTS

Tab. 24 Dp-1 ORFs and gene product annotation

(Coordinates) gives the genomic coordinates of identified ORFs. Note that *orf27* to *35* is encoded by the reverse complement. (Protein length) is indicated in the third column by the number of amino acid residues. (Annotation) provides the homology-based annotation of Dp-1 gene products. (Best match) blastp best-best hit is given. The last column (e|qcov|ident) indicates the e-values (e), the query sequence coverage length in percent (qcov), and the percentaged number of identical amino acid residues within this region (ident). As described in the text, conserved domain signatures have been considered for protein annotations. These are documented in the e-supplement.

ORF	Coordinates	Length	Annotation	Best match	e qcov ident
1	131..652	173	Queuosine biosynthesis protein QueF	YkvM <i>Bacillus licheniformis</i> ATCC 14580 (NADPH-dependent 7-cyano-7-deazaguanine reductase QueF)	3e-55 80 74
2	662..1348	228	Queuosine biosynthesis protein QueC	Queuosine biosynthesis protein QueC <i>Fusobacterium sp. D12</i>	4e-59 97 53
3	1350..1871	173	Queuosine biosynthesis protein QueD	QueD <i>Aggregatibacter actinomycetemcomitans</i> D7S-1 (Queuosine biosynthesis protein QueD)	6e-22 76 41
4	1864..2658	264	Queuosine biosynthesis protein QueE	QueE <i>Exiguobacterium sp. AT1b</i> (7-cyano-7-deazaguanosine (PreQ0) biosynthesis protein QueE)	2e-54 90 46
5	2534..3295	253	Tetrahydrofolate biosynthesis protein FolE	GTP cyclohydrolase I <i>Bacillus coagulans</i> 36D1	3e-60 85 53
6	3306..3803	165	Queuosine biosynthesis intermediate transporter QueT	hypothetical protein CLORAM_01593 <i>Clostridium ramosum</i> DSM 1402	5e-22 89 36
7	3793..4728	311	Hypothetical protein	gp31 <i>Burkholderia phage BcepGomr</i>	2e-14 63 30
8	4852..5025	57	No similarity		
9	5040..5279	79	No similarity		
10	5346..6419	357	DNA polymerase III beta subunit DnaN, putative	DNA polymerase III subunit beta <i>Bacillus pumilus</i> SAFR-032	3e-06 64 26
11	6419..7195	258	Cas4 RecB like exonuclease, putative	hypothetical protein EUBELI_10071 <i>Eubacterium eligens</i> ATCC 27750	3e-34 81 39
12	7192..7683	163	Holliday junction resolvase RecU, putative	hypothetical protein EUBELI_10070 <i>Eubacterium eligens</i> ATCC 27750	3e-09 88 29
13	7670..8239	189	Hypothetical protein	hypothetical protein EUBELI_10068 <i>Eubacterium eligens</i> ATCC 27750	1e-09 70 28
14	8208..8699	163	dUTPase	deoxyuridine 5'-triphosphate nucleotidohydrolase family protein <i>Clostridium sp. M62/1</i>	1e-25 96 45
15	8699..9859	386	Recombination protein RecA	RecA <i>Eubacterium eligens</i> ATCC 27750	3e-54 94 33
16	9805..10218	137	NAD-dependent DNA ligase, putative (Lig)	hypothetical protein CLJ_0249 <i>Clostridium botulinum</i> Ba4 str. 657	1e-11 48 47
17	10215..11240	341	DNA polymerase III gamma/tau subunit DnaX	possible DNA-directed DNA polymerase <i>Leuconostoc mesenteroides</i> subsp. <i>cremoris</i> ATCC 19254 (possible DNA-directed DNA polymerase DnaX)	4e-61 73 47
18	11242..12081	279	DNA polymerase III, delta' subunit HolB, putative	DNA-directed DNA polymerase <i>Anaerocellum thermophilum</i> DSM 6725	5e-07 49 31
19	12074..12967	297	DNA polymerase III delta subunit HoloA, putative	hypothetical protein RUMOBE_04030 <i>Ruminococcus obeum</i> ATCC 29174	4e-08 92 26
20	13020..13163	47	No similarity		
21	13160..14404	414	Metal-dependent phosphohydrolase HD, putative	hypothetical protein BACCAP_00466 <i>Bacteroides capillosus</i> ATCC 29799	1e-35 35 42
22	14423..14800	125	Hypothetical protein	hypothetical protein ISM_09311 <i>Roseovarius nubinihibens</i> ISM	0.25 71 28
23	14800..15084	94	Terminase-like protein	terminase large subunit <i>Streptococcus pneumoniae</i> SP19-BS75	2.3 43 34
24	15081..15476	131	Sigma factor (region 4), putative	phage protein <i>Streptococcus pneumoniae</i> TCH8431/19A	0.039 35 46
25	15533..15989	139	Hypothetical protein	hypothetical protein EUBELI_10048 <i>Eubacterium eligens</i> ATCC 27750	5e-12 64 41
26	15973..16284	103	Hypothetical protein	hypothetical protein ORF028 <i>Pseudomonas phage 73</i>	2e-04 72 42
27	16308..16709	133	Hypothetical protein	peptidase M56, BlaR1 <i>Planctomyces maris</i> DSM 8797	0.064 83 28
28	16859..17338	159	Hypothetical protein	periplasmic binding protein/LacI transcriptional regulator <i>Polaromonas naphthalenivorans</i> CJ2	0.35 55 27
29	17521..17715	64	Hypothetical protein	sulfite reductase <i>Sulfolobus acidocaldarius</i> DSM 639	0.13 68 39
30	17712..17891	59	Hypothetical protein	hypothetical protein CRC_01648 <i>Cylindrospermopsis raciborskii</i>	2.8 91 39
31	18026..18778	250	Hypothetical protein	PrfB <i>Pantoea ananatis</i> LMG 20103 (Peptide chain release factor 2)	2.0 91 22
32	18780..19151	123	Hypothetical protein	GA19226 <i>Drosophila pseudoobscura pseudoobscura</i>	0.2 61 35
33	19161..19529	122	Hypothetical protein	dynein, axonemal, heavy polypeptide 3 <i>Pan troglodytes</i>	1.4 72 27
34	19423..19995	190	Hypothetical protein	hypothetical protein Dtox_2450 <i>Desulfotomaculum acetoxidans</i> DSM 771	9e-14 92 31

RESULTS

ORF	Coordinates	Length	Annotation	Best match	e qcov ident
35	20094..20411	105	Phosphotransferase KptA/Tpt1, putative	hypothetical protein RUMGNA_00196 <i>Ruminococcus gnavus</i> ATCC 29149	3e-06 85 32
36	21512..22252	246	Hypothetical protein	predicted protein <i>Paenibacillus</i> sp. oral taxon 786 str. D14	0.37 41 27
37	22230..23621	463	Terminase, large subunit	phage terminase, large subunit, PBSX family <i>Clostridium lentocellum</i> DSM 5427	8e-120 94 47
38	23674..25434	586	Portal protein, SPP1 Gp6-like	hypothetical protein CTC01087 <i>Clostridium tetani</i> E88	8e-74 87 33
39	25340..25777	145	Zinc finger domain protein, putative	hypothetical protein CTC01098 <i>Clostridium tetani</i> E88	0.003 54 33
40	25743..26738	331	Minor capsid protein, putative	conserved hypothetical protein <i>Clostridium botulinum</i> E1 str. 'BoNT E Beluga'	2e-40 92 34
41	26943..27611	222	Minor capsid protein	similar to phage phi1e RORF204 (minor capsid protein) <i>Staphylococcus phage phiETA</i>	4e-10 65 30
42	27627..28004	125	Hypothetical protein	hypothetical protein Parl_06935 <i>Paenibacillus larvae</i> subsp. <i>larvae</i> BRL-230010	0.9 72 29
43	28017..29096	359	Major capsid protein	phage protein <i>Paenibacillus larvae</i> subsp. <i>larvae</i> BRL-230010	4e-73 98 42
44	29108..29449	113	Rho-like domain lipoprotein, putative	transcription termination factor Rho <i>Chromohalobacter salexigens</i> DSM 3043	1.9 31 50
45	29451..29768	105	Hypothetical protein	hypothetical protein CBO2348 <i>Clostridium botulinum</i> A str. ATCC 3502 (Hypothetical phage protein)	6e-07 95 34
46	29765..30154	129	Hypothetical protein	hypothetical protein CLJ_B2524 <i>Clostridium botulinum</i> Ba4 str. 657	0.03 79 27
47	30154..30507	117	Hypothetical protein	hypothetical protein Thit_1383 <i>Thermoanaerobacter italicus</i> Ab9	3e-12 85 41
48	30516..30893	125	Hypothetical protein	extracellular solute-binding protein, family 1 <i>Burkholderia multivorans</i> CGD1	8.2 68 28
49	30896..31675	259	Listeria-Bacteroides repeat domain protein	hypothetical protein CLJ_B2521 <i>Clostridium botulinum</i> Ba4 str. 657	8e-41 70 47
50	31699..32154	151	Hypothetical protein	hypothetical protein CLI_3089 <i>Clostridium botulinum</i> F str. <i>Langeland</i>	4e-09 89 29
51	32120..32341	73	Hypothetical protein	hypothetical protein CLD_1510 <i>Clostridium botulinum</i> B1 str. <i>Okra</i>	0.056 72 29
52	32386..35835	1149	Tail length tape measure protein	phage minor tail protein <i>Enterococcus faecium</i> 1,231,410	4e-160 81 37
53	35847..36686	279	Tail protein, putative	cellulose synthase-like family C1 protein <i>Hordeum vulgare</i> subsp. <i>vulgare</i>	0.26 14 47
54	36698..40390	1230	Antireceptor	ORF38 <i>Streptococcus</i> phage 7201	1e-152 64 46
55	40401..42440	679	Tail protein, putative	phage structural protein <i>Streptococcus pneumoniae</i> SP18-BS74	0 100 59
56	42490..42759	89	Hypothetical protein	hypothetical protein CGSSp18BS74_10714 <i>Streptococcus pneumoniae</i> SP18-BS74	1e-21 87 66
57	42774..43202	142	Hypothetical protein	hypothetical protein SAG1839 <i>Streptococcus agalactiae</i> 2603V/R	1e-31 95 49
58	43189..43413	74	Holin	holin bacteriophage Dp-1	4e-34 100 100
59	43413..44303	296	Endolysin Pal	endolysin Pal bacteriophage Dp-1	2e-172 100 100
60	44595..45299	234	Hypothetical protein	endonuclease, putative <i>Streptococcus pneumoniae</i> SP3-BS71	0.004 41 28
61	45350..46987	545	Superfamily II DNA/RNA helicases, SNF2 family	Superfamily II DNA/RNA helicases, SNF2 family <i>Eubacterium rectale</i> DSM 17629	3e-93 85 41
62	46956..47201	81	Sigma 70 subunit, putative	RNA polymerase, sigma 70 subunit, RpoD family <i>Shewanella</i> sp. W3-18-1	1.9 43 47
63	47200..47541	113	Resolvase domain protein, putative	Resolvase domain protein <i>Cyanothece</i> sp. PCC 7424	2.0 81 26
64	47542..47961	139	Hypothetical protein	hypothetical protein DET0620 <i>Dehalococcoides ethenogenes</i> 195	0.25 47 25
65	48082..48561	159	Hypothetical protein	unnamed protein product <i>Kluyveromyces lactis</i>	2.8 71 25
66	48718..49362	214	DNA replication protein DnaC	DNA replication protein <i>Eubacterium rectale</i> DSM 17629	8e-21 91 31
67	49624..50961	445	Replicative DNA helicase DnaB	replicative DNA helicase <i>Eubacterium eligens</i> ATCC 27750	3e-59 93 38
68	50955..51974	339	DNA primase DnaG	DNA primase <i>Eubacterium eligens</i> ATCC 27750	5e-41 97 31
69	52033..52647	204	Sporulation sigma factor SigK, putative	sporulation sigma factor SigK <i>Thermoanaerobacter pseudethanolicus</i> ATCC 33223	9e-07 69 26
70	52762..53490	242	Hypothetical protein	hypothetical protein EUBELI_10080 <i>Eubacterium eligens</i> ATCC 27750	6e-20 73 32
71	53538..55877	779	DNA polymerase I	DNA polymerase 1-3'-5' exonuclease and polymerase domains <i>Eubacterium rectale</i> DSM 17629	2e-114 80 41
72	55855..56388	177	Membrane protein, putative	hypothetical protein BFZC1_22719 <i>Lysinibacillus fusiformis</i> ZC1	2e-13 93 30

tRNAs

The Dp-1 genome sequence was searched for tRNAs using tRNAscan-SE (Schattner et al., 2005). No tRNA genes were detectable, indicating that Dp-1 does not introduce self-encoded tRNAs into the host as other phages like coliphage T4 and T5 do (Scherberg and Weiss, 1972).

Genomic context, prediction of gene modules, and a predicted transcriptional map

Phage genomes are usually organized as gene modules (other terms are operons or gene batteries). Inside such a module, ORFs of corresponding gene products that function in a similar cellular process are co-localized. From single transcriptional start sites these ORFs are co-expressed as polygenic mRNAs. In phages transcription of these batteries is regulated stringently post infection. Shortly after infection operons are expressed that encode for proteins responsible for DNA replication which is an early event. Later, operons encoding structural phage proteins and maturation factors are expressed. Finally, lytic enzymes occur and lead to the cell's burst. This hierarchic principle of phage gene expression allows the definition of functional gene clusters based on the function of known ORF products located in a gene battery and also allows a rough prediction when these proteins reach their maximal expression levels. Furthermore, if ORFs of unknown gene products are co-localized within a certain operon of known function, these ORFs likely function in the same cellular process.

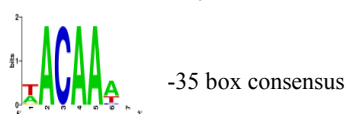
Promoter prediction: To get a first insight into the transcriptional organization of the Dp-1 genome I looked for putative transcriptional start sites by using various prediction tools (2.2.6). However, the results were not clear due to common problems with prediction of false-positive and -negative nucleic acid elements. I looked manually for sigma70 consensus sequences (TATAA(T) -10 box and TTGACA -35 box consensus sequences) in the intergenic regions of predicted ORFs. Several -10 box elements were identified but surprisingly only one in the context with a conserved -35 consensus sequence (promoter 10 (P10)). In total, 12 highly conserved TATA-boxes and thus 12 promoters were found upstream of ORFs that indicate the presence of putative transcription start sites (Tab. 25). Because of the lack of a common -35 TTGACA consensus sequence I checked for presence of another common -35 element. All promoters (except P10) exhibit the consensus sequence "ACAA" in -35 regions. Presumably, this motif could represent a Dp-1 specific -35 element. Since P10 is the only promoter with a conserved -10 and -35 box element Dp-1 genome transcription could be initiated from this site by the host sigma70 factor. Interestingly, the downstream module encodes a putative sigma70 factor that could jump over to activate expression of other Dp-1 operons. However, this is speculative and no reliable gene-product indicators are localized within this gene battery that suggest that this module can be transcribed in the early phase.

RESULTS

Tab. 25 Predicted promoters

(P) indicates the promoter number. (Pos-35) and (Pos-10) indicate the genomic position of -35 and -10 boxes, respectively. (Promoter sequence) element core consensus sequences are labeled in bold. Note, that P10 is the only promoter that contains a conserved sigma70 -35 box consensus sequence. Promoter 1 and 6 contain a double TATA-box motif. Promoter 6, 9, and 12 exhibit an extended spacer sequence between the predicted -35 and -10 region. (ORF) indicates the first ORF which is localized downstream the promoter. Based on the function of module-encoded gene-products a general module function and expression timing is assumed in the last two columns. The weblogo at the tabel's bottom indicates the conservation degree of the nucleotide composition of the putative -35 box (P10 was not considered). Image was made with Weblogo (Crooks et al., 2004).

P	Pos-35	Pos-10	Promoter sequence		ORF	Module function	Expression timing
			-35 box	-10 box			
1	43	64	AACAAA A- N ₁₅	- TATAATATAAG	1	Queuosine related	early
2	4778	4811	TACAAA A- N ₂₇	- TATAA T	8	DNA replication related	early
3	12967	12998	AACAAA T- N ₂₅	- TATAAG	20	Virion - DNA packaging	middle
4	21454	21431	TACAA T- N ₁₇	- TATAA T	35	Unknown	unknown
5	21474	21493	TACAA TG- N ₁₂	- TATAA A	36	Virion - DNA packaging	middle
6	23587	23647	TACAAA T- N ₅₃	- TATAATATAAG	38	Virion - capsid	late
7	29660	29688	AACAAA T- N ₂₂	- TATAAG	46	Virion - tail	late
8	35762	35782	AACAA TG- N ₂₉	- TATAA T	53	Virion - antireceptor	late
9	43059	43107	TACAA CT- N ₃₉	- TATAAG	58	Lysis	late
10	44524	44557	TTGACAA - N ₁₆	- TATAA T	60	Unknown	unknown
11	48600	48621	TACAAA A- N ₁₅	- TATAA T	66	DNA replication related	early
12	52693	52745	GACAA AC- N ₄₅	- TATAA A	70	DNA replication related	early



Terminator prediction: In bacteria transcriptional termination is mediated by two mechanisms, namely Rho-dependent and Rho-independent (intrinsic) termination. Rho-independent termination is achieved by hairpin structures that fold in parallel to RNA synthesis on the RNA strand and lead to the stall of the RNA polymerase on its substrate. These sites are predictable (Lesnik et al., 2001). Remarkably, in *S. pneumoniae* (and few bacterial species else) no Rho factor homolog is present while it is highly conserved in most bacterial species (Washburn et al., 2001). Rho is the key factor for Rho-dependent termination. The absence of an ortholog in *S. pneumoniae* suggests that this bacterium deals with transcriptional termination exclusively by the intrinsic strategy. However, termination/anti-termination factors NusA, B, and G orthologs are present in the *S. pneumoniae* genome (Fraser et al., 2000). The Dp-1 genome also does not encode for a well-conserved Rho homolog (except gp44, a lipoprotein with a partial Rho factor domain, see above), making it likely that the phage uses exclusively Rho-independent transcription termination.

Regarding these findings I screened the Dp-1 genome for intrinsic terminators *in silico*. Each ORF was checked towards the presence of short hairpin structures (loop length cutoff was chosen as 3 to 10 nt in length and the stem 4 to 18 bp in length) with mFold (Zuker, 2003) 20 bases up- and 180 bp downstream of ORF stop codons. Putative candidates were checked with the Tm Server (Zuker, 2003)

RESULTS

and hairpins were considered with $\Delta G \leq -4$ kcal/mol and a melting temperature $\geq 70^\circ\text{C}$. Furthermore, I checked for the presence of an 11 base 5'-A rich and 12 base 3'-T(U)-rich region which flank the hairpins of Rho-independent termination sites (Lesnik et al., 2001). In total, 11 putative Rho-independent termination sites were retrieved in the Dp-1 genome (Tab. 26, Fig. 17). All of them contain the T-regions. However, in some cases the A-regions contain less adenosine as suggested for termination sites of *E. coli* (Lesnik et al., 2001).

In most cases the predicted terminators fit well with the promoter context, e.g., the queuosine cluster that is embraced by a single promoter and terminator (Fig. 17). In other cases the predicted terminators disrupt expected transcriptional read-throughs, e.g., downstream *orf40* or *orf35* (Fig. 17). These could be over-read by transcriptional anti-termination. Experimental evidence comes from the *Pneumococcus* phage Cp-1. Martín and colleagues predicted hairpins and determined transcripts experimentally (Martin et al., 1996b). Mostly, the transcripts that were detected interfered with predicted terminator positions. This indicates that a read-through must occur at certain termination sites, although the anti-termination mechanism is unknown. The authors did not focus on this issue. Moreover, in the Dp-1 genome some intrinsic terminators are unexpectedly absent, e.g., directly downstream the large terminase subunit (*gp37*) (Fig. 17). Putative hairpins were also detectable in such cases but were clearly below the threshold.

Tab. 26 Predicted Rho-independent termination sites

(ORF) and (ORF end) the next ORF upstream the predicted terminator and genomic position of its stop codon. (Terminator position) indicates the genomic positions of the terminator. The fourth column shows the terminator sequence. The letter format given in the headline corresponds to the base residues. These are separated by “-“. The mRNA sequence is given. A 3'-spacer region can be 0, 1, or 2 bases in length. The last two columns list the calculated free energies (ΔG) and melting temperatures (T_m) for the corresponding stem stretches. Note, that the termination site after *orf45* contains a double hairpin structure surrounded by an A- and T-region.

ORF	ORF end	Terminator position	A region		hairpin		T region			ΔG kcal/mol	T_m °C
			A region	stem/loop	stem	spacer	T box proximal	T box distal	extra T box		
7	4728	4719..4760	ACU	UUUUU	AGA-U	AAGAGC	UUUUCG	CUCUUA	--UUUUU-UUUA-AAA	-8.2	81.9
26	16284	16272..16313	GCA	AAAAU	AAAU-	AGACCU	UUUCU	AGGUCU	-A-UUUUU-AUUA-UUG	-7.7	80.7
40	26738	26728..26770	AA	AUU	AUUAG-	GCUCGG	UUCAA	UACCGAGU	-C-UUUUU-GUCU-AUA	-6.2	80.3
45	29768	29880..29954	AAA	AGGAU	AAAA-	GCGAAUG	AAGUCGU	AGCAGACG	ACCUUGUUUGU -U-UAGUUGAUUUUCAACUG--UCCU-GACC-UUU	-19.0	84.0
57	43202	43316..43355	CCA	UUGCAC	UU-	CUUGCA	ACUUUG	CAGG--	UACUG-UUCU-AGG	-6.2	74.1
59	44303	44300..44341	UU	AAAAU	AUAG-	AGAGGAGG	AAGCUCUUUU	-C-UUAAU-AUUG-UUU		-4.9	70.5
70	53490	53483..53521	GAU	UCUAACA	U-	GAGGGCG	CGAGCC	CCU--UUUAU-UAUU-GAU		-12.4	100.6
72	56388	56434..56474	UUG	AAAAAG	UA-	GUCAGGA	AAAAU	UCCUGAU--UAUUU-UUUU-UAC		-8.1	79.7
27	16308	16311..16270	AU	AAU	AAAAAU-	AGACCU	AGAAA	UAGGUCU	-A-UUUAU-UUUG-CUU	-10.4	94.6
28	16859	16856..16828	GC	UAAACU	AAAA-	AGCUCU	AUUUU	UAGGGCU--UUUUA-UUUU-CAA		-7.1	79.6
35	20094	20167..19986	CG	AGUGAC	GAA-	GAGGGG	CAAC	CUGCCCUUC--UUAUG-GCUC-AAA		-11.6	90.6

3.1.2 Some remarks on Dp-1 protein systems

Homology-based annotation of the Dp-1 proteome was suitable to identify certain functions for many putative gene products. Here, some general remarks about Dp-1 protein systems are made.

Queuosine biosynthesis enzymes

A surprising finding was that Dp-1 encodes for a wide and highly conserved set of Que enzymes that are involved in the synthesis of queuosin (Q). Q is a hypermodified nucleoside derivative of guanosine (7-deazaguanosine) occupying the anti-codon wobble position of the tRNAs specific for Asp, Asn, His, and Tyr. It is present in these tRNAs in bacteria and eukaryotes and represents the most complex nucleotidyl modification that is known (Iwata-Reuyl, 2003; Morris and Elliott, 2001). Eukaryotes get Q from their diet or intestinal flora but in prokaryotes Q is synthesized *de novo* from GTP via a complex series of reactions that are not fully understood. The consequence of Q in the mentioned tRNAs is to improve codon/anti-codon base pairing and thus the accuracy and efficiency of translation (Meier et al., 1985).

A bacteriophage specific BLAST search on ACLAME database V0.4 (Leplae et al., 2010) was done using as query the protein sequences of the queuosine biosynthesis cluster of bacteriophage Dp-1 (gp1 to gp7). Numerous bacteriophages infecting diverse bacterial hosts encode genes needed to produce enzymes for queuosine biosynthesis (Tab. 27) but none of these bacteriophages seems to contain a complete queuosine-synthesis cluster. The queuosine cluster of Dp-1 is apparently the most complete of bacteriophage queuosine clusters even if it lacks the putative gene encoding the protein QueA (see below). The bacteriophage Dp-1 contains the genes of queuosine precursor transporter QueT (orf6), four genes for QueF, QueE, QueC, and QueD, (orfs 1, 2, 3, 4) needed for the synthesis of the queuosine precursors preQ0 and preQ1 and a putative tRNA-guanine-transglycosylase (TGT, gp7) needed for the replacement of guanine at the wobble position of tRNA with the queuosine precursor preQ1. TGT was detectable in a PSI-BLAST in the third iteration with the best-best hit of *Ignicoccus hospitalis* KIN4/I TGT (e-value=3e-12). Regarding the operon context gp7 might represent a TGT homolog with low sequence conservation compared to annotated TGT homologs. Thus, gp7 was not considered in the previous sections as TGT and was annotated as hypothetical protein.

Dp-1 lacks the S-adenosylmethionine/tRNA ribosyltransferase-isomerase QueA which is required for the transformation of the tRNA associated preQ1 into epoxy-queuosine (epoxyQ). However, QueA as well as others Q-related enzymes (Tab. 27) are encoded by *S. pneumoniae* which on the other hand lacks QueC, D, E, and T homologs and has a TGT showing no similarity to the Dp-1 encoded TGT.

The presence of the queuosine cluster is thus possible due to two related reasons: the need for tRNA modification to ensure the accuracy of bacteriophage translation and a possible use of a bacteriophage specific queuosine-tRNA modification system to monopolize the tRNAs specific for Asp, Asn, His, and Tyr and divert these tRNAs from the host protein synthesis.

RESULTS

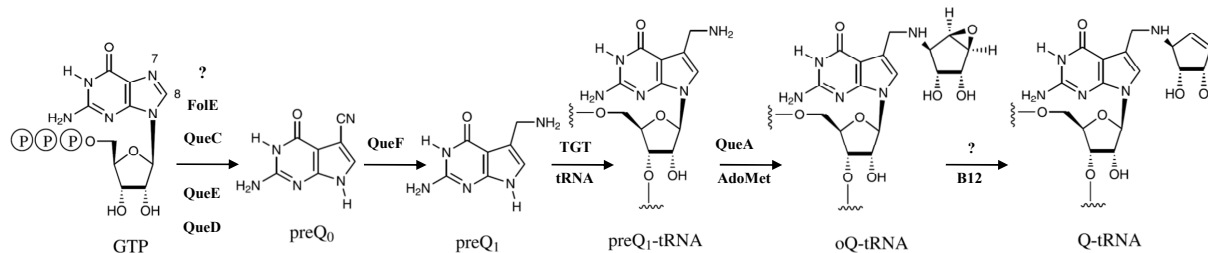


Fig. 20 Q biosynthesis

A simplified queuosine biosynthesis pathway is given. Responsible enzymes and cofactors for the corresponding synthesis steps are indicated. Question marks label the participation of putative unknown enzymes. The figure was drawn after (Iwata-Reuyl, 2003).

Tab. 27 Que synthesis enzymes of Dp-1

Presence or absence of Dp-1 homologs are given by the gp number in the first column, *S. pneumoniae* (SPN) Q enzymes are indicated in the last column. The second column lists the number of ACLAME blast hits of phage-related Que enzymes. In the fourth column the putative function of the corresponding enzyme is given. Descriptions were adopted from corresponding *E. coli* homologs provided by ECOCYC database (<http://ecocyc.org/>).

Dp-1 gp	Number ACLAME hits	Dp-1 Q biosynthesis protein	Putative function in queuosine biosynthesis	SPN homolog
1	1	QueF	NADPH-dependent nitrile oxidoreductase activity, forming a primary amine, catalyzes a later step of queuosine biosynthesis, the formation of preQ ₁ from preQ ₀ . Also known as 7-cyano-7-deazaguanine reductase.	SP1777
2	19	QueC	Biosynthesis in the step(s) leading from GTP to the formation of preQ ₀	-
3	7	QueD	Converts 7,8-dihydroneopterin triphosphate (H ₂ NTP) to 6-carboxy-5,6,7,8-tetrahydropterin (CPH4), CPH4 is likely an intermediate in the pathway of queuosine biosynthesis	-
4	12	QueE	Exact function unknown, involved in the biosynthesis of preQ ₀ with QueD and QueC	-
5	11	Tetrahydrofolate biosynthesis protein FoIE	GTP cyclohydrolase I is an allosteric enzyme that catalyzes the first step in the biosynthesis of tetrahydrofolate. A foIE deletion mutant lacks the queuosine ribonucleoside in tRNAs, indicating that foIE is required for preQ ₀ biosynthesis from GTP	SP0291
6	0	Queuosine intermediate transporter QueT	Membrane transporter for preQ ₀ precursors	-
7	1	tRNA-guanine transglycosylase (TGT), putative	Multimeric enzyme ensuring the replacement of tRNA guanine at the position 34 with the queuosine precursor preQ ₁	SP2058
-		S-adenosylmethionine tRNA ribosyltransferase-isomerase QueA	The queA gene encodes an S-adenosylmethionine:tRNA ribosyltransferase-isomerase, which catalyzes formation of the 2,3-epoxy-4,5-dihydroxycyclopentane ring of epoxyqueuosine (oQ)	SP1416

DNA replication and recombination systems

The Dp-1 genome encodes for a wide set of proteins that are involved in DNA replication and recombination:

DNA polymerase I: DNA polymerase I is encoded by *orf71*. It is highly similar to DNA polymerase I from *Eubacterium rectale* (Tab. 24). It consists of the C-terminal 5'-3' polymerase domain. However, this signature is interrupted between position 609 to 680. The 3'-5' proofreading exonuclease domain is present at the N-terminus (Fig. 19). *E. coli* DNA polymerase I is involved in DNA repair and usually functions in filling up nucleotide lesions on DNA molecules. It acts in nucleotide excision by removing UvrABC nuclease complex (Orren et al., 1992), MutHLS-mediated base excision, replacement of thymidine dimers (Dorson et al., 1978), and in post-replicative repair of gaps that occur in RNA priming regions (replication is done by DNA polymerase III) and double strand breaks (Sharma and Smith, 1987). Thus, Dp-1 DNA polymerase I might function as repair DNA polymerase also during Dp-1 genome replication. Efficient DNA replication could be facilitated by DNA polymerase III.

DNA polymerase III: Several homologs of DNA polymerase III subunit were detected in the Dp-1 genome (Fig. 21B and C). Notably the core polymerase subunits are not present. This suggests that the host core subunits might be used while Dp-1 encodes its own clamp loader subunits. As a consequence, a hybrid DNA polymerase III complex could manage Dp-1 genome replication. However, neither in Dp-1 nor in *S. pneumoniae* TIGR4 homologs are present for the subunit theta (holE), psi (holD), and chi (holC). Absence of these subunits in the TIGR4 genome might be substituted by homologs with low sequence conservation. Other factors that are necessary for DNA polymerase III dependent replication are given in Fig. 21B. Five of them are encoded by the virus and the host genome. In contrast, the Dp-1 genome does not encode for an RNaseH homolog (removes RNA primers), DnaA (initiates replication), and Ssb (binds and stabilizes ssDNA strands). While Dp-1 self-encoded proteins could be involved in Dp-1 genome replication, the uniquely encoded *S. pneumoniae* replication proteins might be used additionally for creation of a functional and processive DNA polymerase III system.

Origin of replication: The genome sequence region around bp 20,695 exhibits a very low G+C content of 15% and is of particular interest because a putative origin of replication is detectable between bp 20,663 and 21,223 (Fig. 17, *ori1*). Here, three degenerated DnaA boxes (TTATCCACA consensus, with two mismatches per box) are present at positions 20,793, 20,805, and 20,837 (Ori-Finder web-based application (Gao and Zhang, 2008)). The first DnaA-box locates in the reverse strand which coincides with a non-coding region of the reverse transcribed genomic cluster. Interestingly, another potential origin of replication has been previously reported in the Dp-1 genome by García et al. between the positions 56,348 and 56,464 bp where two perfectly identical direct repeats have been

RESULTS

found (TTA TCT TGC AGT CAA TTG CTT CGA GAT ATT TGA AAA AGT AGT CAG GAA AAT TCC TGA TTA T with a TTAT overlap) (García et al., 2005) (Fig. 17, *ori2*). It is noteworthy that no DnaA box is detectable in this region. It is possible that both putative origins are functional and that their presence reflects an unknown aspect of Dp-1 genome replication.

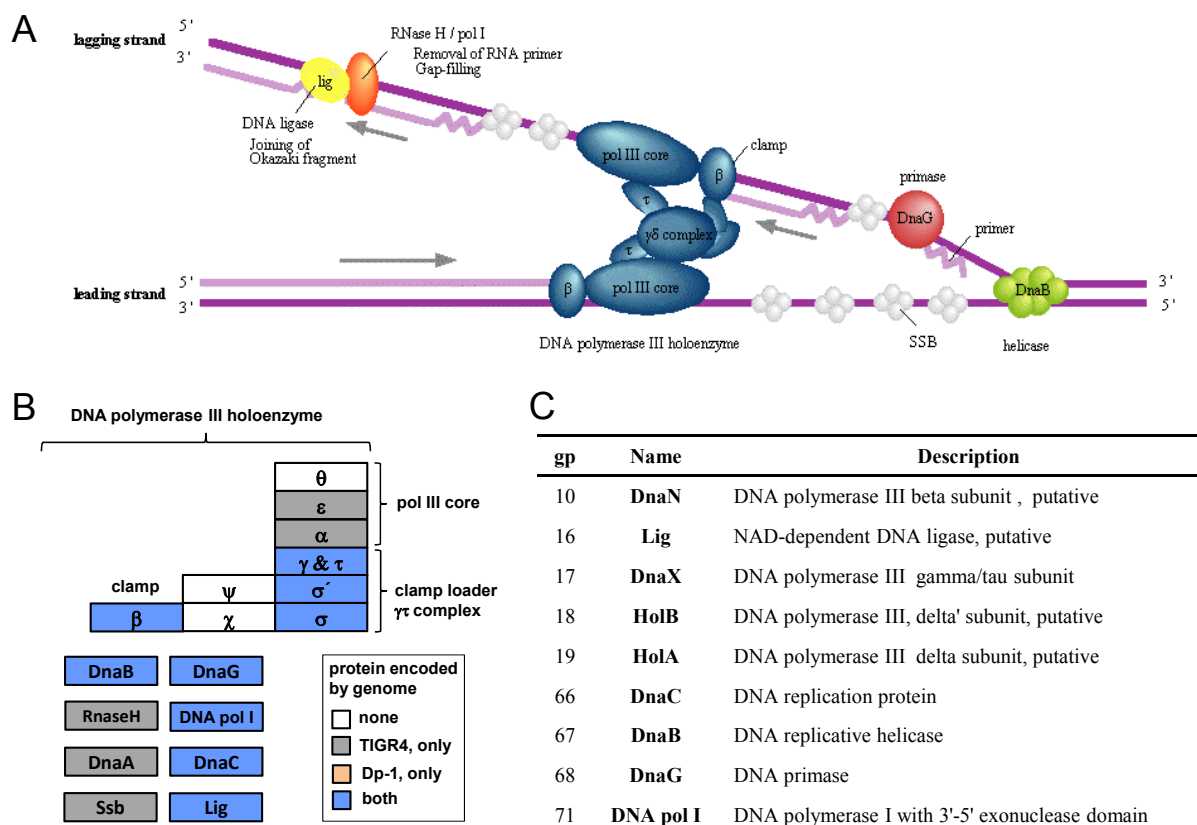


Fig. 21 Comparison of the DNA polymerase III system of Dp-1 and its host

(A) Illustration of a canonical DNA polymerase III system. Image taken from KEGG database (reference pathway entry “DNA replication”). (B) Subunit composition of DNA polymerase III system of Dp-1 and *S. pneumoniae* (after KEGG database, see above). The core and holoenzyme subunits are given. Additional factors that play a role are given below. Different colors indicate the genomes the subunits are encoded by. *S. pneumoniae* TIGR4 genome was used as host reference strain. (C) Homology-annotated Dp-1 proteins that might participate in DNA polymerase III-dependent DNA replication. Column “gp” indicates the systematic gp number. DNA pol I is shown since it is involved in removing RNA primers from the lagging strand by its exonuclease function after the action of RnaseH onto RNA primers (Ogawa and Okazaki, 1984).

DNA recombination: Four putative recombination protein homologs were detected in the Dp-1 genome. Since Dp-1 is a lytic phage and does not undergo the integration of its DNA molecule into the host genome the reason for the presence of these proteins is uncertain. Gp11, gp12, gp15, gp63 encode for putative recombination proteins (Cas4/RecB-like exonuclease, Holliday junction resolvase RecU, RecA, and a resolvase domain protein, respectively). Presence of these proteins eventually underline that homologous recombination could be an important process for Dp-1. This could take place during or after replication of the Dp-1 genome since here intermediary structures are known, e.g., from T4 phage to appear on the concatemeric DNA that necessarily had to be resolved prior to the DNA packaging event (see below). In parallel homologous recombination could be necessary for DNA repair and to restart stalled replication forks. Moreover, the need for DNA-recombination

systems might be due to Dp-1 making use of a recombination dependent replication model alternatively to the *ori*-dependent model. T4 makes use of such a replication model: essentially, a ssDNA 3' genome end is invaded into a homologous dsDNA region forming a D-loop. From here the D-loop is converted into a replication fork while the ssDNA 3' end is used as primer for leading and the opposite strand for lagging strand synthesis. This results in intermediary DNA structures at the initiation site that have to be resolved (Kreuzer, 2000; Martinez-Jimenez et al., 2005; Mosig et al., 2001).

No integrase homolog, the indicator for temperate phages, was detected in the Dp-1 genome in agreement with the findings that Dp-1 is exclusively lytic (Lopez et al., 1977).

Structural components of Dp-1

The Dp-1 genome encodes for a large set of putative structural protein homologs (details can be found in Tab. 24). Homologous head proteins were found: the major head protein (gp43) is the major structural head component. The portal protein gp38 belongs to the SPP1 portal protein family. A portal protein complex normally appears at one vertex of the capsid and forms a hole through which DNA is packed and the contacts with tail components are mediated. Homologs of minor head proteins (gp40, gp41) could function as minor structural components of the capsid or as scaffolding proteins during head assembly. Several tail proteins were found including putative structural components like gp53 and gp55. The tail length tape measure protein gp52 could act as the tail length ruler during tail assembly. The tail tip antireceptor gp54 presumably mediates the attachment to the host cell. It consists of an N-terminal putative antireceptor domain and several collagen G-X-Y triple motifs (Fig. 19). Antireceptor proteins from *Siphoviridae* phages infecting low-GC-content gram-positive bacteria usually contain repeated G-X-Y motifs at their C-termini (Lucchini et al., 1998). Dp-1 antireceptor is supposed to bind to phosphatidyl-choline in the lipoteichoic acid. This motif appears to be characteristic of tropocollagen molecules and its biological function is to provide elasticity. The large terminase subunit is encoded by *orf37* and mediates the DNA translocation into the pro-capsid during DNA packaging. A Dp-1 homolog of a small terminase subunit is notably absent.

Dp-1 DNA transcription

Many phages encode their own RNA polymerase. For instance, T7 phage uses the host RNA polymerase early after infection and later its own more processive polymerase while inhibiting the host RNA polymerase (Nechaev and Severinov, 1999). In the Dp-1 genome no homologs of RNA polymerases or RNA polymerase core subunits could be detected. Presumably, Dp-1 transcription is dependent on the *S. pneumoniae* host RNA polymerase system. However, three sigma factor homologs were identified. Gp24 contains a sigma factor 4 region which is known to mediate contacts with the -35 promoter box of sigma 70-like sigma factors (Campbell et al., 2002). Gp62 is another putative sigma 70 factor and gp69 is a homolog of the sporulation sigma factor (SigK) group. All three ORFs are localized in different gene modules. No transcriptional repressor homologs have been detected for

Dp-1. This indicates that Dp-1 might regulate its transcription just by its sigma factors using the host RNA polymerase.

Lytic cluster

The Dp-1 genome encodes for the two components that are needed for the lytic step: the membrane protein holin is encoded by *orf58*. It contains a predicted signal peptide and two transmembrane regions (Fig. 19). Holins with two transmembrane helices are classified as S²¹ family holins in contrast to the second holin family S^λ whose members consist of three transmembrane regions (Wang et al., 2000). The lytic enzyme is the amidase Pal (gp59). It has an N-terminal amidase domain and C-terminal choline cell wall binding repeats (Fig. 19). More detailed explanations about Pal can be found in the introduction section.

Other notable gene products

These are membrane-associated proteins of unknown function: *orf72* encodes for an unknown membrane protein containing six putative transmembrane helices (Fig. 19). Gp44 contains a predicted lipoprotein membrane anchor signature (Fig. 19). Since Dp-1 contains lipids as structural components in its head (Lopez et al., 1977), membrane proteins might be necessary to mediate the integration of the host membrane during head assembly. While *orf44* is localized in the gene module that encodes for structural head proteins, *orf72* is co-localized with the DNA polymerase I gene. Here, the functional link is missing on the genomic level and gp72 might be responsible for any other task than virion assembly.

3.1.3 Intra-viral protein-protein interaction networks of Cp-1 and Dp-1

The sum of all PPIs of an organism (interactome) helps to functionally link proteins to certain cellular pathways. In this section I describe the intra-viral protein-protein interaction (PPI) networks of the lytic bacteriophage Cp-1 and Dp-1 that infect *S. pneumoniae*. Phage PPI networks are strongly underrepresented in the literature and since publication of the intra-viral T7 interactome in 1996 (Bartel et al., 1996) no phage interactome has been explored. These results help not only to extend and establish bacteriophages in interactomics but also give novel insights into the interplay of individual proteins of these phages.

To screen towards the Cp-1 and Dp-1 interactome, the Yeast Two-Hybrid- (Y2H) system was used since it is an excellent method to screen an organism's proteome reliably and quickly. Therefore, I first cloned all ORFs as full-length ORF constructs of both phages into an entry vector library based on the site-specific Gateway® recombination system (Invitrogen) (2.2.1.3). For the screens I used two different Y2H expression vector systems (pDEST32/pDEST22 (Invitrogen) and pGBKT7g/pGADT7g (Clontech)) since it was shown already that the application of alternative expression vector systems can clearly increase the screen sensitivity and thus the number of detected PPIs (Rajagopala et al., 2009). The individual phage baits were screened in one-on-one Y2H matrix tests against their corresponding prey array by using a robot-supported workstation (Biomek2000, Beckman Coulter) (2.2.5.6).

The binary Dp-1 interactome

The primary screens detected a total of 232 redundant interactions among the 72 Dp-1 proteins: 99 were identified with the pDEST32/22 and 133 with the pGBK/GADT7g expression vector system. I retested all these interactions by Y2H retest experiments (2.2.5.9) and reproduced a total of 69 (69.7%) of the pDEST32/22 and 109 (82.0%) of the pGBK/GADT7g protein pairs (Fig. 22). In total 156 unique, non-redundant³ intra-viral protein-protein interactions were reproducible. On average each Dp-1 protein binds ~2.2 interaction partners and 57 proteins exhibit at least one interaction. Thus, a large fraction of 79.2% of the proteome participates in the interaction network (Fig. 23). The full list of identified intra-viral Dp-1 PPI data is given in the Tab. 35 in the supplement; a full list of redundant PPIs can be found in the e-supplement.

The pDEST32/22 system identified 49 (31.4%) and the pGBKT7g/pGADT7g system identified 92 (59.0%) unique interactions of the 156 non-redundant interactions. Surprisingly, only 15 interactions overlapped between both vector systems (9.6%). This number confirms the findings of Rajagopala and colleagues although the overlap found in this study is about three times larger (Rajagopala et al., 2009). With the pGBKT/GADT7g expression system I obtained nearly twice as many unique

³ “non-redundant” means that (i) positively tested protein pairs are counted as one interaction while they can appear in both vector systems or (ii) they can appear among their different fusion proteins since the proteins are tested twice in one vector system as DBD- and AD-fusions. This can lead to a redundant interaction if a test pair binds as DBD- as well as AD-tagged protein with a certain counterpart

RESULTS

interactions (Fig. 22A). However, with the pDEST system the more specific interactions were detected. The pDEST32/22 interaction network consists of 51 nodes connected by 64 edges. This corresponds to an average node degree of 1.26 (the node degree is the number of interactions a certain protein has). In contrast the pGBKT7g/GADT7g network contains 46 proteins connected via 107 interactions. Here, a protein binds to 2.33 partners in average. Although the pDEST system detected fewer PPIs, it included five additional proteins in the network and thus may be more specific. It is unknown why such great differences in the interaction patterns occur with the usage of alternative Y2H expression vector systems. It is assumed that the plasmid copy number has an impact on the final protein level (*2μ ori* produces high copy numbers in pGBKT7g/pGADT7g compared to a low copy number of centromeric *ori* in the pDEST plasmids) or variations in the DBD- and AD-linker regions could lead to differed steric properties of the fusions (Rajagopala et al., 2009).

In the retest experiments I tested yeast against various concentrations of 3-AT in the readout medium (0, 0.1, 0.25, 0.5, 1, 2.5, 5, 10, 25, and 50 mM) and introduced a scoring scheme that helps to compare the quality and stability of the PPIs within this dataset. 3-AT is a competitive inhibitor of the reporter gene product HIS3 (Hilton et al., 1965) and thus allows to semi-quantify the interaction strength. Correlation between interaction affinities *in vitro* and yeast growth on readout medium was demonstrated by Estojak and colleagues for *lacZ* and *LEU2* reporter genes (Estojak et al., 1995). I determined the highest concentration 3-AT for each interaction pair on which yeast growth was observable and used this value to define an interaction-specific 3-AT score. For self-activating baits I calculated the score as follows: $3AT\ score = 3AT_{max} - 3AT_{background}$

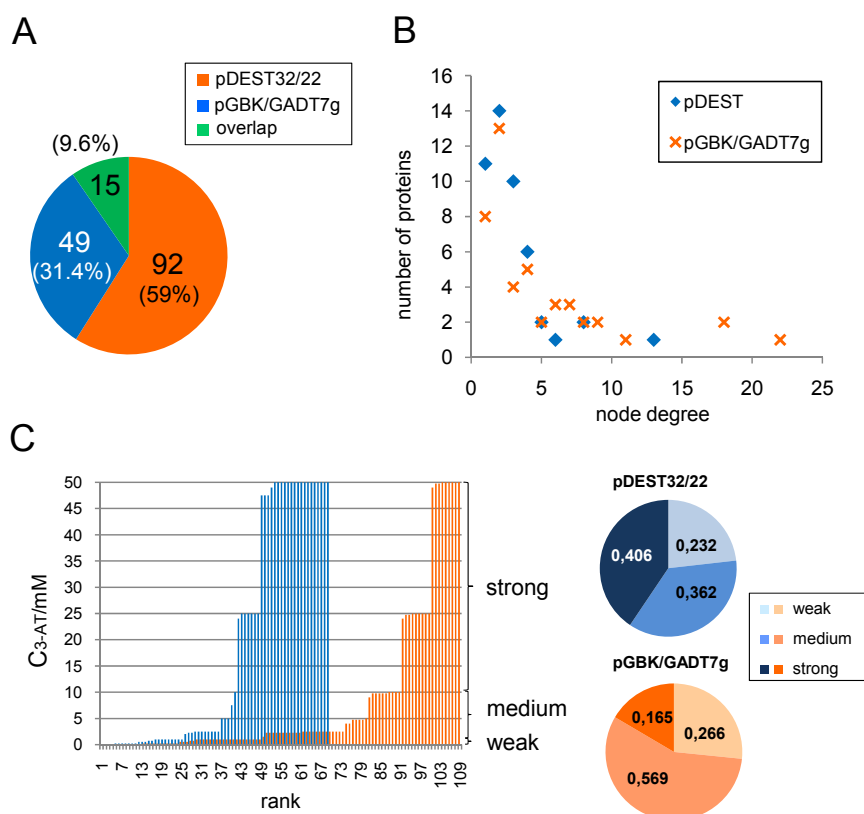


Fig. 22 Dp-1 PPI and network properties
(A) Absolute and relative number of detected PPIs and the applied Y2H expression vector systems.
(B) Number of detected inter-actions (node degree) plotted against the number of proteins with a certain number of interaction partners.
(C) Rank order plot of 3-AT scores of individual PPIs. The pie charts summarize the relative number of PPI 3-AT scores split into groups of suggested weak, medium, and strong interaction signals.

RESULTS

3-AT_{max} represents the highest concentration of 3-AT where colonies appeared on the readout medium test plate. 3AT_{background} is the minimal inhibitory concentration of 3-AT that is needed to suppress self-activation growth of a certain bait. This test was carried out in parallel. For simplification the PPIs were grouped as being of weak (<1 mM), medium (≥1 mM and <10 mM) or strong binding (≥10 mM) (see. Fig. 22C). Interestingly, the pDEST32/22 vector system resulted in stronger interactions (40.6%) compared to pGBKT7g/GADT7g (16.5%) that shows in general more medium and weak interactions.

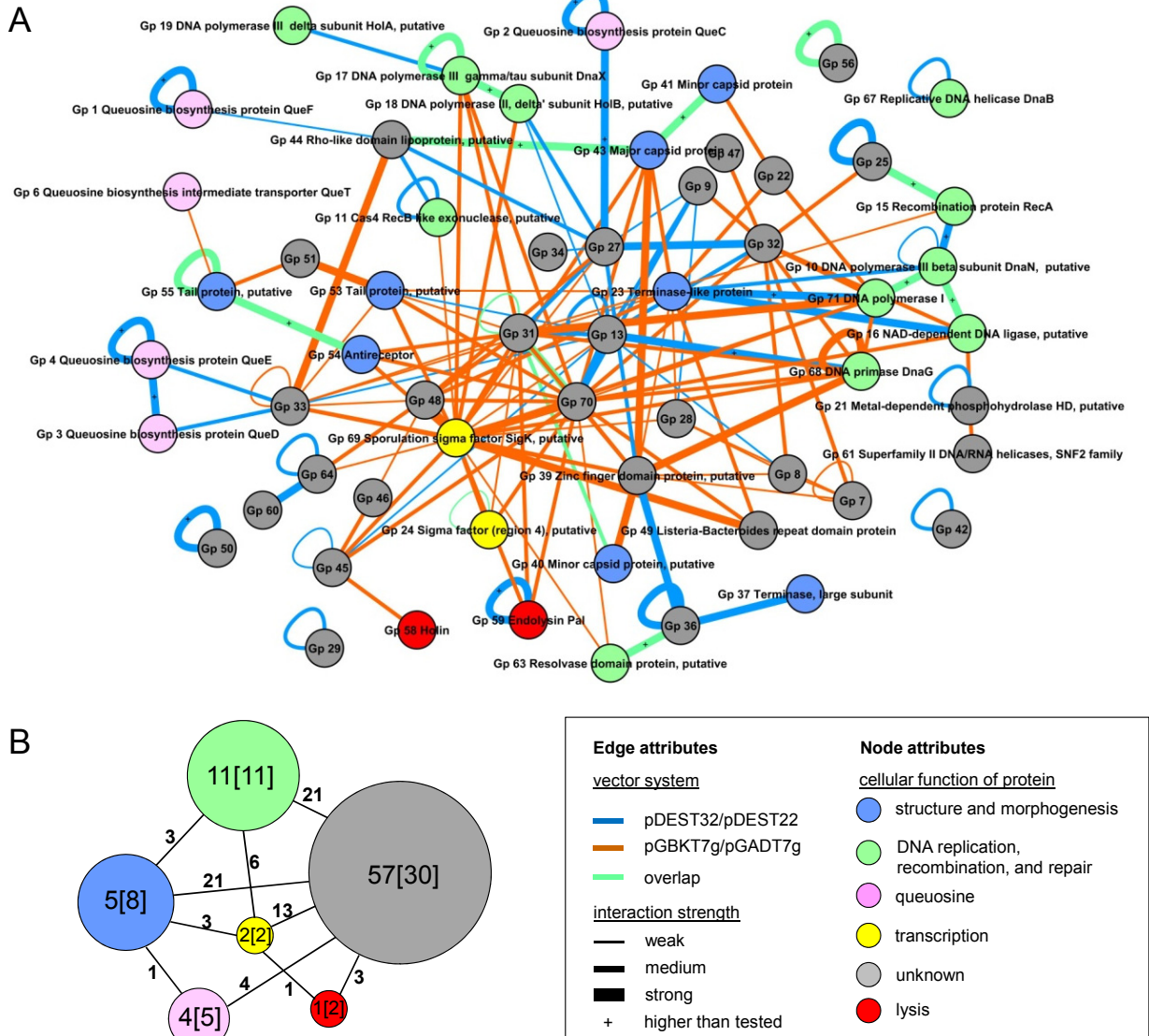


Fig. 23 Intra-viral Dp-1 PPI network connects most of the Dp-1 proteins

(A) Full PPI network. The nodes represent proteins and the edges represent the protein interactions. Properties are given in the figure legend. The interaction signal strength is indicated by the edge width and was determined by 3-AT titration (explained in the text). “+” labels interactions that were detectable on the highest tested concentration of 3-AT (50 mM). Figure was drawn with Cytoscape (Shannon et al., 2003). (B) Network is simplified by interactions among and between different gene product groups with an assumed general function (refer to Fig. 17C). Numbers of interacting proteins of a certain group and their interaction are summarized. The number in the group node represents the PPI number that occurs within this group, number in brackets indicate the number of proteins in a group. Numbers on the edges correspond to the interactions that occur between certain functional protein groups.

As it can be seen from Fig. 22B the network includes 30 proteins of unknown cellular function and links them on the interaction level to various protein groups. The two interacting sigma factors seem to play a central role in the network taken by their high interaction frequency with other protein groups although they may function primarily in transcription. Especially SigK (gp69) acts as a hub protein and shows the highest PPI number with 22 interactions. The average PPI number within a same group is never higher than 1 except for proteins of unknown function. For instance, DNA replication proteins interact on average with one interaction partner of their own functional class but undergo on average 1.9 PPIs with unknowns. This means that unknowns have the highest impact on the network in terms of their interaction number. Since most of them undergo binding with proteins of different functions in parallel an interpretation is difficult. I analyzed the network's functional properties considering the genomic context by a sophisticated statistical model (Fig. 25).

The binary Cp-1 interactome

The primary Y2H screens detected 34 redundant PPIs, 20 by the pDEST32/22 and 14 by the pGBKT7g/pGADT7g system. In Y2H retest experiments 10 PPIs (50%) were reproduced in pDEST32/22 and 12 (70.6%) in the pGBKT7g/pGADT7g expression vector system. Ignoring the AD/DBD fusion direction and the vector systems a total of 17 unique PPIs were identified (Fig. 24, see also Fig. 26A). 15 Cp-1 proteins occur in the network and thus a minimum of 53.6% of the Cp-1 proteins undergo protein-protein contacts. Since the absolute interaction number is low, I did not focus on network and PPI properties in detail as for the Dp-1 intra-viral PPI network. However, the same tendency regarding the different vector systems and 3-AT scores are obvious (Fig. 24). A full list of identified intra-viral Cp-1 PPIs is given in Tab. 28 (full documentation is given in the e-supplement).

The small number of intra-viral PPIs does not mean that more interactions failed to be detected since the Cp-1 genome encodes for only 28 proteins in contrast to Dp-1 with 72. The Dp-1 proteome has 2,628 unique interaction possibilities while Cp-1 has maximally 378. Based on the detected PPI number, Dp-1 uses 5.9% of its PPI spectrum while Cp-1 uses 4.5%, respectively.

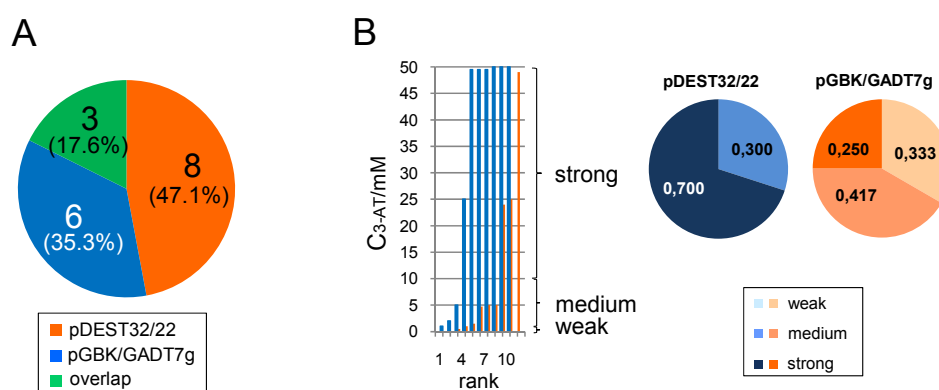


Fig. 24 Cp-1 PPI properties

(A) Absolute and relative number of detected PPIs and the applied Y2H expression vector systems. The numbers indicate the absolute and relative PPI numbers. (B) Rank order plot of 3-AT scores of individual PPIs. The pie charts summarize the relative number of PPI 3-AT scores split into groups of suggested weak, medium, and strong interaction signals.

RESULTS

Tab. 28 Intra-viral PPIs of Cp-1

The table represents a non-redundant list of Cp-1 intra-viral PPIs that were reproducible in Y2H retest experiments. DBD-/AD- (bait/prey) interaction direction and the vector system the PPIs were detected with are combined. (VS) indicates the positively tested vector system ((D) pDEST32/22 or (G) pGBKT7g/pGADT7g). The 3-AT score represents the concentration range of 3-AT in the readout medium where yeast growth was visible compared to self-activation background growth of that bait. “+” indicates that yeast grew still on the highest tested concentration (50 mM 3-AT).

Protein A	Protein B	VS	3-AT score/mM
Gp5 DNA polymerase	Gp6 Hypothetical protein	D	25
Gp8 Possible scaffolding protein	Gp10 Connector protein	D	49.5 ⁺
Gp8 Possible scaffolding protein	Gp13 Endoprotease of major head protein	D	2
Gp8 Possible scaffolding protein	Gp6 Hypothetical protein	G	1.5
Gp8 Possible scaffolding protein	Gp8 Possible scaffolding protein	D G	49.5 25 ⁺
Gp8 Possible scaffolding protein	Gp9 Major head protein	D D G G	49.5 ⁺ 49 ⁺ 50 ⁺ 25 ⁺
Gp9 Major head protein	Gp14 Hypothetical protein	G	5
Gpc Hypothetical protein	Gpc Hypothetical protein	G	5
Gp10 Connector protein	Gp10 Connector protein	G	1
Gp11 Lower collar protein	Gp11 Lower collar protein	G	4.75
Gp12 Hypothetical protein	Gp15 Hypothetical protein	D	50 ⁺
Gp14 Hypothetical protein	Gp14 Hypothetical protein	D	1
Gp16 Hypothetical protein	Gp15 Hypothetical protein	D G	0.1 50 ⁺
Gp17 Tail protein N	Gp11 Lower collar protein	G	0
Gp19 Tail protein C	Gp11 Lower collar protein	G	0
Gp19 Tail protein C	Gp17 Tail protein N	G	0.5
Gp20 Terminase	Gp20 Terminase	D	5

Comparison of the genomic contexts and the intra-viral PPI networks reveals a modular organization of phage interactomes

Expression of phage genes is stringently regulated after the phage has entered the host cell. Phage genes are organized in gene clusters (operons) encoding proteins involved in similar functional processes and having a close similar timing of transcription (e.g., (Martin et al., 1996b)).

Thus, I investigated for Dp-1 whether the detected protein-protein interactions correlate with their putative gene cluster function (Fig. 25). Therefore, I assumed that a module (operon) ranges from a predicted promoter start site to the next promoter. This was done since the predicted termination sites were not adequate to define a clear operon in some cases.

The number of PPIs that occur within the same functional cluster was counted, e.g., interactions among proteins involved in DNA replication are defined as “in function” interactions whereas PPIs between different functional gene batteries are “out of function”. As expected, the majority of interactions (64.6%) appear to be “in function” (only interactions from gene clusters with a predicted known function were considered for this calculation). Nevertheless, there are still 35.4% PPIs that are “out” due to binding between proteins from different functional clusters, indicating that there is active cross-talk between different functional modules. For instance, proteins encoded by expression units for

structural components show 27 intra-functional interactions. Proteins encoded by gene clusters involved in DNA replication undergo 29 intra-functional interactions. These two groups cross-communicate via 27 additional interactions and might reflect a close connection between replication and virion assembly. Most of the intra-functional interactions within the same group are significantly overrepresented (Fig. 25B).

The same tendency becomes obvious if the putative expression timing is considered (Fig. 25C). I grouped the predicted transcriptional units into four co-regulated functional classes that are putatively expressed in an early, middle, late, and an unknown stage. Here, 59.4% of the interactions are “in time” whereas 40.6% interact “out of time” (only interactions from gene clusters with a predicted expression timing were considered for this calculation). “In-time” interactions among all different groups (except unknowns) are statistically overrepresented significantly (Fig. 25C). PPIs that are “in” could underline their *in vivo* relevance since simultaneous protein presence is the requirement for physical binding. For “out” interactions the *in vivo* relevance is unclear, e. g. DNA replication proteins that bind with structural proteins. However, gene expression is a dynamic process and remnants of proteins that are present already in the early phase might have a regulatory relevance (or vice versa) on gene products that are predominantly present in the middle or late phase.

Some examples for “in time” and “in function” interactions are illustrated in Fig. 25D. Probably co-expressed queuosine biosynthesis proteins QueD and QueE undergo protein contacts and thus might reflect the reconstitution of a heteromeric enzyme complex necessary for synthesis of preQ₀. Also DNA polymerase III subunits interact among each other. Here gp17, gp18, and gp19 genes are genomic neighbors and associate also on the protein level. Gp71 (DNA polymerase I) which is encoded by a terminally located gene battery binds gp68 (primase) as well as gp10 DNA polymerase III patch clamp (DNA pol. III beta subunit). The latter is encoded by a different module. Here, it becomes clear that proteins involved in similar cellular processes can cross-communicate also beyond the genetic organization. In case of gp71 and gp10 the DNA polymerase I might function in recognition of stalled DNA polymerase III by gp10 binding to remove RNA primers from the heteroduplex DNA. I will focus on the modular genome organization and PPI properties in more detail in the discussion section and will give some comments on “out of time” and “out of function” PPIs.

RESULTS

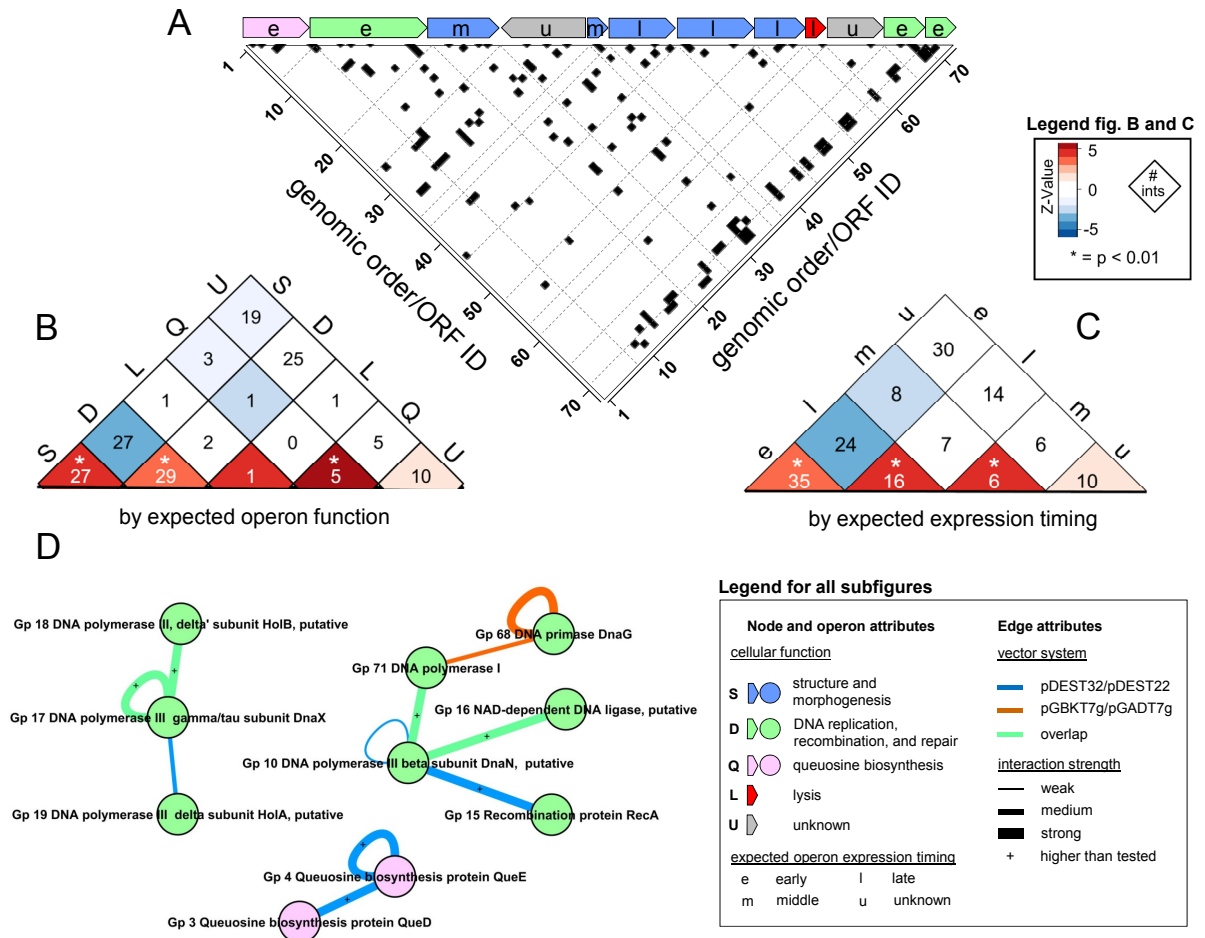


Fig. 25 Relationship of the Dp-1 genome context and the intra-viral PPI network

(A) Dp-1 whole genome matrix plotted against the identified PPIs. Detected PPIs are indicated as black boxes. Putative gene battery borders are highlighted by dashed lines and the putative transcriptional organization is indicated at the top. The color codes are explained in detail in the legend. (B,C) Simplified, collapsed genome matrices. All predicted modules of a certain general function (B) or of a certain expression timing (C) have been combined (e.g., all operons of the function “structure and morphogenesis” or of expected expression timing “early” are summarized as one class, respectively). Letters surrounding the matrices indicate a certain group. For abbreviations see legend. The number of interactions between various gene classes was counted for the real intra-viral PPI network of Dp-1 and for randomized versions of this network. Overrepresentation of a link compared to 10,000 randomized networks was assessed by calculating a Z-score. The Z-score color gradient (see upper legend) indicates over- or underrepresentation of the PPI number within a certain gene class or between different gene classes. The numbers in the various matrix quadrants represent the absolute PPI number of the real network. Statistical significance for over- or underrepresented PPI numbers among certain gene classes is indicated in the corresponding quadrants by an asterisk ($p < 0.01$). Statistics were done in collaboration with Björn Titz, Crump Institute for Molecular Imaging, Los Angeles. (D) Some PPIs that occur within a certain gene class, here “in time” and “in function”. The examples indicate interactions among DNA replication proteins and Q enzymes. Nodes represent the proteins and edges the interactions between protein pairs. The interaction strength is indicated by the edge width which was determined by 3-AT titration (see main text). “+” labels the interactions that were still detectable on the highest tested concentration of 3-AT. Figure was drawn with Cytoscape (Shannon et al., 2003). Interaction and protein properties are indicated in the legend.

The small Cp-1 PPI network (Fig. 26A) does not allow for a statistical comparison of the genome organization with interactome properties as done for Dp-1. However, the tendency becomes obvious that modularity of the genome can be recovered on the PPI level. Nearly all PPIs occur “in function”. Compared to Dp-1 that exhibits also an agile cross-communication among different gene classes, Cp-1

RESULTS

seems to be organized even more modular on the PPI level. This suggests that more complex biological systems exhibit more cross-talk on the protein level.

In the genome matrix quadrant II (Fig. 26B), representing the module where genes of structural proteins are accommodated, 13 interactions can be found, e.g., among the tail proteins N and C and the lower collar protein. Another partial reconstitution of the virion is represented by the interaction of the scaffolding protein with the major head protein (MHP) and connector protein. Represented by quadrant I (Fig. 26B) a PPI was detected among the Cp-1 DNA polymerase and the hypothetical protein gp6 which is encoded by the same gene module. Although the function of gp6 is unknown the interaction links gp6 to the DNA polymerase and thus might function in Cp-1 DNA synthesis. Only one PPI occurs “out of function”: gp6 binds to the scaffolding protein gp8 and links replication with virion morphogenesis. However, the physiological meaning here is unclear. Some expected PPIs were not detected in the screens for Cp-1 (e.g., MHP protease with MHP or tail proteins with the antireceptor) as well as others for Dp-1. Putative false-negative PPIs will be discussed later (4.1.1).

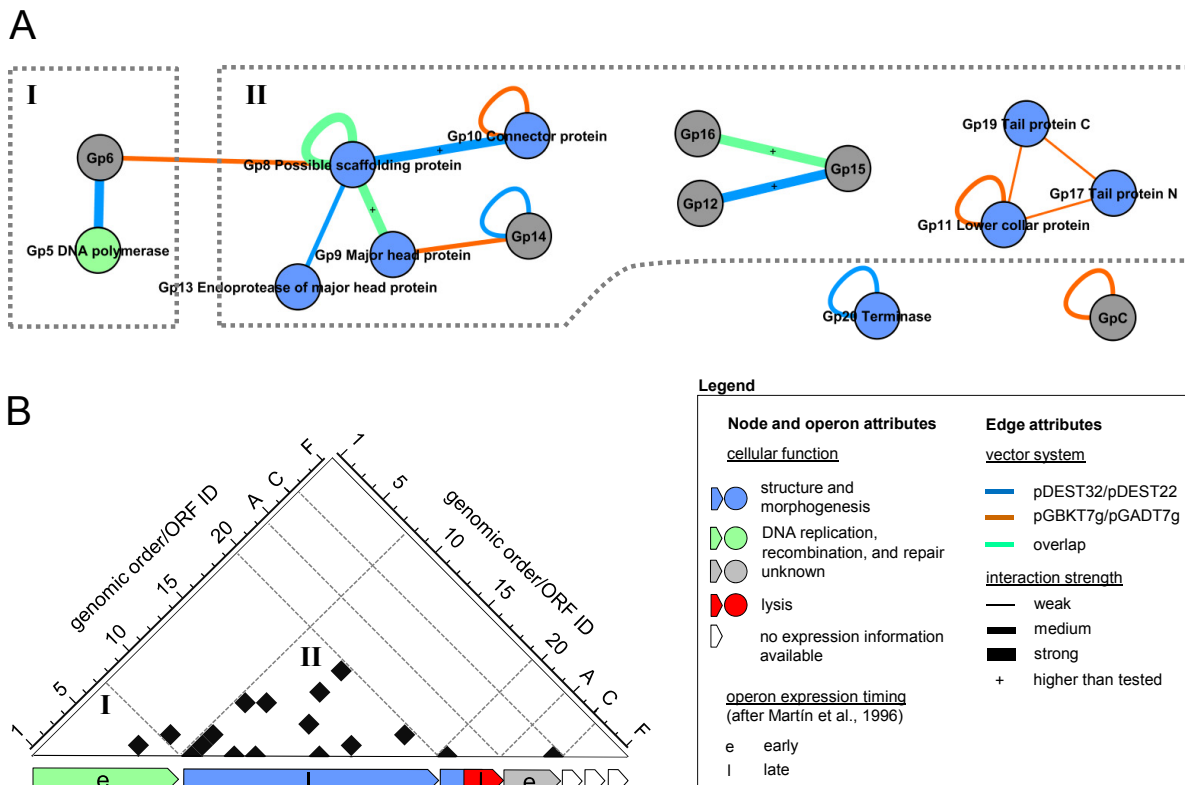


Fig. 26 Intra-viral PPI network of Cp-1 and relationship to its genomic context

(A) All identified intra-viral PPIs presented as interaction network. Nodes represent the proteins and edges represent the interactions between protein pairs. Roman numerals and boxes indicate PPIs that occur among the corresponding, simplified transcriptional units given in (B). The properties in (A) are explained in the legend. The interaction signal is indicated by the edge width which was determined by 3-AT titration (explained in the main text). “+” labels the interactions that were detectable at the highest tested concentration of 3-AT. Figure was drawn with Cytoscape (Shannon et al., 2003). (B) Cp-1 genome matrix vs. identified PPIs. Detected PPIs are highlighted by black boxes. Operon borders are highlighted by dashed lines and were drawn after Fig. 6. Roman numerals in the quadrants correspond to those in the boxes given in (A). Corresponding expression units are illustrated at the bottom. Properties of (B) are explained in the legend.

3.1.4 Protein-protein interactions of Cp-1 and Dp-1 with their host *S. pneumoniae*

In the previous section I identified systematically intra-viral PPIs of Cp-1 and Dp-1. However, the phages' reproduction is dependent on their host. Thus, phages cannot be regarded as isolated biological entities. After they have entered the host cell they use the host cell to produce large quantities of phage proteins by using the host's translation machinery.

The phage-host dependence can be tested on different cellular levels. For instance, Qimron and colleagues searched genome-wide for genetic links between *E. coli* and T7 affecting phage reproduction. Surprisingly, only a small number of *E. coli* ORFs turned out to affect T7 growth. Four genes had a deleterious effect when over-expressed. Another 11 host genes were essential for T7 growth once they were knocked out (Qimron et al., 2006). This indicates that in fact a small subset of host proteins has a significant influence on T7 growth. However, the methodology of this study focused on genetic links and might not have been sensitive enough to find all host links, e.g., those that fine-tune and optimize phage reproduction or that are not examinable at all since the host genes are essential for the host itself.

Alternatively, the exploration of the PPI level provides the advantage to unravel other virus-host relationships, e.g., such host factors that do not have a significant impact on phage reproduction under laboratory conditions.

A physical association among phage-host protein pairs can specifically imply which host pathways could be manipulated, e.g., a virus protein might neutralize a toxic host protein by binding or vice versa. Alternatively, the phage might be capable to recruit host proteins towards phage-related processes by binding to fill up its "genomics gaps" by host proteins.

The aim of this project was to screen proteome-wide for phage-host PPIs. This has never been done before. Such studies are critical to understand the phage-host relationship on a global level since PPIs can indicate which host pathways are relevant for a phage and thus necessarily might be modified, disrupted, or stimulated by the phage.

Proteome-wide identification of phage-host PPIs by Y2H screens

To identify proteome-wide phage-host PPIs I screened with the Y2H system all individual Cp-1 as well as Dp-1 baits as pDEST32-DBD fusions against an *S. pneumoniae* TIGR4 strain prey array, kindly provided by Seesandra V. Rajagopala (JCVI). The host prey array contained 1,705 individual pDEST22-AD preys which represent 76% of the *S. pneumoniae* genome. Due to the large size of the prey array and to speed up the screening procedure I decided to perform the screens by a mini-prey pooling strategy (2.2.5.7). I therefore collapsed the 18 individual 96-format prey array plates into a single plate resulting in prey pools containing 18 individual prey clones per test position. These were mated against individual phage baits. Positives were identified by a yeast colony PCR followed by sequencing of the PCR product (2.2.5.8). The primary pooling screens detected a total of 130 phage-host interaction pairs (a full list is given in the e-supplement). Since pooling screens are known to

RESULTS

produce a higher false-negative rate I screened in addition 25 Cp-1 and five Dp-1 baits against the *S. pneumoniae* full-prey matrix array as one-on-one tests resulting in the detection of 63 PPIs. Combined, this resulted in a total of 186 individually detected interaction candidates. I subsequently retested 170 of them in Y2H retest experiments. 49 interactions were clearly reproducible. 121 were not reproducible and mostly caused by high experimental background of a few bait proteins (randomly appearing colonies). The baits that have been screened by both strategies revealed 12 reproducible interactions. 7 of them were detected by both methods, 5 only with one-on-one screens, and none in addition with the pooling strategy. This implies that the one-on-one tests were more sensitive and that the intrinsic false-negative error rate of the pooling screens is about 42%. Although I adjusted nearly equal cell numbers of the preys by pipetting during the preparation of the pooling plate, some preys might have been over- or underrepresented (Fig. 27). This could explain the calculated higher false-negative rate. The screens were done only as phage-DBD → host protein-AD direction and not vice versa due to the immense working effort. Moreover, a second vector system was not considered, e.g., as for the intra-viral screens the pGBKT7g/pGADT7g system. Thus, the 49 detected phage-host interactions might represent only a small subset of all possible interactions and I expect that more PPIs among phage and host proteins take place *in vivo*.

However, these results represent the most comprehensive dataset among phage-host PPIs. In summary, the retests revealed 49 unique virus-host PPIs, of which 11 are attributed to Cp-1 and 38 to Dp-1. This involves 36 uniquely binding *S. pneumoniae* proteins, seven Cp-1 and 19 Dp-1 proteins. All detected interactions are listed in Tab. 29 and shown in Fig. 28.

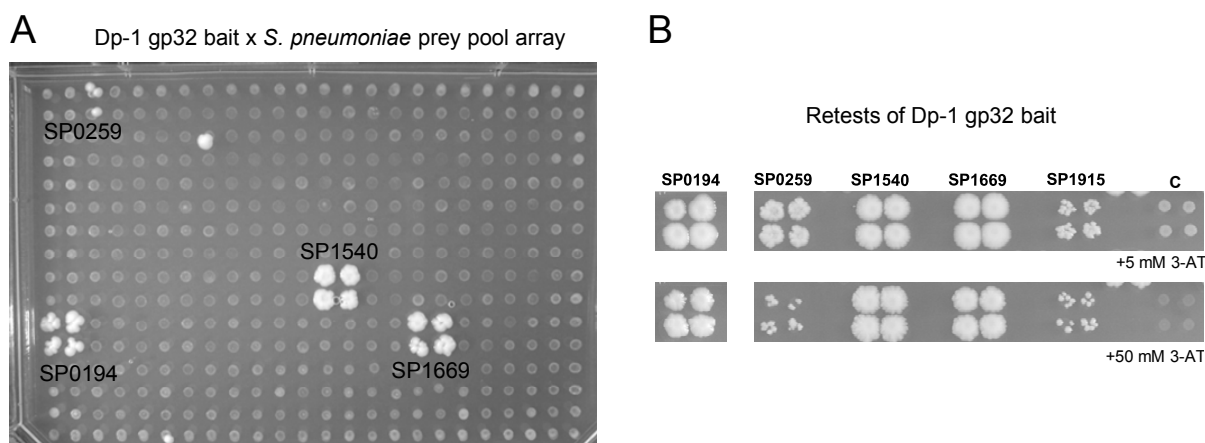


Fig. 27 A screen and retest example

(A) Y2H screen of Dp-1 gp32 bait against *S. pneumoniae* mini-pool array consisting of 18-prey pools per test position. The -Leu-Trp-His readout medium plate contained 5 mM 3-AT to suppress gp32-DBD self-activation and was incubated for 7d at 30°C. Tests were done as quadruplicates. (B) One-on-one Y2H retests of gp32 in quadruplicate format. Note, that the SP1915 interaction was also found in another screen. Gp32 self-activated up to 2.5 mM 3-AT in the readout medium. Representatively, plates containing the minimal inhibitory (5 mM) and the highest tested 3-AT concentrations (50 mM) are shown. Dp-1 gp32 is a protein of unknown function. Interacting host proteins are (SP0194) conserved hypothetical protein, (SP0259) Holliday junction DNA helicase RuvB, (SP1540) single-strand DNA binding protein, (SP1669) MutT/nudix family protein, and (SP1915) hypothetical protein. Although yeast growth on the SP0259 screen test position was weak, the retest results were clear. This was probably caused by underrepresentation of the SP0259 prey cell number in the pool.

RESULTS

Tab. 29 Phage-host PPIs detected by Y2H screens and verified by Y2H retests

The table lists all identified interactions between Cp-1/Dp-1 and *S. pneumoniae* protein pairs, verified by Y2H retest experiments. The last column represents the 3-AT score, the concentration range of 3-AT/mM in the readout medium where yeast growth was visible compared to self-activation background growth of a certain bait. “+” indicates that yeast grew still on the highest tested concentration (50 mM 3-AT).

Phage	Protein	Phage protein description	Host protein	Host protein description	3-AT score
Cp-1	Gp10	Connector protein	SP1354	ribosomal protein L7/L12	2.5
Cp-1	Gp10	Connector protein	SP1881	glutamate racemase Murl	50 ⁺
Cp-1	Gp16	Hypothetical protein	SP0259	Holliday junction DNA helicase RuvB	1
Cp-1	Gp17	Tail protein N	SP0979	oligoendopeptidase F	50 ⁺
Cp-1	Gp22	Lysozyme (Cp11)	SP1208	uridine/cytidine monophosphokinaseUdk	47.5 ⁺
Cp-1	Gp6	Hypothetical protein	SP1713	transcriptional regulator, NrdR family	50 ⁺
Cp-1	Gpb	Hypothetical protein	SP0859	membrane protein (DUF979)	25
Cp-1	Gpb	Hypothetical protein	SP1213	conserved domain protein (DUF2130)	50 ⁺
Cp-1	Gpb	Hypothetical protein	SP1980	cmp-binding-factor 1 Cbfl	25
Cp-1	Gpb	Hypothetical protein	SP2168	putative fucose operon repressor FcsR	2.5
Cp-1	Gpc	Hypothetical protein	SP0979	oligoendopeptidase F	2
Dp-1	Gp12	Holliday junction resolvase RecU, putative	SP2168	putative fucose operon repressor FcsR	0.1
Dp-1	Gp14	dUTPase	SP2125	conserved hypothetical protein (DUF59)	50 ⁺
Dp-1	Gp16	NAD-dependent DNA ligase, putative	SP0259	Holliday junction DNA helicase RuvB	50 ⁺
Dp-1	Gp18	DNA polymerase III, delta' subunit HolB, putative	SP1584	GTP-sensing transcriptional pleiotropic repressor CodY	24.5
Dp-1	Gp29	Hypothetical protein	SP2012	glyceraldehyde 3-phosphate dehydrogenase (Gap)	0.25
Dp-1	Gp31	Hypothetical protein	SP0259	Holliday junction DNA helicase RuvB	0.5
Dp-1	Gp31	Hypothetical protein	SP1153	hypothetical protein	0.25
Dp-1	Gp31	Hypothetical protein	SP2168	putative fucose operon repressor FcsR	0.2
Dp-1	Gp32	Hypothetical protein	SP0194	conserved hypothetical protein (DUF1292)	47.5 ⁺
Dp-1	Gp32	Hypothetical protein	SP0259	Holliday junction DNA helicase RuvB	47.5 ⁺
Dp-1	Gp32	Hypothetical protein	SP1540	single-strand DNA binding protein	47.5 ⁺
Dp-1	Gp32	Hypothetical protein	SP1669	MutT/nudix family protein	47.5 ⁺
Dp-1	Gp32	Hypothetical protein	SP1915	hypothetical protein (LytTr DNA-binding domain)	47.5 ⁺
Dp-1	Gp33	Hypothetical protein	SP1088	DNA repair protein RadC	0.5
Dp-1	Gp34	Hypothetical protein	SP0446	acetolactate synthase, small subunit IlvN	9
Dp-1	Gp34	Hypothetical protein	SP2157	alcohol dehydrogenase, Fe-containing (ADH)	2
Dp-1	Gp39	Zinc finger domain protein, putative	SP0259	Holliday junction DNA helicase RuvB	50 ⁺
Dp-1	Gp4	Queuosine biosynthesis protein QueE	SP2036	PTS system, IIA component (PTS EIIA2)	25
Dp-1	Gp44	Rho-like domain lipoprotein, putative	SP0446	acetolactate synthase, small subunit IlvN	9.75
Dp-1	Gp44	Rho-like domain lipoprotein, putative	SP1050	putative transcriptional regulator (HTH domain)	4.75
Dp-1	Gp44	Rho-like domain lipoprotein, putative	SP1536	conserved hypothetical protein (methyl transferase domain)	4.75
Dp-1	Gp44	Rho-like domain lipoprotein, putative	SP1575	conserved hypothetical protein (DnaB/D domain)	9.75
Dp-1	Gp44	Rho-like domain lipoprotein, putative	SP1725	sucrose operon repressor ScrR	2.25
Dp-1	Gp44	Rho-like domain lipoprotein, putative	SP2157	alcohol dehydrogenase, Fe-containing (ADH)	2.25
Dp-1	Gp47	Hypothetical protein	SP0687	ABC transporter, ATP-binding protein	49 ⁺
Dp-1	Gp48	Hypothetical protein	SP1746	conserved hypothetical protein (HD phosphohydrolase domain)	2.5
Dp-1	Gp48	Hypothetical protein	SP2168	putative fucose operon repressor FcsR	0.5
Dp-1	Gp51	Hypothetical protein	SP1672	recombination protein RecR	0.5
Dp-1	Gp58	Holin	SP1505	membrane protein (DUF20)	1
Dp-1	Gp58	Holin	SP1606	glycosyl transferase, family 2	25
Dp-1	Gp58	Holin	SP1731	conserved hypothetical protein (DUF1212)	2.5
Dp-1	Gp58	Holin	SP1751	putative transporter, CorA family	25
Dp-1	Gp60	Hypothetical protein	SP2024	PTS system, IIA component (PTS IIA)	10
Dp-1	Gp72	Membrane protein, putative	SP1606	glycosyl transferase, family 2	50 ⁺
Dp-1	Gp9	No similarity	SP0259	Holliday junction DNA helicase RuvB	50 ⁺
Dp-1	Gp9	No similarity	SP1395	putative phosphate transport system regulatory protein PhoU	50 ⁺
Dp-1	Gp9	No similarity	SP1504	TPR domain protein	0.5
Dp-1	Gp9	No similarity	SP2168	putative fucose operon repressor FcsR	50 ⁺

***Dp-1* and *Cp-1* attack the host interaction network mainly at different positions**

The detected phage-host interactions reveal that *Dp-1* and *Cp-1* attack the host mainly by different PPIs (Fig. 28). Since both phages are remarkably different and belong to different phage families, these results were expected. The findings might indicate that both phages evolved different strategies to manipulate the host system by attacking or recruiting different protein binding partners.

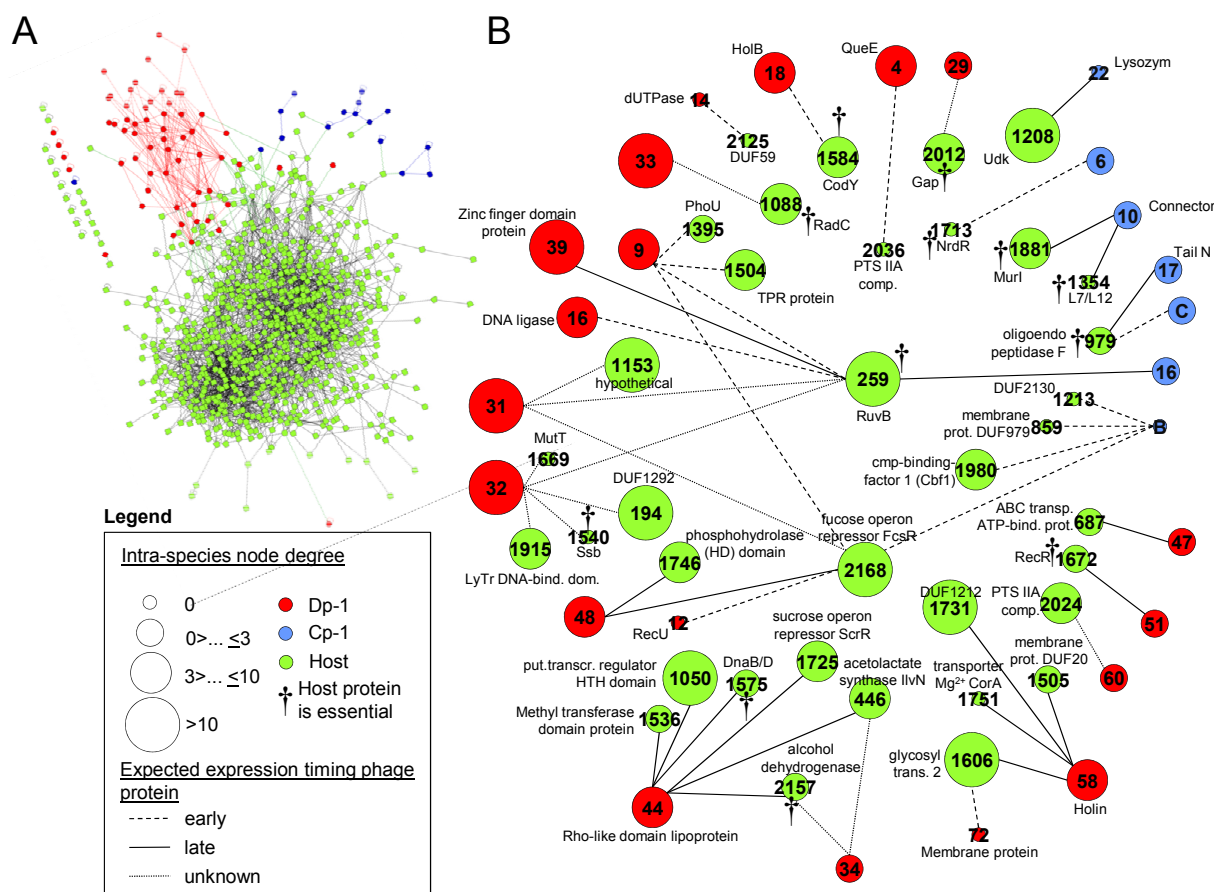


Fig. 28 *Dp-1* and *Cp-1* interactions with the host PPI network

(A) The phages' intra-viral networks are combined with a *S. pneumoniae* PPI network. The host network was constructed based on Y2H interaction data, kindly provided by Seesandra V. Rajagopala (JCVI, Rockville, MD) and Russ Finley (CMMG, Wayne State University School of Medicine, Detroit, MI). Nodes symbolize the proteins and edges the PPIs. Figure was drawn with Cytoscape (Shannon et al., 2003). (B) The network from (A) is simplified and the detected phage-host PPIs are highlighted. Numbers in the nodes correspond to the systematic ORF numbers of the corresponding species. Properties are given in the legend. Gene product descriptions (if available) are given in the figure. The intra-species node degree of certain proteins is indicated by the node size (number of interactions of a protein that can be counted in the intra-viral or intra-host PPI network). Expected expression timing of phage proteins post infection is indicated by different edge properties as given in the legend. Expression timing of *Dp-1* proteins is based on predictions (3.1.1) and of *Cp-1* used from (Martin et al., 1996b). Essential host proteins that bind phage proteins are labeled by †. Information about essential host proteins was integrated from several studies done with *S. pneumoniae* (Lee et al., 1999; Song et al., 2005; Thanassi et al., 2002; van Opijnen et al., 2009; Zalacain et al., 2003).

Interestingly, *Cp-1* and *Dp-1* gene products show overlapping interactions with two host proteins. On the one hand, this is the Holliday junction DNA helicase RuvB (SP0259) which is bound by five different *Dp-1* proteins and gp16 (hypothetical protein) of *Cp-1*. This suggests that the host's homologous recombination system might be critical for both phages. On the other hand, a putative

fucose operon repressor (SP2168) is bound by four different Dp-1 proteins and the Cp-1 protein gpC. This could mean that reprogramming of the fucose operon gene expression by SP2168 blockage might be an important event for both phages. While the latter two mentioned host proteins are the only ones attacked by both phages, there are other host proteins that have more than one phage interaction partner of the same phage species. These cases are included in Fig. 28B. For instance, the host oligoendopeptidase F (SP0979) is bound by Cp-1 Tail N protein (gp17) and gpC of unknown function. GpC was shown to be expressed early post infection whereas gp17 is expressed late (Martin et al., 1996b). This indicates that there could be a need for the phage to block this enzyme, e.g., to prevent degradation of phage proteins all over the time the phage is present the host. This could be facilitated through targeting the same host protein by various phage proteins that are maximally expressed at different time points.

Comparison of host and phage PPI network properties

An unpublished *S. pneumoniae* Y2H interactome is available (references are given in the text of Fig. 28). I compared various network properties to check whether the intra-viral networks differ in certain aspects from the host network.

Interaction specificities: In Fig. 29A I plotted the intra-host network protein node degree against the number of host proteins that mediate interactions with the tested phage proteins. Surprisingly, nine host proteins undergo interactions with phage proteins but are not present in the host network. This indicates that phage proteins are capable to bind host targets that do not undergo necessarily physical contacts with other host proteins. Thus, phages can very specifically attack host proteins, e.g., Cp-1 gp6 of unknown function binds to the NrdR host repressor that usually blocks expression of ribonucleotide reductase genes and primarily binds to its DNA repressor binding site (discussed later in more detail) (Torrents et al., 2007). On the other hand, phage proteins also bind to host hub proteins. Hubs are thought to play an important physiological role while having a high binding frequency (Dosztanyi et al., 2006). For instance, the conserved hypothetical protein SP0194, conserved in 99 bacterial species (COG3606) (von Mering et al., 2007), interacts with 38 various host proteins and with one phage protein (hypothetical protein gp32 of Dp-1). In summary, the majority of 31 phage PPIs are mediated by host proteins with an intra-host node degree <10 indicating specific phage-host PPIs in contrast to 18 phage-PPIs that involve host hub proteins (node degree ≥ 10). This underlines that the phages can specifically attack proteins that do not undergo PPIs in the host as well as hub proteins. In the latter case the phage-host interaction is very specific since only a single phage protein binds to a single host hub protein.

Cp-1 and Dp-1 attack frequently essential host proteins: Five previously published studies determined the essentiality of 480 *S. pneumoniae* genes which is approx. 20% of the genome (Lee et al., 1999; Song et al., 2005; Thanassi et al., 2002; van Opijnen et al., 2009; Zalacain et al., 2003). I used this

information to check how often the phage proteins bind to essential host proteins. Therefore, the ratio of interacting essential and interacting non-essential host proteins (or the number of interactions, respectively) of different PPI datasets was calculated (Fig. 29B). Surprisingly, proteins of both phages bind more often to essentials than the host proteins themselves. Dp-1 shows 24.1% and Cp-1 even 200.3% more targeted essential host proteins than the host itself. Cp-1 exhibits as many interactions with essential host proteins as with non-essentials (see also Fig. 28B). This finding becomes more dramatic if the relative interaction number is considered: Dp-1 shows 66.5% and Cp-1 284.3% more interactions with essential host proteins compared to the host-host PPIs. This is because the same essential host proteins can be multiply bound by various phage proteins. The calculations indicate that both phages bind frequently to essential host proteins and thus might target predominantly essential host processes.

Phage proteins undergo more homomeric interactions than host proteins: Homomeric interactions play a crucial role in many protein functions, e.g., constitution of homomeric, quaternary enzyme complexes, or structural protein complexes. Dp-1 homomers occur 2.1 times and Cp-1 homomers 6.1 times more frequently than homomeric interactions among *S. pneumoniae* proteins (Fig. 29C). Thus, the smaller a biological system is the more important self-interactions could be. Especially in the case of the bacteriophages, protein homomerization seems to play an important biological role in virion assembly. For instance, I detected various homomerization PPIs among structural components of Cp-1 and Dp-1 (see above). Also three out of six proteins of the Que gene cluster of Dp-1 undergo detectable self-interactions as an example for assembly of homomeric enzyme complexes.

Phage-interactions with functional host protein classes: To check whether certain host protein groups are preferentially bound or not bound by phage proteins I used the TIGR CMR cellular main role classification of *S. pneumoniae* proteins (Davidsen et al., 2010) and counted the number of interacting host proteins. As shown in Fig. 29D there are certain host protein groups that are targeted by both phages, e.g., proteins of DNA metabolism, proteins that have regulatory functions or that are involved in cell envelope biogenesis. This could mean that manipulation of these certain processes is important for both phages although host binding partners can differ. Other groups are targeted only by one phage, e.g., four interacting host proteins involved in transport and binding were detected for Dp-1 but none for Cp-1 proteins. As a consequence, taking advantage of host take-up systems by Dp-1 might represent a special property of this phage whereas Cp-1 might have no effect on this process.

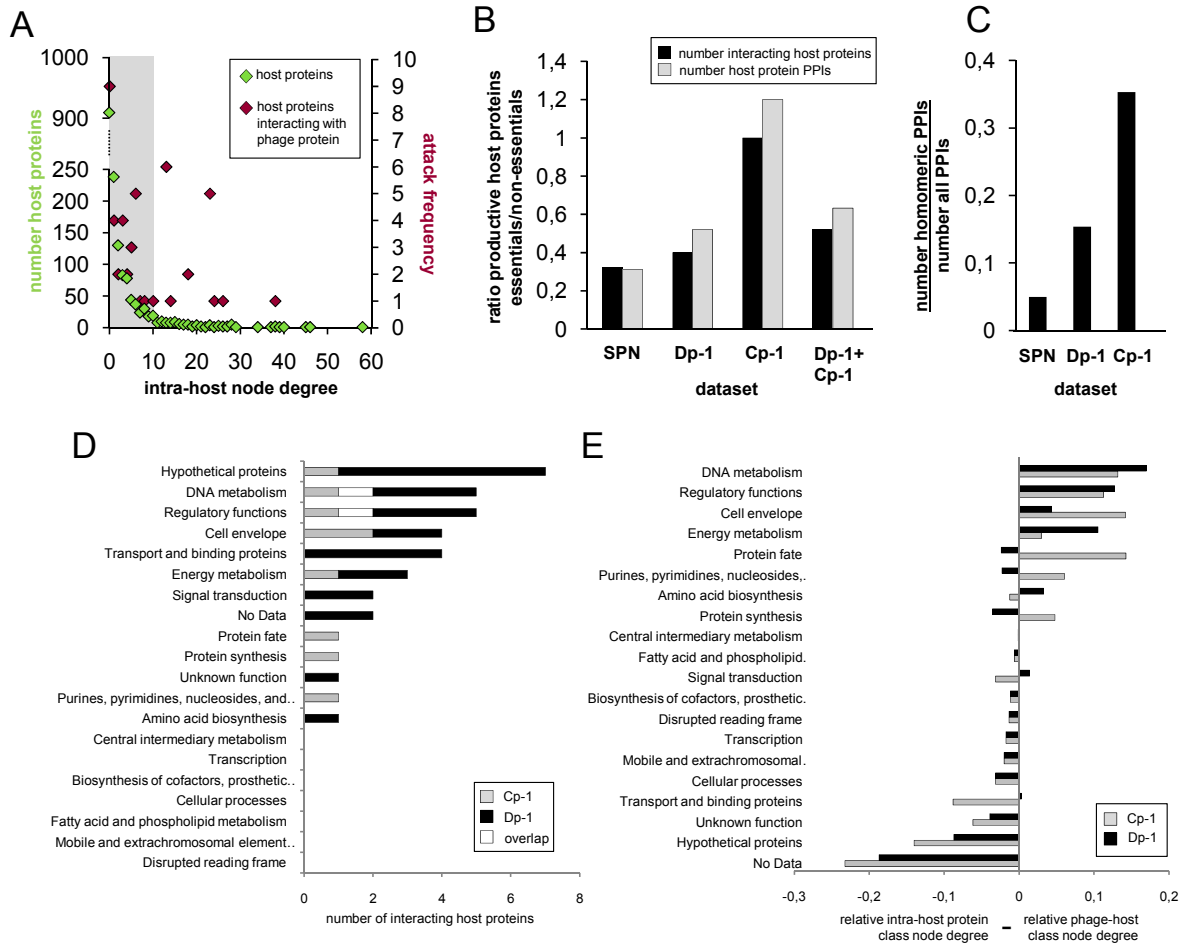


Fig. 29 Comparison of host-phage PPI network properties

(A) The number of PPIs (node degree) among *S. pneumoniae* proteins compared to the number of host proteins that interact with phage proteins. The intra-host protein node degree is sorted by ascending values (X-axis). The Y-axis to the left represents the number of host proteins (in green) with a certain node degree (a node degree of 0 was considered). The X-axis to the right represents the number of detected phage-host PPIs of a host protein (in purple) with a certain intra-host node degree (attack frequency) and indicates the interaction specificity of phage-protein interacting host proteins. The section labeled in grey highlights the area including specifically interacting host proteins (defined by a node degree <10). Proteins with node degrees ≥ 10 were considered as hub proteins.

(B) Bacteriophages bind essential host proteins more frequently than the host itself. PPI datasets are compared regarding the essentiality of productive host proteins (productive means that a host protein exhibits minimum one PPI in a dataset). PPI datasets are the intra-host network (SPN), Dp-1-host PPIs (Dp-1), Cp-1-host PPIs (Cp-1), and the combined phage-host PPIs of Cp-1 and Dp-1 (Dp-1 + Cp-1). Bars in black represent the number of interacting essential host proteins divided by the number of interacting non-essential host proteins. Bars in grey indicate the number of PPIs that are mediated by interacting essential host proteins divided by the number of interactions that are mediated by non-essentials.

(C) The phage PPI modules exhibit a higher number of homomeric interactions than the host. The different datasets represent intra-species networks. The number of homomeric interactions per dataset was counted and divided by the number of all PPIs of the network including homomeric and heteromeric PPIs.

(D) Absolute number of detected phage-host PPIs differentiated by the main role of host proteins (TIGR cellular main role, (Davidsen et al., 2010)).

(E) Host protein classes that interact more frequently with other host proteins than with phage proteins or vice versa: the sum of interactions among a certain host protein group (class node degree) was counted and divided by the number of all interactions of the intra-host network (the sum of all class node degrees). These were subtracted from the following ratio: the number of PPIs that a certain host protein class mediates with phage proteins was counted and divided by the sum of all phage-host PPIs counted for Cp-1 or Dp-1. A protein class with a negative value indicates that this class binds relatively more frequently with any other host proteins than with Cp-1 or Dp-1 proteins; groups with positive values indicate that they bind relatively more frequently with Cp-1 or Dp-1 proteins than with other host proteins.

I wondered whether certain host protein classes interact relatively more frequently with other host proteins or with phage proteins (and thus, if phage-host protein interactions preferentially are overrepresented in certain classes or vice versa) (Fig. 29E). Some host protein classes seem to be ignored by the phages: e.g., while “hypotheticals” interact frequently in the host PPI network they relatively do not exhibit many PPIs with phage proteins. Note, that in Fig. 29D this protein group shows the highest absolute PPI number with phage proteins. But given the fact that in the host network these proteins are involved in mediating 987 links, the absolute number of seven PPIs with phage proteins in this class is negligible. On the other hand, host proteins involved in DNA metabolism and regulatory functions do not mediate as many interactions in the *S. pneumoniae* network but undergo a relatively higher interaction number with phage proteins. This indicates that the phages interact with these certain classes relatively more frequently than the host itself.

Comparison of T7 phage genetic host links with Dp-1 and Cp-1 virus-host PPIs

The detected Cp-1/Dp-1-host PPIs were compared to the available genetic T7 phage-host links identified by Qimron and colleagues (Qimron et al., 2006). This was done to check whether there are overlaps among the detected PPIs and the genetic information available from T7 and thus if proteins from different hosts might be important for phage reproduction in general.

As already mentioned, T7 growth is dependent on the presence of 11 non-essential *E. coli* genes, including proteins involved in LPS biosynthesis (not relevant for phages infecting gram-positives) and a thioredoxin. T7 growth is inhibited by over-expression of four host genes (dGTP triphosphatase, the hsdR subunit of the EcoKI restriction enzyme, the rcsA activator of capsule synthesis genes, and uridine-cytidine kinase). From these 15 host genes that have an effect on T7 reproduction 7 have 22 orthologous proteins in *S. pneumoniae* (cluster of orthologous groups (COGs) (von Mering et al., 2007)). The latter number results from paralogs present in *S. pneumoniae*. Surprisingly, one out of these 22 proteins undergoes direct binding with a phage protein identified in this study: the Cp-1 lysozyme Cpl1 binds the *S. pneumoniae* uridine-cytidine kinase (Udk) which was shown by Qimron et al. to have a deleterious effect on T7 growth when it is over-expressed. Since Cp-1 is a T7-like virus, this overlap might indicate that Udk could be of general relevance for podovirus reproduction.

Verification of phage-host PPIs by LuMPIS pull down assays

The Y2H system is known to be one of the most sensitive methods to screen for PPIs (Braun et al., 2009; Chen et al., 2010; Venkatesan et al., 2009). Nevertheless, there are often doubts left about the quality of Y2H interaction datasets since it is thought that many interactions pass the screen (false-negatives) and many interactions are spurious (false-positives). The latter can be caused by randomly appearing yeast colonies, a heterologous experimental situation (e.g., expression of alien proteins in yeast), or by misfolded proteins that lead to sticky interaction. Random effects can be excluded by systematic Y2H retesting (as done here). Sticky interactions can be efficiently filtered from large PPI datasets since an unusual high protein node degree indicates a sticky behaviour. However, since the

RESULTS

Cp-1 and Dp-1 genomes have tiny gene numbers compared to other organisms such a simple filtering approach was not adequate. Thus, the data validation was done experimentally.

First, I started to express a set of recombinant proteins from Cp-1, Dp-1, and *S. pneumoniae* in *E. coli* with the aim to use the lysates in classical pull down assays and to purify proteins for *in vitro* experiments. However, the majority of expression tests resulted in exiguous protein amounts indicating that the phage and *S. pneumoniae* proteins act either toxic in *E. coli* or cannot be expressed well because of the ORFs' low GC-content. Regarding the latter two reasons I decided to check the interactions in a heterologous system: LuMPIS, a luminescence-based MBP pull-down interaction screening system (Vizoso Pinto et al., 2009), is a variation of the LUMIER system (luminescence-based mammalian interactome mapping system) (Barrios-Rodiles et al., 2005) and can be applied in a 96-format. It makes use of an MBP-tagged bait to pull down GFP-luciferase-tagged prey proteins that can be detected by luminescence readout. It was shown to function particularly well for the expression of ORFs with a low GC-content in mammalian cells (Vizoso Pinto et al., 2009). This technology allowed to test the interactions while avoiding problems related to protein expression in bacteria. Here, the aim was to validate the identified PPIs dataset by an alternative method that differs from the initially used Y2H method (genetic reporter in an “*in vivo*”-like situation in contrast to direct “*in vitro*” binding).

I checked systematically a subset of the interactions that were identified in this work, namely the 49 individual phage-host PPIs. Host proteins were consequently used as MBP-tagged baits and phage proteins as GFP-luciferase-tagged preys (GFP-luc).

37 PPIs (75.5%) were above the cutoff of the luminescence intensity ratio (sample relative compared to MBP negative control) $LIR > 3$ (Fig. 30). The results indicate that the dataset is of high quality including a wide range of PPIs that are reproducible by this alternative methodology. A conservative threshold of $LIR = 3$ is commonly used for LUMIER/LuMPIS assays (Barrios-Rodiles et al., 2005; Vizoso Pinto et al., 2009). The remaining interactions that were below the threshold might be false-positives due to their non-reproducibility. This leads to the assumption that the subset tested includes $\sim 1/4$ false-positives. Since there were no crucial methodical differences compared to the intra-viral PPIs this number is probably similar for the Cp-1 and Dp-1 intra-viral PPI datasets. However, it was shown that many interactions which are not detectable by one method can be detected by others (Braun et al., 2009). This means that the interactions below the LIR threshold may be reproducible by other methods and represent LuMPIS false-negatives.

In Fig. 30A/B I aligned the 3-AT scores of the Y2H retest experiments with decreasing LIR values. This was done to see if the signal intensities correlate or if there are general differences between Y2H and LuMPIS signal intensities. A linear correlation between LIR and the 3-AT score was expected. As shown in Fig. 30D the 3-AT and LIR values correlate weakly but positively. In general, high 3-AT scores are connected to high LIR values and vice versa. However, there are many exceptions

RESULTS

underlining that certain interactions exhibit preferentially higher signals in Y2H or by LuMPIS. This is not unexpected since the methods differ substantially.

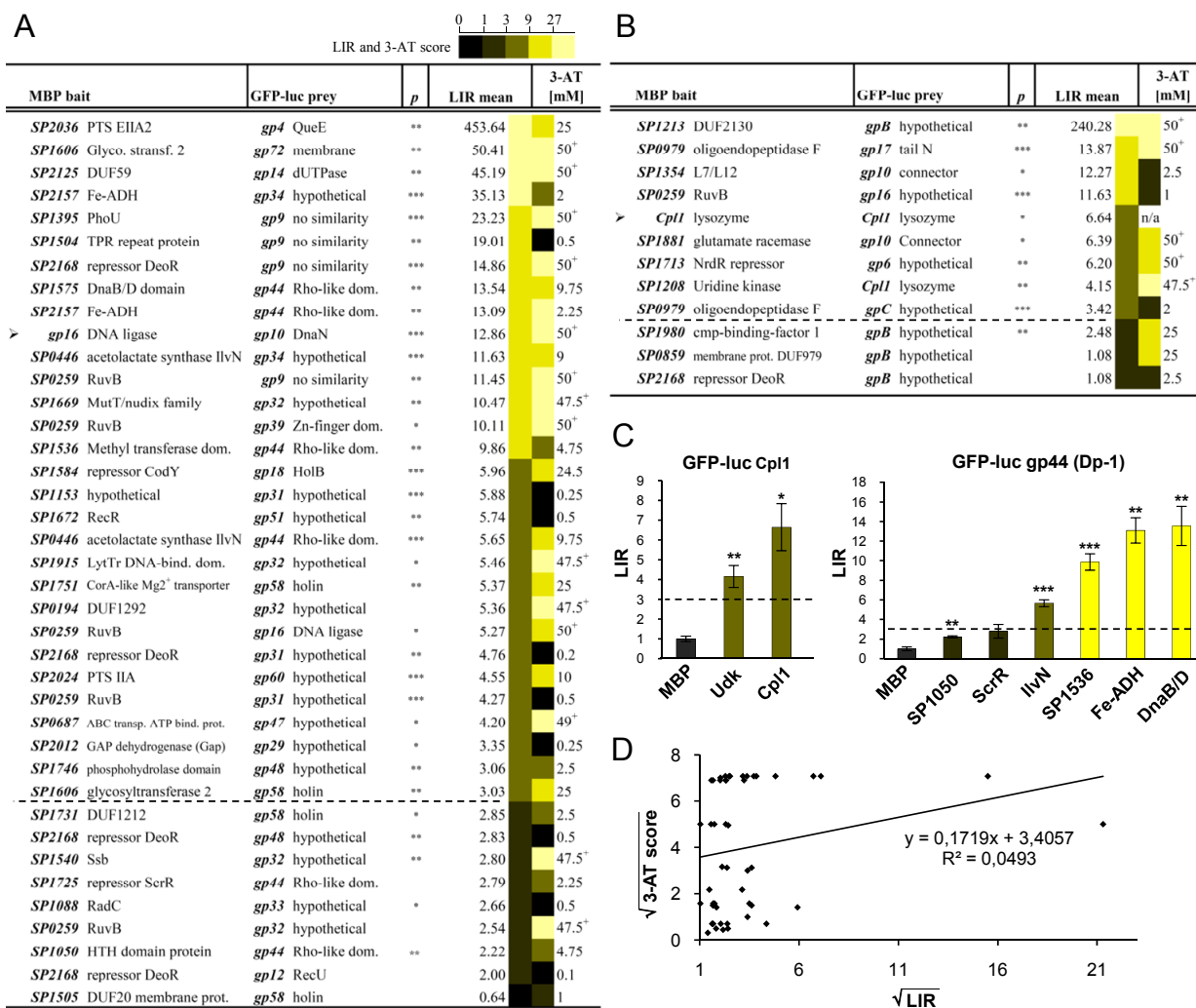


Fig. 30 LuMPIS verification of phage-host PPIs identified in the Y2H screens

(A, B) LuMPIS assay results for Dp-1- and Cp-1-host protein interactions. Test pairs are sorted by decreasing LIR mean values. Tests were done as quadruplicates and thereof the LIR mean value was calculated. The dashed lines flag the LIR threshold ($=3$). (p) indicates the mean error probability between an MBP negative control and a single test pair calculated by a T-test. Significant p -values are symbolized as follows: (***) $p < 0.001$; (**) $0.001 \leq p < 0.01$, (*) $0.05 < p < 0.1$. Furthermore, the LIR mean value of each individual PPI is compared with the 3-AT score determined by 3-AT titration in the Y2H retests (see above). A color gradient indicates the signal intensity of LuMPIS and Y2H (see Fig. legend above Tab. A). (>) labels additionally tested phage-phage PPIs. The Dp-1 gp16-gp10 PPI was identified in the previous intra-viral Y2H screen by both vector systems and reciprocally as AD/DBD-fusion proteins. Thus, the interaction was assumed as very reliable and used as a positive control. The Cp-1 PPI Cpl1-Cpl1 was detected by Y2H screens but could not be clearly reproduced in the Y2H retest. However, based on protein cross-link assays (see below) a homomerization interaction was expected and thus tested by this alternative method. (C) Two test examples presented as histograms. GFP-luc tagged phage preys were tested against MBP only (negative control) and a set of MBP-tagged host preys. Error bars indicate the standard errors of the LIR mean values. Asterisks indicate significant low error probabilities as given under in (A) and (B). (D) Regression analysis of LIR mean values of individual PPIs plotted against their corresponding 3-AT scores. LuMPIS assays were done in collaboration with Maria Guadalupe Vizoso, Max von Pettenkofer Institute, München.

3.1.5 Characterization of the Cp-1 lysozyme - host uridine-cytidine kinase interaction

In the Y2H screens an unexpected interaction was detected: the Cp-1 lysozyme binds with *S. pneumoniae* uridine-cytidine kinase (Udk), an enzyme involved in the pyrimidine salvage pathway (Parks and Agarwal, 1973). It transfers the γ -phosphoryl group from ATP or GTP onto uridine or cytidine resulting in UMP or CMP, and ADP or GTP (Ahmed and Welch, 1979) (Fig. 33 Fig. 33A). Udk is a non-essential enzyme. As already mentioned above, a genetic link between T7 phage and *E. coli* Udk was found by Qimron and colleagues (Qimron et al., 2006) indicating that this enzyme might play a general role in podovirus reproduction. Thus, I focused on analysis of this particular PPI. The interaction is “unexpected” since the lysozymes are in fact known to be responsible for the lysis step by hydrolyzing the peptidoglycan layer whereas a secondary extra-lytic function would be surprising.

The Cpl1 cell wall binding module mediates the interaction with Udk

The Cpl1 structure itself has been solved (Hermoso et al., 2003). It consists of three domains that can be divided into two functional modules (Fig. 31D): the N-terminal hydrolase domain is responsible for peptidoglycan degradation. It is connected via a short acidic linker to the C-terminal cell wall binding module that consists of cell wall binding domain cI and cII. The latter bind to choline in the peptidoglycan and are thought to orientate the enzymatic domain correctly.

In order to map the domains involved in this PPI, I tested by Y2H experiments several Cpl1 truncation constructs. I systematically truncated Cpl1 into nine fragments by cloning (Fig. 31D) and tested them against the Udk full-length construct. Udk is a globular single-domain protein (Fig. 32C) and thus not useful to generate truncated constructs since this might have destroyed the protein's integrity.

No positives were detected for constructs that contained the catalytic domain indicating that it is not involved in contacting Udk (except the full-length clone). However, constructs that carried the hydrolase domain were self-activators. Consequently, the catalytic domain is responsible for the Y2H self-activation property.

Surprisingly, four truncations involving the cell wall binding module were tested positively. The smallest construct that was still able to maintain the PPI represents a minimal construct of cell wall binding domain cI (residue 200-281). In contrast, when the cII minimal construct (282-339) was tested, the interaction was lost, indicating that cI is the interacting subdomain of the cell wall binding module.

In comparison, the T7 lysozyme only consists of a catalytic domain (no cell wall binding domain is present given the difference between gram+ and gram- hosts). This is interesting since Cpl1 mediates the contact via the cI domain which is not present in the T7 lysozyme. This could indicate that Cp-1 and T7 might have evolved different, analogous strategies either to bind to the host Udk or to affect it by a completely different mechanism. There exists no information if the T7 lysozyme can bind on the protein level to Udk. Furthermore, the Dp-1 cell wall hydrolase PalI that contains homologous cell wall binding repeats was tested negatively with Udk (Fig. 31C, consider also Fig. 19) indicating that

RESULTS

the PPI with Udk is a species specific property of Cp-1 and might be not relevant for siphoviruses. Moreover, I also checked for Udk if it is able to bind with itself since this was expected from literature information of the human Udk homolog (Suzuki et al., 2004). This was confirmed (Fig. 31B, last test position).

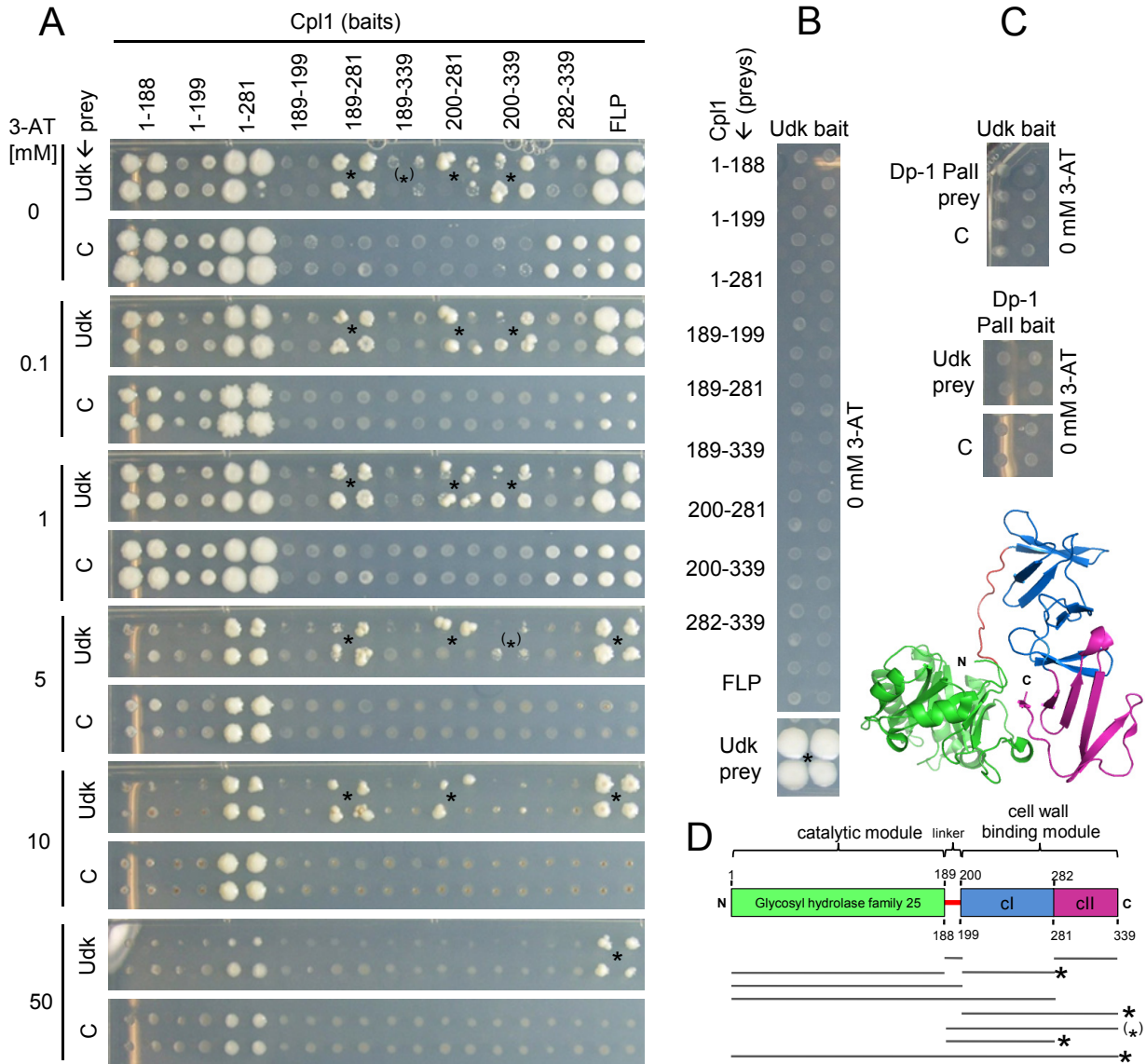


Fig. 31 Domain interaction mapping of Cpl1-Udk PPI

Cpl1-Udk interaction mapping by Y2H experiments. **(A)** Several Cpl1 truncation constructs tested as baits against the Udk prey. The figure shows images taken from diploid yeast on $-His$ readout medium after 7 d growth at $30^{\circ}C$ on varying concentrations 3-AT. Pairs were tested in the pDEST32/22 expression vector system as quadruplicates. Given numbers indicate the amino acid range of certain constructs (D). (FLP) Cpl1 full-length construct. **(C)** Self-activation control with empty pDEST22 prey plasmid and the bait construct. Asterisks indicate positively tested constructs on the corresponding concentration of 3-AT. Asterisks in brackets label weakly appearing colonies which are clearly above the background. **(B)** The same test constellation but bait-prey tag direction is interchanged. A homomerization PPI of Udk itself was detected (Udk/Udk). **(C)** reciprocal test of Dp-1 cell wall hydrolase (Pall, gp59) with Udk. **(D)** 3D-structure of Cpl1 and illustration of the linear Cpl1 domain composition. Colors are the same in the linear and 3D-image. Bars below indicate the tested Cpl1 constructs and their amino acid residue range. Asterisks highlight positively tested constructs that interact as fragments with Udk. 3D-structure was drawn with Pymol after (Hermoso et al., 2003) (PDB entry 1H09).

Cpl1 binds to Udk in vitro

The Cpl1-Udk interaction was detected in the Y2H screens and was reproducible in a Y2H retest experiment. Furthermore, it was confirmed alternatively by LuMPIS (Fig. 30C) indicating that the interaction is also stable *in vitro*.

To learn more about this PPI, I purified Cpl1 and Udk as N-terminally His₆-tagged proteins for biochemical assays. This was done since the human Udk ortholog is known to function as a tetrameric complex (Suzuki et al., 2004). I wanted to check whether Cpl1 has an influence on the quaternary state of *S. pneumoniae* Udk and thus if it can destroy the enzyme's function just by interfering with the integrity of its quaternary structure.

To determine the quaternary structures of Cpl1 and *S. pneumoniae* Udk I chemically cross-linked the proteins with glutaraldehyde either individually or in a mix of both proteins (Fig. 32A) (2.2.2.9). The Udk sample showed a main band between 100 and 130 kDa. This band is in agreement with the size of a tetra-homomeric complex. The molecular weight (MWG) of a monomer is approx. 28.8 kDa and of a tetramer 115.2 kDa and fits between the corresponding marker bands. Some less intense smaller bands occurred, that could represent free homo-, di-, and trimers. In contrast, cross-linked Cpl1 (MWG of homomer is approx. 42.0 kDa) led to a ladder-like pattern indicating that self-interaction also occurs here while the vast majority of the protein forms high molecular weight aggregates. When Cpl1 and Udk were combined all Udk bands disappeared indicating that the Udk complexes are incorporated into the Cpl1 aggregates. However, Udk mono- to tetramers seem to be mainly cross-linked with the dominant high MWG aggregates of Cpl1 and thus no novel heteromeric bands appeared on the gel running below 170 kDa. In summary, these results show that (i) Udk of *S. pneumoniae* is preferentially organized as a tetramer similar to the human homolog, (ii) that it interacts with Cpl1, (iii) and that it is incorporated into the Cpl1 oligomeric complexes. Thus, Cpl1 might have no direct influence onto the homotetrameric structure of Udk and vice versa.

Because of this finding I applied the purified proteins to gel filtration since the ladder-like pattern of Cpl1 could have been also attributed to the cross-link *in vitro* situation. As shown in Fig. 32B the majority of Cpl1 migrates as a monomer and only a small portion as a high-MWG complex. Thus, the ladder-like pattern was probably an artifact of the cross-linking condition. The main peak of the Udk sample appeared later than Cpl1 and likely represents the tetramer. The large intensity differences between Udk and Cpl1 are because Udk contains less aromatic amino acid residues compared to Cpl1 (Cpl1 has thus a stronger absorption). When Udk and Cpl1 were mixed stoichiometrically, approx. half of Cpl1 and Udk was converted into high-MWG complexes, supporting the results of the cross-linking assay.

In summary, both proteins show homomerization interactions but do also bind to each other when they are mixed *in vitro* or tested by Y2H or LuMPIS assays. These results show that Cpl1 is able to convert Udk into high-MWG aggregates. However, it remains unclear what the functional consequences of

this interaction is. It is likely that the Udk tetramer is converted or incorporated into Cpl1 oligomers *in vitro* and thus that the tetramers are maintained.

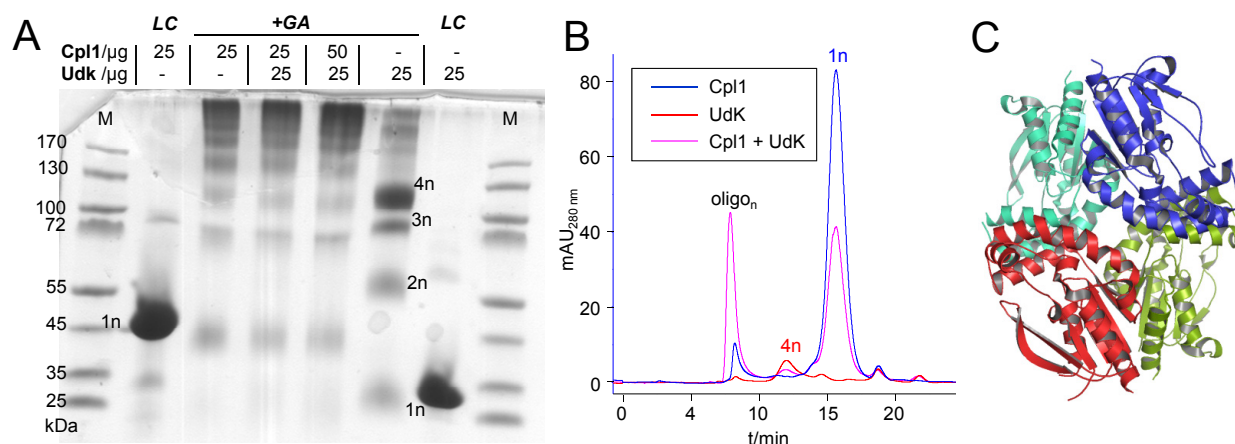


Fig. 32 Effects of Cpl1 onto the quaternary structure of Udk

(A) Cpl1/Udk protein cross-link assay with glutaraldehyde (GA). Purified, N-terminally His₆-tagged proteins were chemically cross-linked with GA and then separated on a tricine-SDS-PAGE gradient gel, stained afterwards with Coomassie Brilliant Blue R-250. (LC) loading controls of non-GA-treated protein samples, (M) protein marker. The quaternary protein states are labeled (e.g., by 1n, 2n, ...). The assay was repeated 3 x with the same result. (B) Gel filtration of Udk and Cpl1. The experiment was done under neutral pH conditions (PBS pH7.4). Equal protein amounts of 100 µg were applied to the column individually or in combination. Expected quaternary structures are indicated in the corresponding colors. Gel filtration was done in collaboration with Daniel Nelson and Caren Stark, UMBI, Rockville, MD. (C) Quaternary 3D-structure of human Udk ortholog (UCK). The figure shows UCK in a ligand unbound state. Different subunits are indicated by different colors. Figure was drawn with PyMOL after (Suzuki et al., 2004) (PDB entry 1UFQ).

Cpl1 does not affect the Udk enzyme activity in vitro

Although the gel filtration and cross-linking experiments demonstrated that the integrity of the functional Udk tetramer might be not influenced by Cpl1, I tested if the Cpl1 interaction has an effect on Udk's enzymatic activity. For instance, it is possible that Cpl1 blocks the substrate binding pocket directly or allosterically. If so, differences should have been measurable by testing the enzymatic activity of Udk in the presence of Cpl1. Qimron and colleagues already assumed that the genetic link of T7 with *E. coli* Udk might be an indicator that at least T7 could manipulate the pyrimidine salvage pathway although it is uncertain which T7 component might be responsible (Qimron et al., 2006).

I applied the purified proteins to an enzymatic assay by using an experimental setup established for the human uridine-cytidine kinase (Ahmed and Welch, 1979) (Fig. 33). The products/educts were determined qualitatively by thin layer chromatography assuming that Cpl1 affects Udk activity drastically. While the *S. pneumoniae* Udk is able to convert uridine to UMP (shown for the first time), Cpl1 itself had no crucial effect on Udk activity after several time points even when tested in a 10 fold molar excess. At least under *in vitro* conditions, no effects are detectable. This leads to the conclusion that (i) Cpl1 does not block Udk activity *in vitro*, (ii) and thus the pyrimidine salvage pathway might not be influenced *in vivo*, (iii) although the interaction was demonstrated to occur *in vitro*, the binding epitope is likely to be outside the Udk-homomerization sites and the active center. I conclude that

RESULTS

Cpl1 does not destroy the Udk tetramers since the tetrameric integrity is essential to maintain its enzymatic activity (Suzuki et al., 2004). Thus, Cpl1 has no effect onto the quaternary structure and the enzymatic activity. Although this interaction is reproducible by several independent methods the physiological relevance might be not to manipulate the host's pyrimidine salvage pathway.

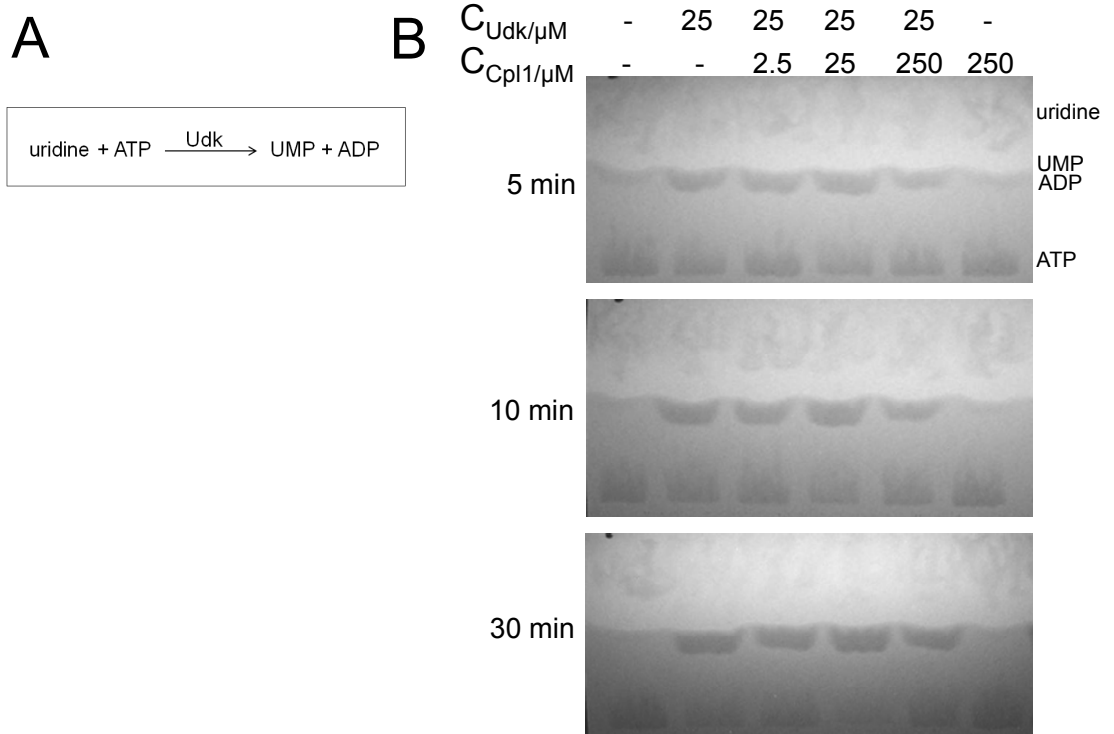


Fig. 33 Competitive Udk enzyme assay

(A) Udk reaction. Udk transfers the gamma-phosphoryl group from ATP onto uridine resulting in UMP and ADP. (B) Competitive enzyme assay. Purified proteins (N-terminally His₆-tagged) were used individually or combined while keeping a constant Udk concentration. Cpl1 was applied with an equal molar concentration and a 10 fold molar excess or depletion. Samples were taken after 5, 10, and 30 min and separated by thin layer chromatography on DEAE cellulose with 0.03 N HCl mobile phase. The educts and products were visualized under UV light (254 nm). The educt/product bands are indicated for the 5 min samples, respectively. The single images were compressed 50% in height.

3.2 Functional analysis of the conserved hypothetical protein YbeB

E. coli protein YbeB and its orthologs such as plant Iojap are proteins of hitherto unknown function that are nearly universally conserved from bacteria to human. Here, I show that YbeB family members bind to ribosomal protein L14 in most tested species, including *E. coli*, *Treponema pallidum*, *Synechocystis PCC 6803*, *Streptococcus pneumoniae* as well as human mitochondria, and *Zea mays* chloroplasts.

Although an *E. coli* YbeB knock-out has no obvious phenotype, GFP and β -galactosidase reporter proteins get translated more rapidly than in wild-type cells. Thus, YbeB appears to be a negative regulator of translation *in vivo* by inhibiting ribosomal subunit assembly, given its association with L14 on the 50S-30S interface.

Most proteins of wide phylogenetic distribution are well-characterized. However, a small group of highly conserved hypothetical proteins remain uncharacterized, usually because they have no obvious phenotypes when mutated. Unfortunately, even a thousand completely sequenced genomes have not helped to annotate some of these highly conserved genes (Kazuta et al., 2008).

Escherichia coli YbeB (b0637) is a conserved hypothetical protein that belongs to cluster of orthologous genes (COG0799) and can be found from bacteria to man. Galperin and Koonin ranked YbeB and its orthologs third among uncharacterized proteins for further experimental studies (Galperin, 2001). However, a few clues as to YbeB's function have emerged over the years. Most importantly, it was shown to co-migrate exclusively with the mature *E. coli* large ribosomal subunit and several studies identified interaction partners of YbeB orthologs in various species (Butland et al., 2005; Gavin et al., 2006; Jiang et al., 2007; Parrish et al., 2007; Rain et al., 2001; Titz et al., 2008). However, no study has followed up on these preliminary observations.

3.2.1 High evolutionary distribution of YbeB homologs by low sequence conservation

YbeB orthologs are widely distributed: The cluster of orthologous group COG0799 currently includes at least 506 YbeB-related protein sequences in 505 species (STRING 8.2) (von Mering et al., 2007). HMM signatures are more sensitive and are able to detect the YbeB/DUF143 signature (Pfam entry PF02410, Pfam V24.0) in 1,497 individual protein sequences distributed among 834 species (Finn et al., 2010). *E. coli* YbeB is a core representative of this family since it consists almost exclusively of the DUF143 domain (105 amino acids, (Tab. 30)). Given that YbeB interacts with L14 (L14p/L23e, Pfam entry PF00238) (this has to be mentioned here in advance) I wondered whether the DUF143/YbeB family members are as widespread as this ribosomal protein and projected both DUF143 and L14 HMM signatures onto the iTol tree of life (Hunter et al., 2009; Letunic and Bork, 2007) (Fig. 34).

While L14 is virtually universally present (with only one possible exception, namely the unicellular eukaryote *Thalassiosira pseudonana*), YbeB is only slightly more restricted: in Eubacteria DUF143

orthologs are present in nearly all species except *Mycoplasma*, *Clostridia*, and *Buchnera* and very few other exceptions. In Eukaryota DUF143 signatures are present in nearly all representative species in the tree. The signature is conspicuously absent in Archaea. Pfam HMMs are highly sensitive for detection of conserved sequence patterns. Since YbeB homologs exhibit overall a low sequence conservation (see below), there might be still the chance that the species that lack YbeB/DUF143 homologs contain such a “homolog” with an extra low sequence conservation but with a conserved 3D-structure and failed a detection by the HMM.

YbeB orthologs are less conserved on the sequence level: While YbeB’s DUF143 signature shows nearly evolutionary omnipresence, its amino acid sequence is surprisingly not as well-conserved as in ribosomal proteins (Fig. 35A). The DUF143 core sequence exhibits only three amino acid residues which are absolutely conserved in all sequences. In contrast, an alignment of L14 from the same species reveals 10 absolutely conserved residues (Fig. 35B). While the strict conservation of L14 can be explained by its functional necessity to undergo conserved interactions with ribosomal protein L19 as well as 23S and 16S rRNA, the functions of the few residues that are conserved in YbeB remain unknown. This suggests that YbeB function requires fewer conserved sites. Interestingly, the sequence length of YbeB homologs (105-612 amino acids) is much more flexible than that of L14 (122-145 amino acids) (Tab. 30). Eukaryotic DUF143 orthologs typically contain an N-terminal signal peptide and thus have variable N-termini. The C-terminus of DUF143 proteins is also flexible in length. For example, the human ortholog C7orf30 has 40 additional residues that are not covered by the DUF143 signature. The main exception is the yeast ortholog ATP25 which contains 366 C-terminal residues. Its precursor polypeptide is cleaved after residue 292 by an unknown mechanism as soon as it has been imported into the mitochondria, resulting in two cleavage products: the C-terminal fragment specifically stabilizes mRNA of ATPase subunit C (Oli1) while the function of the DUF143 containing product is unclear (Zeng et al., 2008).

Eukaryotic YbeB orthologs localize to organelles of eubacterial origin: Given the conservation of YbeB/DUF143 in Eubacteria, I wondered whether this protein family would be localized to plastids and mitochondria in Eukaryotes. I used WoLF PSort to predict the subcellular localization of all known eukaryotic proteins containing a DUF143 and found an overall high support for localization in eukaryotic organelles of eubacterial origin: 82% of the eukaryotic family members were predicted to be localized to mitochondria or chloroplasts (Horton et al., 2007). For the yeast ortholog ATP25, the localization to mitochondria was also experimentally confirmed (Zeng et al., 2008) and the *Zea mays* ortholog, Iojap, was found in chloroplast fractions (Han and Martienssen, 1995).

YbeB orthologs appear mainly as single-domain proteins: In Tab. 31 I summarized representative proteins which contain the DUF143 domain in combination with other domains or motifs. In the vast majority of 1,874 cases (99.2%), DUF143 signatures occur as single domain proteins as in *E. coli*

RESULTS

YbeB. Interestingly, in nine bacterial homologs the DUF143 domain co-occurs with an N-terminal cytidyltransferase domain. Most of these proteins are annotated as putative nicotinate (Nicotinamide) nucleotide adenylyltransferase (nadD), a key enzyme for synthesis of NAD⁺. However, there is no experimental evidence that links DUF143 to this pathway. Whether the remaining domain fusions shown in Tab. 31 are functional proteins or evolutionary artifacts remains to be seen, although the occasional fusion to nucleic acid-binding domain is suggestive of a DNA- or RNA-related function.

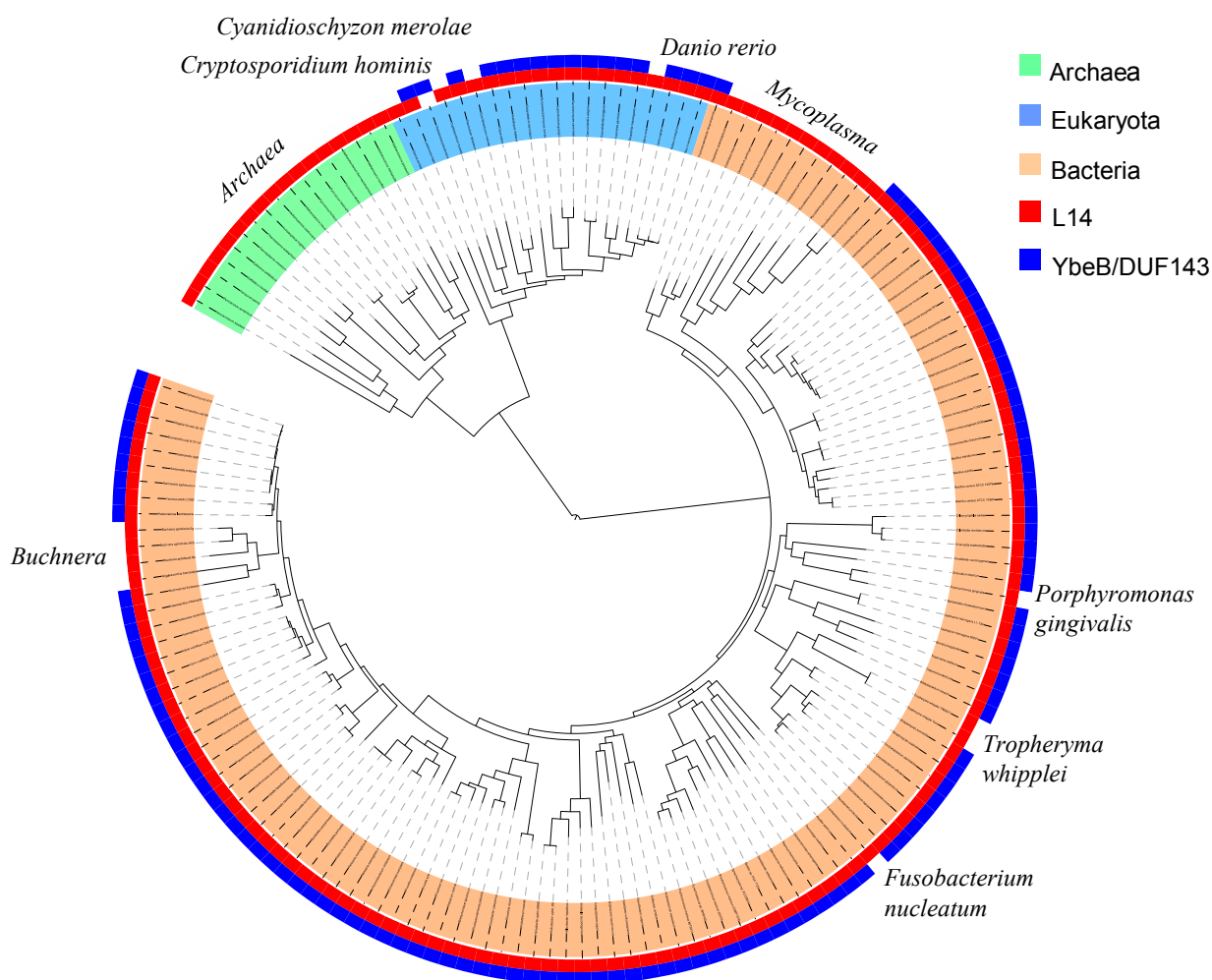


Fig. 34 YbeB/DUF143 orthologs are widely distributed but absent in a few clades

Presence of an YbeB ortholog in a species is labeled in blue, L14 in red. Only a few groups seem to lack YbeB/DUF143, most notably the archaeal, *Clostridium*, and *Mycoplasma* clades (highlighted by their names in the figure). DUF143 (Iojap-related) and L14 signatures were projected onto the iTol tree of life (V1.7) (Letunic and Bork, 2007) using entries IPR004394 and IPR000218 from Interpro V26.0 (Hunter et al., 2009). Phylogenetic analysis was done in collaboration with Björn Titz, Crump Institute for Molecular Imaging, Los Angeles.

Tab. 30 Study-related YbeB and L14 orthologs

(Architecture) illustrates the domain architecture (including predicted signal peptides indicated by rectangles) while Pfam DUF143 is the position of the HMM signature. (PDB entries) known 3D-structures. (SigP) indicates predicted signal peptides. (Total protein length) in amino acid (aa) residues. Information is based on Uniprot entries. Protein descriptions are completed with information from species-specific databases (SGD, EcoCYC, EMBL). References for localization and signal peptides: ⁽¹⁾ experimental evidence *in vivo* (this work), ⁽²⁾ experimental (Ishihama et al., 2008), ⁽³⁾ experimental evidence by chloroplast fractionating, *Zea mays* (Han and Martienssen, 1995), ⁽⁴⁾ experimentally by mitochondrion isolation/LC-MS in yeast (Reinders et al., 2006), ⁽⁵⁾ by plastid genome sequence, plastid-encoded (Markmann-Mulisch and Subramanian, 1988); predicted signal peptides ⁽⁶⁾ by SignalP 3.0 (Emanuelsson et al., 2007); ⁽⁷⁾ by TargetP 1.1 Server (Emanuelsson et al., 2007), predicted to localize to chloroplasts by PSORT II (Nakai and Horton, 1999).

YbeB homologs

Species	Primary locus tag/ name/ Uniprot Ac	Gene product description	3d	Localization	Pfam DUF143	SigP	Architecture	Protein length/aa
<i>Synechocystis</i> PCC 6803	slr1886/-/P73658	Putative Iojap protein			30-129			154
<i>E. coli</i> K12	b0637/ybeB/P0AAT6	50S subunit associated protein		Cytoplasm ²	6-104			105
<i>Treponema pallidum</i>	TP0738/-/O83720	Putative uncharacterized protein			6-104			111
<i>S. pneumoniae</i> TIGR4	SP1744/-/Q97P97	Iojap-related protein			6-104			117
<i>Homo sapiens</i>	C7orf30/-/Q96EH3	Uncharacterized protein		Mitochondria ¹	93-194	1-22 ⁶		234
<i>Zea mays</i>	Iojap/-/Q41822	Iojap		Plastides ^{3,5}	116-213	1-62 ⁷		228
<i>S. cerevisiae</i>	YMR098C/ ATP25/Q03153	Mitochondrial protein required for stability of Oii1p (Atp9p) mRNA and for the Oii1p ring formation		Mitochondria ⁴	104-246			612
<i>Chromobacterium violaceum</i>	CV0518/-/Q7P0P8	Putative uncharacterized protein	2o5a		6-104			122
<i>Bacillus halodurens</i>	BH1328/-/Q9KD89	BH1328 protein	2id1		6-104			117



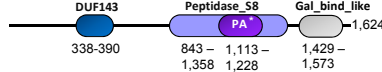
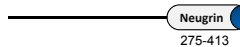
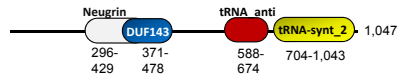


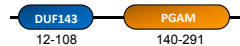
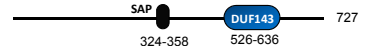
L14 homologs

Species	Primary locus tag/ name/ Uniprot Ac	Gene product description	Pfam L14p/L23e	Protein length/aa	SigP	Localization
<i>Synechocystis</i> PCC 6803	slr1806/ rplN/P73310	50S ribosomal protein L14	1-122	122		
<i>E. coli</i> K12	b3310/ rplN/P0ADY3	50S ribosomal protein L14	1-123	123		Cytoplasm ²
<i>Treponema pallidum</i>	TP0199 rplN /O83229	50S ribosomal protein L14	1-122	122		
<i>S. pneumoniae</i> TIGR4	SP0219/rplN/P0A473	50S ribosomal protein L14	1-122	122		
<i>Homo sapiens</i>	MRPL14/L14 _{mt} /Q6PIL8	39S ribosomal protein L14, mitochondrial	31-145	145	1-19 ⁶	Mitochondria ¹
<i>Zea mays</i>	rpl14/-/P08529	50S ribosomal protein L14, chloroplastic	1-123	123		Plastides ⁵
<i>Saccharomyces cerevisiae</i>	YKL170W/MRPL38/P35996	54S ribosomal protein L38, mitochondrial	1-138	138		Mitochondria ⁴
<i>Chromobacterium violaceum</i>	CV4176/rplN/Q7NQG2	50S ribosomal protein L14	1-122	122		
<i>Bacillus halodurens</i>	BH0144/rplN/Q9Z9K4	50S ribosomal protein L14	1-122	122		

RESULTS

Tab. 31 DUF143/Iojap signatures appear mainly as single-domain proteins

Signatures based on domain information from Interpro 26.0 (Hunter et al., 2009). (Frequency) gives the number of proteins with this domain architecture. (Domain architecture) numbers indicate the sequence range of the corresponding domain. NA binding „X“ (last column) indicates that one co-occurring domain is known to bind to nucleic acids.

Representative species	Uniprot accession	Protein description	Frequency	Domain architecture	NA binding
<i>E. coli</i> K12	P0AAT6	50S subunit associated protein	1,874		
<i>Campylobacter hominis</i>	A711F9	Putative nicotinate nucleotide adenylyltransferase	9		
<i>Micromonas pusilla</i> CCMP1545	C1MXG4	Predicted protein	1		X
<i>Malassezia globosa</i>	A8Q3H1	Putative uncharacterized protein	1		
<i>Ustilago maydis</i> (Smut fungus)	Q4P5K6	Putative uncharacterized protein	1		X
<i>Eubacterium saphenum</i> ATCC 49989	C7GZ26	Metal dependent phosphohydrolase	1		
<i>Thermosinus carboxydivorans</i> Nor1	A1HNX5	Iojap-like protein	1		X
<i>Propionibacterium acnes</i> J139	D1Y9U9	Iojap-like protein	1		
<i>Micromonas sp.</i> RCC299	C1EE68	Predicted protein	1		X

3.2.2 Systematic examination of conserved docking sites of YbeB on the large ribosomal subunit

As already mentioned above, recent large scale studies revealed many PPI partners of YbeB homologs from various organisms. However, no clear link of YbeB function could be indicated since the PPI partners play a role in various cellular processes. All known PPIs for YbeB homologs are listed in Tab. 37. A previous study could show that *E. coli* YbeB co-migrates exclusively with the large ribosomal subunit and thus might functionally participate in protein translation (Jiang et al., 2007). Since the 50S subunit surface offers many possibilities at many positions for factors to bind, I wanted to know where YbeB is able to bind exactly. The expected results could limit YbeB's mode of action on the 50S subunit, e.g., if it might behave as a common chaperone, if interacting with various ribosomal proteins or if it has a specific function at a certain 3D-site on the 50S. To get deeper insights I checked systematically for putative binding sites of YbeB by testing several PPIs known from high-throughput experiments that link YbeB with the translational but also other processes.

Interaction partners of YbeB in T. pallidum: which translational components interact with YbeB?

Initially, I screened the YbeB homolog TP0738 of *Treponema pallidum*, the Syphilis spirochete, for potential binding partners by using a proteome-wide Yeast Two-Hybrid screen as described previously (Titz et al., 2008). This screen detected a total of 62 interaction partners (see suppl. Tab. 36 for a complete list). Next, I retested all 10 interactions that involved translation-related proteins (based on TIGR CMR cellular main role, (Davidsen et al., 2010)). Nine proteins tested positives (Fig. 36A): proteins of the large ribosomal subunit (L9, L14, L29, and L32), InfA (translation initiation factor 1), Def (peptide deformylase), ThiI (tRNA 4-thiouridine synthase), GatC (glu-tRNA amidotransferase, subunit C), and RimM (putative 16S rRNA processing protein). The interaction with L14 was by far the most stable when measured by the concentration of 3-AT, a competitive inhibitor of the reporter gene, HIS3 (Fig. 36A, left). The other interactions were relatively weak, with none of them detectable at more than 1 mM 3-AT as opposed to 50 mM for L14. Furthermore, the L14-YbeB interaction was the only one that was detectable in a reciprocal screen, that is both proteins could be used as bait or prey fusion. Thus, the L14-YbeB interaction is the most stable one in this tested set while all others might be transient interactions which may be physiologically irrelevant.

Which interaction is conserved?

I wondered if any interactions determined in *T. pallidum* as well as others from various published datasets are conserved in other species. Since YbeB is evolutionary widely distributed/conserved I expected that some of its interaction partners have to undergo a conserved interaction with YbeB. To identify these cases I tested by another Yeast Two-Hybrid experiment interologous pairs ("interologs" means conserved pairs) of *E. coli*, except GatC since *E. coli* lacks a homolog. Moreover, I tested eight putative interaction partners that have been identified in a protein complex together with *E. coli* YbeB

RESULTS

and L14 by Butland and colleagues (Butland et al., 2005) and four orthologous pairs detected by Y2H screens in *Campylobacter jejuni* (Parrish et al., 2007). Surprisingly, only the interaction with L14 turned out to be conserved, again detectable as AD/DBD reciprocal pairs and up to 50 mM 3-AT (Fig. 36B). All other tested pairs were negative and are summarized in Tab. 32. Thus, I concluded that L14 is the primary and conserved binding target of YbeB on the 50S ribosomal subunit.

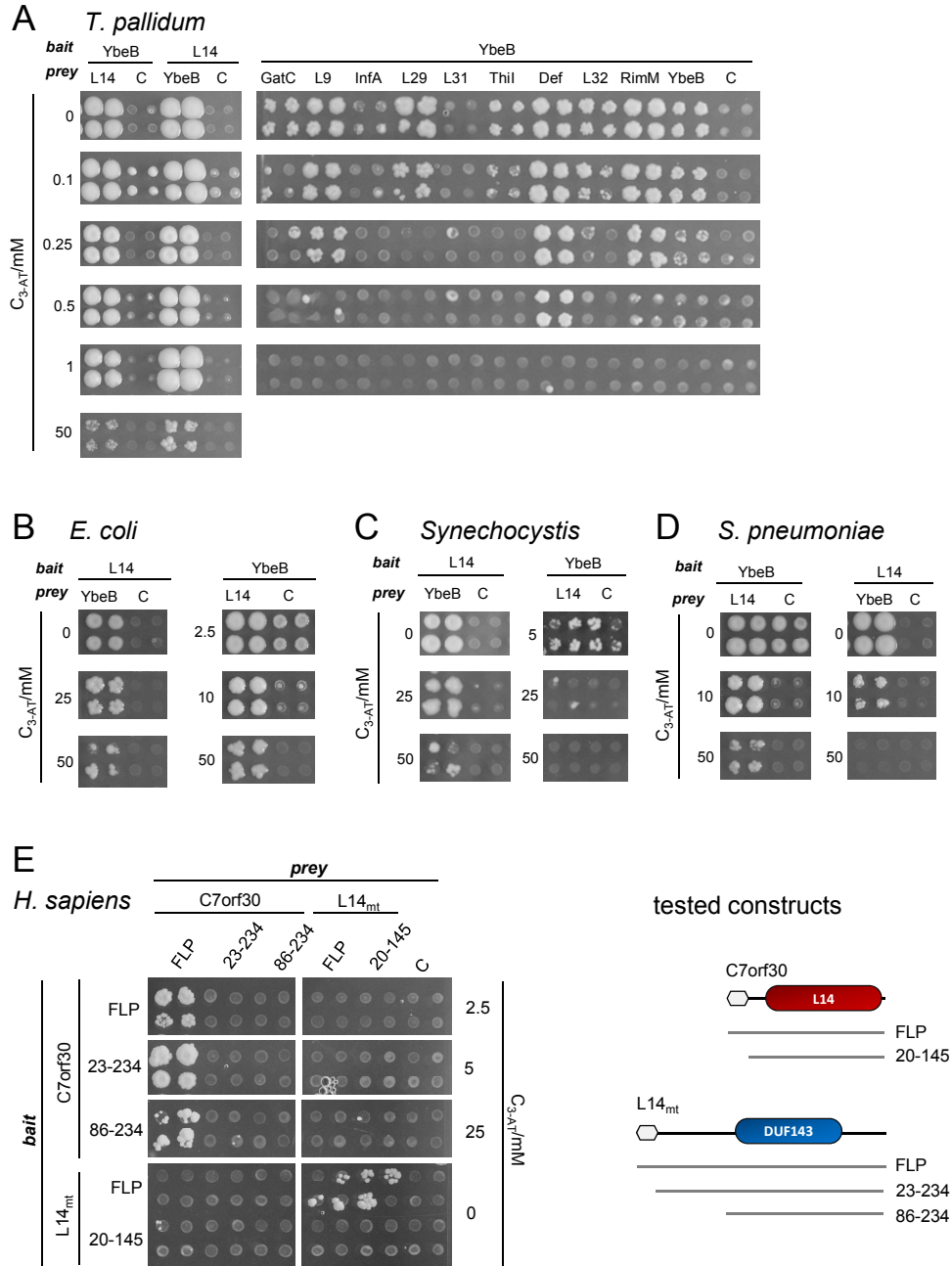


Fig. 36 Mapping of conserved protein docking sites of YbeB with ribosome-related proteins
Panels show Y2H tests with various concentrations of 3-AT (mM) in quadruplicate tests. “C” control (self-activation test with empty prey vector). Expression vectors pGBKT7g/pGADT7g were used. (A) Interactions of *T. pallidum* YbeB (TP0738) with proteins involved in translation. YbeB binds specifically to L14 but also weakly to other ribosomal proteins in Y2H tests. Only YbeB-L14 shows a signal above 0.5 mM 3AT. Interologous tests of (B) *E. coli* L14/YbeB. See also Tab. 32. (C) *Synechocystis PCC 6803*. (D) *Streptococcus pneumoniae TIGR4*. (E) Human YbeB homolog C7orf30 and mitochondrial L14_{mt} as full-length (FLP) and N-terminally truncated constructs (numbers indicate the range of amino acids and are illustrated to the right of (E)).

RESULTS

Tab. 32 Interactions of YbeB that are not conserved in *E. coli*

Here, all putative interaction partners are listed which have been tested negatively with *E. coli* orthologs by Y2H experiments. Orthologs from *T. pallidum* and *C. jejuni* were selected by MBGD orthologous protein groups (Uchiyama et al., 2010). Gene product descriptions were taken from Uniprot entries of *E. coli* homologs. The reference set gives the source of YbeB orthologous interactions they were primarily described in. The interactions listed from Butland et al. are proteins that have been co-purified as protein complex of *E. coli* YbeB.

Name	<i>E. coli</i> locus tag	Reference species	Locus tag ortholog	Gene product description	Reference interaction set		
MiaB	b0661	<i>C. jejuni</i>	Cj1454c	tRNA-i(6)A37 methyltransferase	(Parrish et al., 2007)		
SerS	b0893		Cj0389	Seryl-tRNA synthetase			
HemA	b1210		Cj0542	Glutamyl-tRNA reductase			
S10	b3321		Cj1708c	30S ribosomal protein S10			
DhaK	b1200	<i>E. coli</i>	-	Dihydroxyacetone kinase, N-terminal domain	(Butland et al., 2005)		
YehL	b2119		-	Uncharacterized protein yehL			
YehQ	b2122		-	predicted protein			
L19	b2606		-	50S ribosomal subunit protein L19			
Cca	b3056		-	Multifunctional CCA protein			
L4	b3319		-	50S ribosomal protein L4			
YihU	b3882		-	Uncharacterized oxidoreductase yihU			
L7/L12	b3986		-	50S ribosomal subunit protein L7/L12			
ThiI	b0423		<i>T. pallidum</i>	TP0559		tRNA sulfurtransferase	(Titz et al., 2008), this work
InfA	b0884			TP0097		Translation initiation factor IF-1	
L32	b1089	TP0807		50S ribosomal protein L32			
RimM	b2608	TP0907		Ribosome maturation factor RimM			
Def	b3287	TP0757		Peptide deformylase			
L29	b3312	TP0197		50S ribosomal protein L29			
L9	b4203	TP0060		50S ribosomal protein L9			

Given the wide phylogenetic presence of both YbeB and L14, I wanted to know if the YbeB-L14 interaction is conserved ubiquitously. Therefore, I tested in addition the pair in one Gram-positive bacterium (*Streptococcus pneumoniae* TIGR4) and a Cyanobacterium (*Synechocystis* PCC 6803) (Fig. 36C,D). While the *S. pneumoniae* pair was detectable reciprocally again, the *Synechocystis* interaction was only found in the combination of DBD-L14/AD-YbeB. In both species, however, the interaction seemed to be strong, detectable at up to 50 mM 3-AT again.

Finally, I wondered if these interactions are conserved in human and maize. Both, human and plant YbeB-homologs (C7orf30 and Iojap) contain a C-terminal DUF143 signature, hence their full-length ORFs were cloned as well as N-terminally truncated fragments excluding a predicted signal peptide.

Since human C7orf30 is predicted to have a mitochondrial signal peptide (Tab. 30) I chose to test the mitochondrial L14 homolog, L14_{mt}. I also tested a full-length clone and a construct (residues 20 to 145) that excludes a mitochondrial signal peptide (Tab. 30). Both ORFs are encoded by the nuclear genome (Venter et al., 2001). No clear Y2H signal was detectable between L14_{mt} and C7orf30 (Fig. 36E) although this interaction could be verified in a pull down experiment and by BiFC *in vivo* (see below). Notably, weak signals were detected among L14_{mt} constructs indicating a homomerization of L14_{mt}. Certain C7orf30 constructs appeared to dimerize as well (Fig. 36E).

The YbeB-homolog of *Zea mays* (Iojap) carries the DUF143 signature also near its C-terminus. The ORF was tested as full-length construct and as three N-terminally truncated constructs that exclude a plastidal localization sequence (Tab. 30) (residues 63-228 and 77-228) and a DUF143 core fragment (residues 114-228). Iojap is encoded by the *ij* gene in the nuclear genome (Han et al., 1992). Because Iojap was isolated from plastidal fractions (Han and Martienssen, 1995), I tested the corresponding plastidal L14-homolog RPL14 as full-length clone. RPL14 is encoded by the plastid genome and thus has no signal peptide (Markmann-Mulisch and Subramanian, 1988). However, no positives were detectable (data not shown) since the Iojap constructs self-activated still in presence of 200 mM 3-AT (that is, Iojap baits were able to activate transcription of the HIS3 reporter without preys present).

Cross-species interactions indicate a high L14-YbeB interaction co-conservation

The L14-YbeB interaction appears to be conserved in many if not most species. Because the structures of both L14 and YbeB are known, inter-species interactions may shed additional light on conserved residues or structural features that mediate these interactions (see below). Therefore, I also tested cross-species interactions using Y2H assays with full-length constructs. Indeed, cross-species positives were found in nearly all possible species combinations (13 out of 15 possible combinations) (Fig. 37). While the intra-species pairs of *Zea mays* and human failed to bind again, they surprisingly did exhibit interactions with several prokaryotic counterparts. For instance, the maize Iojap prey bound to all bacterial L14 baits. DBD-C7orf30 interacted with the prey of *T. pallidum* and *S. pneumoniae* L14. These results are remarkable because YbeB is not all that well-conserved on the sequence level but still seems to retain its binding specificity.

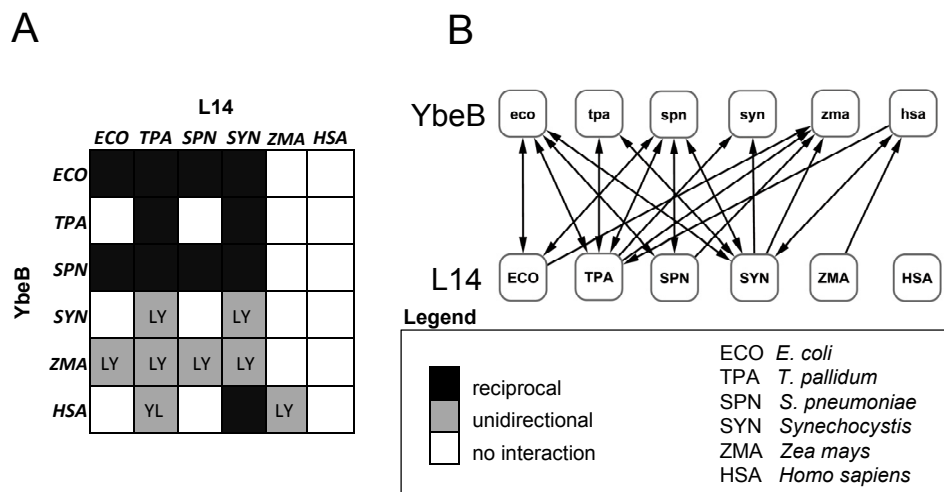


Fig. 37 Y2H cross-species interactions of YbeB and L14 orthologs

(A) Matrix representation. Various YbeB orthologs interact with L14 from other species and vice versa. For example, *E. coli* YbeB interacts with *T. pallidum* L14. Reciprocal means both proteins interact as prey and bait. Unidirectional interactions are positive only in one combination, indicated as bait-prey pair, e.g., LY = L14 as bait and YbeB as prey. (B) Alternative representation as mini-network. The direction of an arrow indicates the fusion protein interaction in bait → prey direction. Figure was drawn with Cytoscape (Shannon et al., 2003).

The YbeB-L14 interaction is conserved in man and plants

Human and maize proteins either did not interact in the Yeast Two-Hybrid tests or were not useful because of auto-activation, so I tested these interactions by pull down experiments (Fig. 38). Human and *Zea mays* proteins were expressed as NusA or MBP fusions. Furthermore, I verified the *E. coli* interaction as GST (YbeB) and MBP fusions (L14). While the *E. coli* pair worked well in co-expression (nearly equal protein levels could be achieved when co-expressed), the human and maize constructs were expressed separately and then the lysates were combined since the MBP fusions were expressed at much higher levels than the NusA constructs. In the pull down experiment I could confirm the *E. coli* interaction and the interactions among maize Iojap/RPL14 and human C7orf30/L14_{mt}.

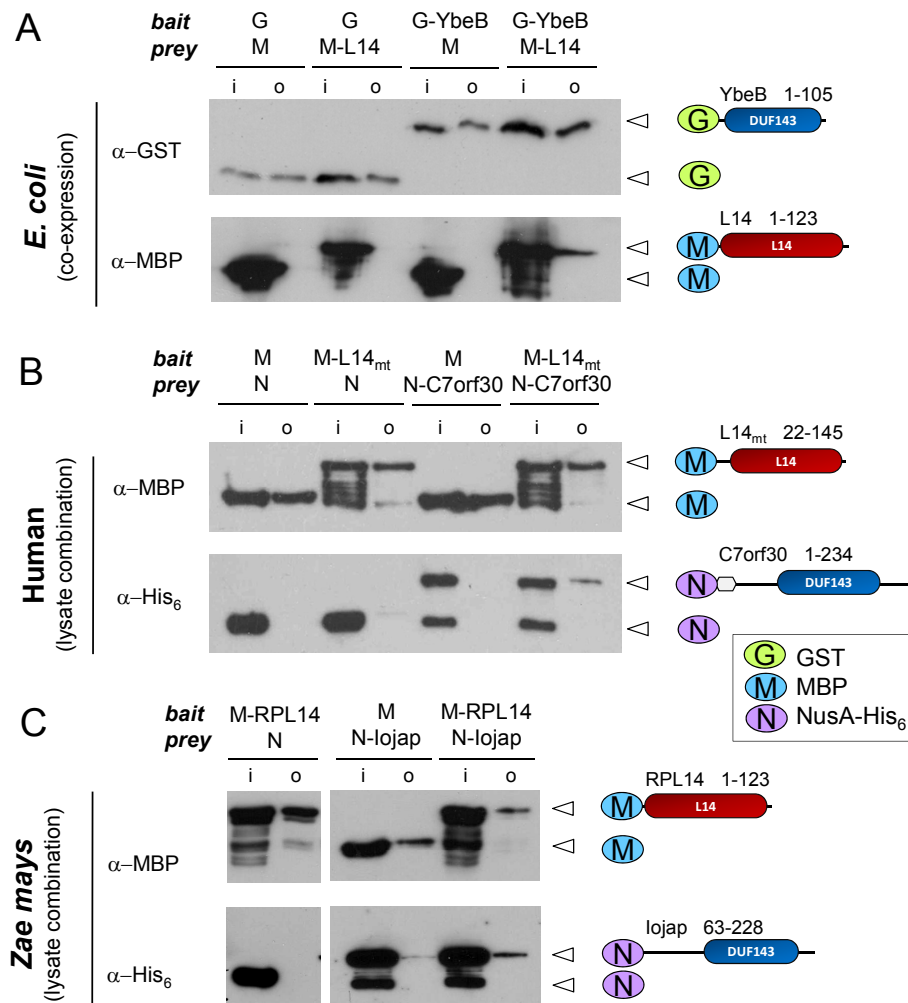


Fig. 38 *E. coli*, maize, and human YbeB-L14 pairs interact *in vitro*

(A) Pull down of *E. coli* YbeB and L14 as GST and MBP fusions, respectively. Constructs were co-expressed in *E. coli* BL21(DE3) and then GST-YbeB was used as bait and isolated by glutathione beads. Proteins were separated by SDS PAGE and detected with α-GST or α-MBP antibodies. (B, C) Pull down of human C7orf30/L14_{mt} and maize Iojap/RPL14 complexes. YbeB orthologs were expressed as NusA fusions and L14 orthologs as MBP fusions and protein lysates were combined. MBP fusions were used as baits and isolated with amylose beads. Samples were separated by SDS-PAGE and detected after Western blotting with α-MBP and α-His₆ antibodies. (i) input samples, (o) output samples. Constructs included the amino acid ranges shown on the right. Note that signal sequences were deleted in human L14 (B) and maize Iojap (C).

The human YbeB/L14 orthologs co-localize into mitochondria and interact in vivo

In order to investigate the *in vivo* relevance of the C7orf30/L14_{mt} interaction, I tried to detect both proteins in human mitochondria. C7orf30 and L14_{mt} were both cloned as full-length ORFs into pcDNA3.1-HA-mCherry (Diefenbacher et al., 2008) with native N-termini. C-terminally the ORFs were tagged by mCherry fluorescent protein. The plasmids were transfected into HeLa cells and then localization was observed by confocal microscopy (Fig. 39). As expected, L14_{mt} localized to the mitochondrial compartment as indicated by MitoTracker mitochondrial staining. C7orf30 also localized to mitochondria exclusively. The overlap of MitoTracker and mCherry signals show clearly that mitochondrial L14 specifically co-localizes with C7orf30 into mitochondria.

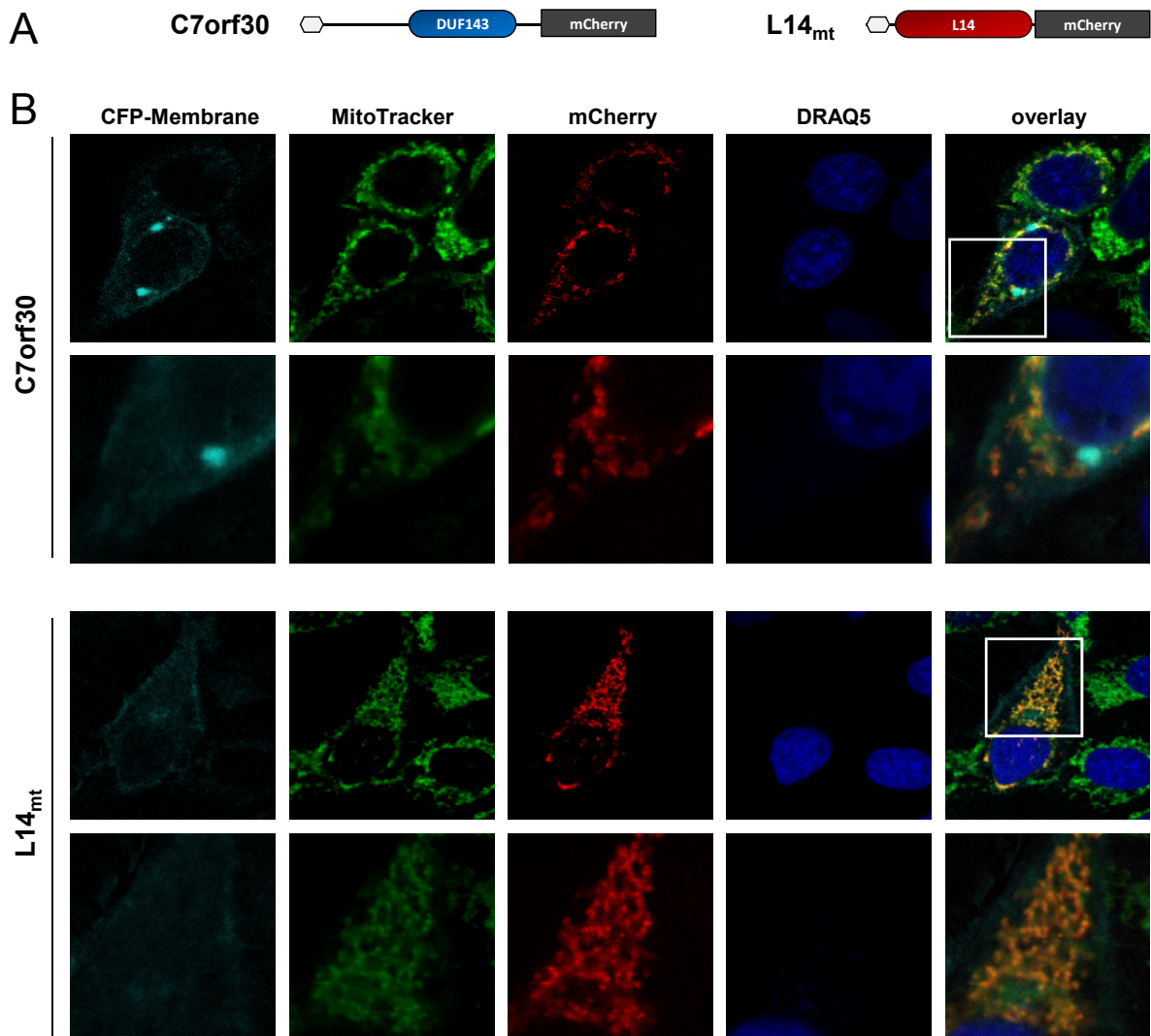


Fig. 39 Human C7orf30 and L14_{mt} co-localize *in vivo* into the mitochondrial compartment
(A) Reporter constructs used for protein localization. Rectangles indicate predicted signal peptides. **(B)** Localization experiment of C7orf30 and L14_{mt} in HeLa cells, respectively. (From left to right) Membranes stained by CFP-membrane (pECFP-Mem, Clontech), mitochondria with MitoTracker® Green FM (Invitrogen), protein localization by mCherry fusions, nuclear staining with DRAQ5, and overlay. Images were acquired by using a confocal laser scanning microscope (Zeiss LSM 510 Meta). Images are 630 x enlarged, blow-ups are a zoomed parts of the indicated cell. The assay was done in collaboration with Markus Diefenbacher, ITG, KIT, Karlsruhe.

RESULTS

Next, I wondered if C7orf30 and L14_{mt} are also capable to interact in mitochondria. Both full-length ORFs were cloned into the vectors pcDNA3.1(+)-HA-VN and pcDNA3.1(+)-HA-VC (Roder et al., 2010) with their native N-termini and C-terminal Venus fluorescing protein moieties (Fig. 40A). After co-transfection C7orf30 and L14_{mt} showed reciprocal complementation (Fig. 40B). Note that in the controls L14_{mt} also interacted with itself while C7orf30 showed no dimerization signal. This partly contradicts the Y2H data which indicated homomerization of both proteins. This difference may be explained by different conditions (yeast nuclei vs. human mitochondria) or by the use of different fusion constructs that may either permit or inhibit an interaction (N-terminal tags used in Y2H while C-terminal tags used in this assay).

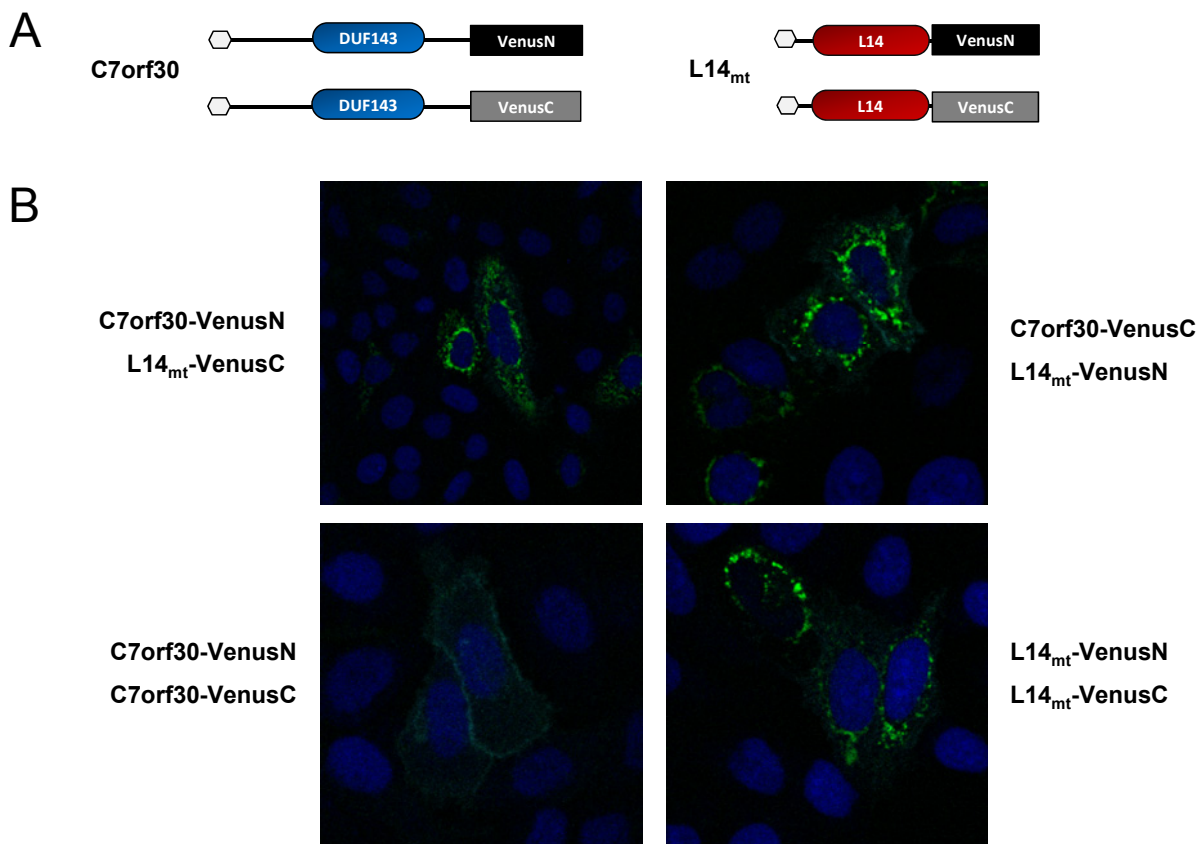


Fig. 40 C7orf30 and L14_{mt} interact *in vivo*

(A) Constructs used for bi-molecular fluorescence complementation (BiFC). Rectangles indicate predicted signal peptides. (B) BiFC in HeLa cells. Overlay images represent BiFC, DRAQ5, and CFP-membrane stained cells. The corresponding BiFC constructs that were combined (co-transfected) for the complementation assay are indicated in the figure. Images show cells at 630 x magnification. The assay was done in collaboration with Markus Diefenbacher, ITG, KIT, Karlsruhe.

Mapping the YbeB binding site on E. coli L14 – YbeB caps conserved and critical L14 amino acid residues

L14 is a nucleoprotein that is localized at the large ribosomal subunit on the of 50S-30S interface (Fig. 41A). L14 was shown to mediate bridges with the 16S rRNA of the 30S subunit during assembly of the functional ribosome. Only three other 50S proteins undergo (beside rRNA-rRNA contacts) such 30S contacts (Gao et al., 2003). Since the protein-interaction of YbeB with L14 turned out to be highly conserved, I wondered if also the amino acid residues that mediate the interaction are conserved as the logical consequence. First, I probed purified recombinant *E. coli* YbeB as GST- and MBP-fusion against an L14 peptide array with linear 15-mer peptides as described in (Titz et al., 2006a). However, I could not identify any interacting peptide (data not shown). Then I tried to identify critical residues by random mutagenesis of L14 and YbeB followed by a reverse two-hybrid assay. L14 and YbeB baits self-activated the URA3 reporter making it not possible for negative selection on 5-FOA containing medium (data not shown). Thus, I finally mutated L14 by site-directed mutagenesis. Therefore, I had to select a few less amino acid residue candidates. Based on the 50S 3D-structure (PDB id: 2AW4) (Schuwirth et al., 2005) I focused on amino acids that (i) were highly conserved (Fig. 35B) and (ii) that were located on the L14 surface that is exposed towards the 30S small subunit interface. These criteria identified two adequate regions: K114 and a stretch of four continuous residues (T97, R98, I100, F101) (Fig. 35B and Fig. 41A,B).

K114 was substituted by a single alanine codon while the T97-F101 stretch was replaced by four alanine codons. Then the constructs were tested again by a Y2H experiment (Fig. 41B). The interaction was lost in both cases, confirming that the interaction must involve the mutated conserved residues. Interestingly, T97 and R98 are supposed to be involved in bridge B8 between the 50S and 30S subunits (Gao et al., 2003) (Fig. 41A). Consequently, YbeB could prevent this interaction on the 3D-level. The highly conserved residue K114 is not known to mediate 16S rRNA contacts but its positively charged residue is exposed on the L14 surface (Fig. 41A). An explanation for the necessity for its conservation might be to specifically mediate and maintain the interaction with YbeB.

Both identified epitopes are non-linear but are located in close three-dimensional proximity (Fig. 41A, Fig. 35B). This could explain the unsuccessful peptide array experiment since here linear peptides are tested: the integrity of the L14 structure may be needed to bind YbeB and thus linear peptides may not be sufficient for an interaction. The L14 mutant with the stretch T97-F101 replacement could lead to deleterious effects onto the 3D-integrity of L14. However, when wild-type and both mutants (K114 and T97-F101) were expressed in *E. coli* as MBP fusions, no differences were visible regarding the solubility of the mutant versions compared to wild-type L14 (wild-type L14 is hardly soluble as GST but as MBP fusion under the applied buffer conditions, data not shown). This indicates that the T97-F101 substitution does not destroy the 3D-integrity.

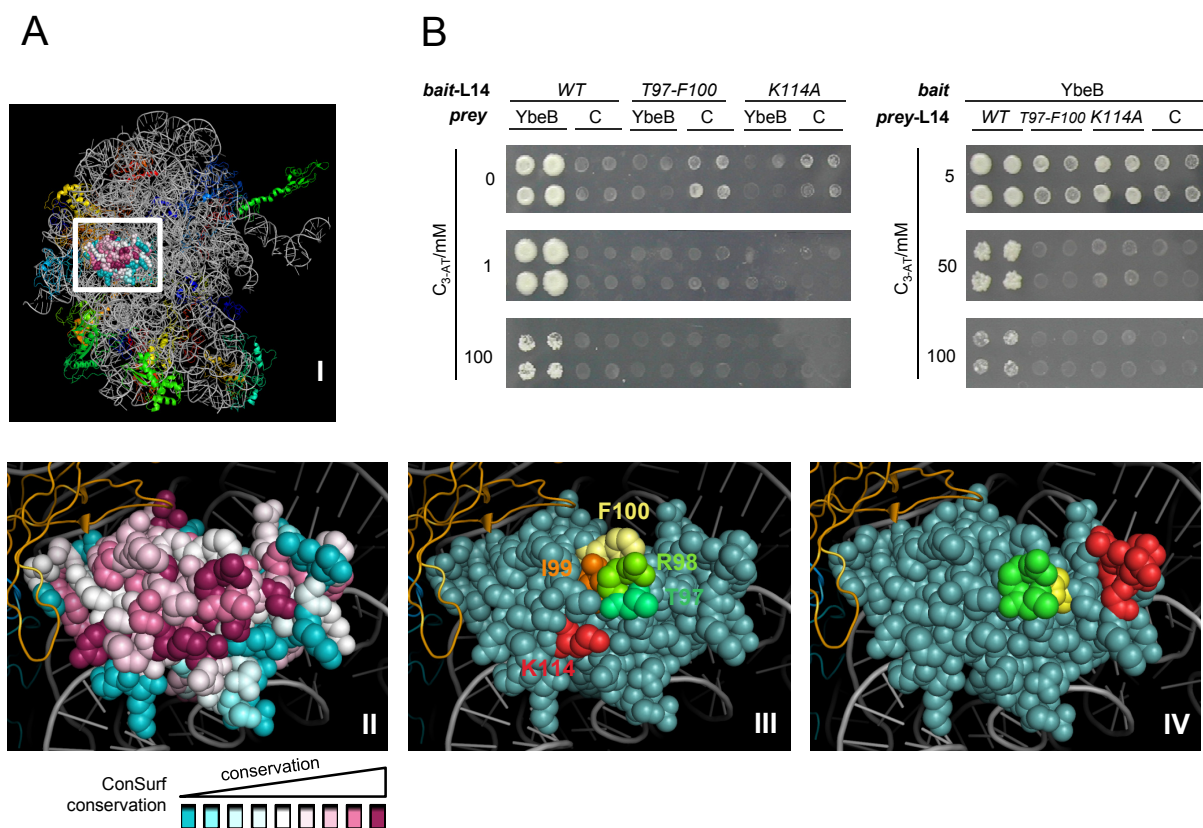


Fig. 41 Conserved amino acids mediate the interaction of L14 with YbeB

(A) L14 in the context of the 50S ribosomal subunit (PDB: 2AW4) (Schuwirth et al., 2005). (I) The large ribosomal subunit. rRNAs are in grey, ribosomal proteins in color. (II-IV) The exposed surface of L14. (II) Conservation of L14 amino acids are highlighted in magenta (highly conserved), grey (moderately conserved), and turquoise (little or no conservation). Conservation scores were determined by ConSurf Surfer (Landau et al., 2005). (III) Amino acid residues substituted by alanine. Mutated residues are highlighted: K114 and T97 to F100. (IV) Critical residues of L14 that contact 16S rRNA as determined by (Gao et al., 2003): L14 amino acid residues involved in bridge B5 are shown in red (E45 – G50) and B8 in yellow (N13) and green (T97, R98). Figures were drawn with PyMOL. (B) Mutated L14 does not interact with YbeB. L14wt: wild-type L14; T98-F100: L14 with T98 to F100 replaced by four alanine residues; K114A: L14 with K114 replaced by alanine. “C” control (self-activation test with empty prey vector). Y2H tests were done reciprocally in both fusion protein directions. The image shows quadruplicate testes. Diploid yeast was incubated for 7 d on readout medium.

3.2.3 YbeB functions as a negative modulator of translation *in vivo*

Gene knock-outs of translation-related proteins often cause growth defects under certain conditions or are lethal regarding the physiological significance of the translational process. In the existing literature there exists no experimental evidence that YbeB orthologs function indeed in translation. To find a YbeB-translation link also *in vivo*, I checked an *E. coli ybeB* knock-out strain by phenotyping and reporter assays.

Absence of YbeB causes a growth defect under competitive conditions

Given its near universal phylogenetic presence, YbeB was expected to be an essential gene. However, its gene knock-out in *E. coli* has no consequence on mutant growth under certain conditions (Baba et al., 2006; Jin et al., 2007). YbeB ORF is located at the first position in an operon including four more ORFs, *rmlH*, *mrdb*, *mrda*, and *rplA* (Mendoza-Vargas et al., 2009). To exclude indirect effects onto the downstream ORFs, I flipped-out the kanamycin cassette that substituted the *ybeB* ORF. This resulted in a truncated *ybeB* ORF encoding a truncated 34 amino acid peptide (Baba et al., 2006).

I tested the mutant towards general stress factors like high salt and high temperature in an agar plate assay but no growth defect was detectable compared to the wild-type (Suppl. Fig. 50). I did not check towards a cold sensitivity phenotype since this was already negatively tested for $\Delta ybeB$ (Jiang et al., 2007). Moreover, the deletion strain demonstrated growth similar – although consistently slightly impaired – to the wild-type strain under standard conditions (consider reference growth curves in Fig. 42B, Fig. 43A, and suppl. Fig. 51). Reduced growth under optimal conditions, cold, or heat sensitivities are typical for KO of accessory proteins of translation and can result in degenerated ribosome and polysome profiles, e.g., as demonstrated by (Jiang et al., 2007; Rasouly et al., 2009). In contrast, gene knock-outs of the ribosomal core components are lethal (Baba et al., 2006). A paralog that could complement YbeB function is not present in the *E. coli* genome (Blattner et al., 1997). I was surprised about the absence of such basic phenotypes in the *ybeB* KO strain, given its conservation and strong association with 50S/L14. In addition, I tested the mutant towards sensitivity against chemical compounds that interfere with translation elongation (spectinomycin, chloramphenicol, gentamycin, and tetracycline) and transcription (rifampicin) in a growth curve assay (Suppl. Fig. 51). However, no clear growth differences were detectable in the $\Delta ybeB$ strain.

In order to be more sensitive, I then measured growth of $\Delta ybeB$ in a competitive time-course assay (Fig. 42). When the KO strain was mixed with wild-type cells at equal cell numbers only 9% of the colonies turned out to be $\Delta ybeB$. This clearly shows that absence of YbeB causes a deleterious growth defect under competitive, more natural conditions, which is difficult to detect by a standard growth assay.

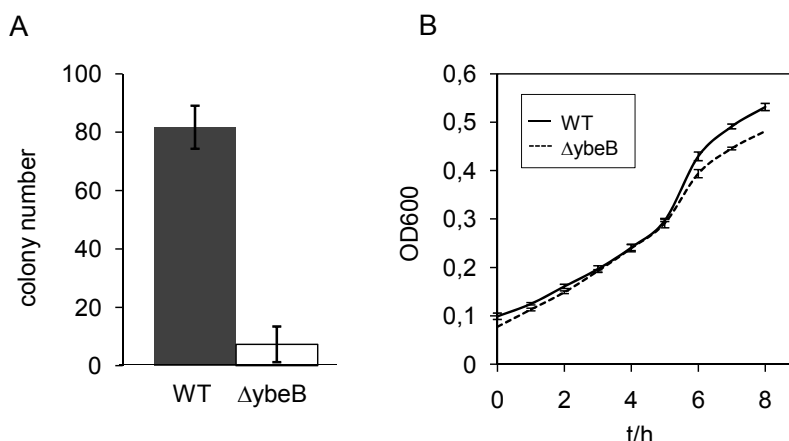


Fig. 42 The *E. coli ybeB* KO leads to reduced fitness

(A) Number of wild-type and mutant colonies after 72 h of competitive growth at room temperature (as determined by genotyping PCR of 3 x 93 randomly selected colonies, see (2.2.3.9) for details). (B) Growth of *ybeB* deletion mutant and wild-type cells are similar when grown separately in 1 x MOPS minimal medium containing 0.2 % glucose.

The loss of YbeB results in premature protein synthesis in vivo

Since no phenotype of *ybeB* KO strain was detectable that indicates a growth defect under non-competitive conditions, I used reporter gene constructs to test whether the absence of YbeB has an effect on translation *in vivo*. *E. coli lacZ* (encoding for β -galactosidase) was cloned under control of an arabinose inducible promoter into plasmid pBAD24 (Guzman et al., 1995). This enzymatic reporter was chosen because of its high sensitivity. Wild-type and $\Delta ybeB$ strains that harbor this construct were induced with low and high levels of inductor (0.002 and 0.2% w/v L-arabinose) in presence of 2 mM ONPG (2-Nitrophenyl- β -D-galactopyranoside). Then the ONPG turnover of growing cultures at 37°C was measured hourly.

Unexpectedly, the mutant exhibited an earlier ONPG turnover compared to WT in presence of 0.002 and 0.2 % inductor (Fig. 43A) suggesting YbeB having a deleterious effect *in vivo* on protein translation. Maximum ratios of $\Delta ybeB$ /WT occurred already in the early phase, 2 h after induction in the presence of 0.002% (ratio 8.17) and after 1 h with 0.2 % arabinose (ratio 7.35) indicating that proteins are produced earlier or faster in the mutant. The plateau and decline of the ONPG curves at later stages might be due to formation of inactivated reporter protein or inclusion bodies. In the control experiment without arabinose there was still but a strongly delayed signal for WT and $\Delta ybeB$ detectable with no striking difference between WT and mutant. *LacZ* expression was probably caused by leakiness of the arabinose-inducible promoter in combination with the high sensitivity of the reporter enzyme. In addition, a control mutant $\Delta hrpA$ (Baba et al., 2006) was assayed in parallel. HrpA is a mRNA helicase which was shown to have strong beneficial effect on translation *in vitro* (Kazuta et al., 2008). However, there was no deleterious effect obtainable at all *in vivo*. The ONPG turnover curve corresponded exactly to that of the wild-type indicating that there is no effect of $\Delta hrpA$ *in vivo*

on translation but underlining that the specific effect is attributed to *ΔybeB* in this assay (data not shown).

This phenotype can be rescued (Fig. 43B) as soon as YbeB is expressed from a plasmid. This leads to a delayed translational signal, confirming the findings that loss of YbeB causes a more rapid onset of protein translation (see below).

To verify these findings and to exclude the indirect effects of enzymatic outputs, the saved cells of these samples were analyzed by Western blotting to visualize β -galactosidase protein levels directly (Fig. 43C). Under non-induced conditions no reporter protein was detectable indicating presence of highly active but cryptic enzyme levels (data not shown). When induced with 0.002 % arabinose, *ΔybeB* already showed clear detectable reporter protein after 2 h while the wild-type did not; induction with 0.2 % arabinose resulted in detectable protein levels in both cases already after 1 h. The mutant strain again produced somewhat higher protein levels already after 1 and 2 h. For later time points protein levels were nearly equal. At early time points the β -galactosidase protein levels correlate with the enzymatic turnover indicating that absence of YbeB leads to production of premature protein levels.

To see whether the differences in resulting protein levels are caused by a translation effect or possibly might be due to different mRNA levels, I checked the presence of *lacZ* reporter mRNA by an RT-PCR (Fig. 43D). Therefore, the reporter strains were grown again under the same conditions and induced with 0.002% arabinose since here the clearest differences were found in the enzymatic readout as well in the β -Gal immunodetection experiment. In the non-induced samples little PCR product was detectable underlining the leakiness of the promoter. However, tests of later time points revealed that there is no difference of mRNA levels visible between WT and *ΔybeB*. These results underline that the production of premature protein levels in *ΔybeB* is specifically based on a translational effect and not caused indirectly on the transcriptional level.

Finally, I verified these observations using a GFP-based reporter assay (Albano et al., 1998) (Fig. 43E). In concordance with the *lacZ* reporter system, the *ybeB* deletion mutant showed a more rapid increase in protein (GFP) levels compared to the wild-type strain. The maximum signal ratio between *ΔybeB* and WT was again observed in the early phase (after 1 ½ hours): a ratio of 2.00 for induction with 0.002 % L-arabinose and a ratio of 1.29 for induction with 0.2 % L-arabinose. The strikingly higher GFP fluorescence intensity of *ΔybeB* at the early time point is clearly visible by fluorescence microscopy (Fig. 43F).

Taken together, by using these reporter assays I was able to demonstrate that *ybeB* deletion results in a more rapid onset of protein translation *in vivo* and thus strongly suggest a functional role of YbeB in protein translation as a negative modulator— possibly, in early translational processes such as 50S/30S ribosome complex assembly.

RESULTS

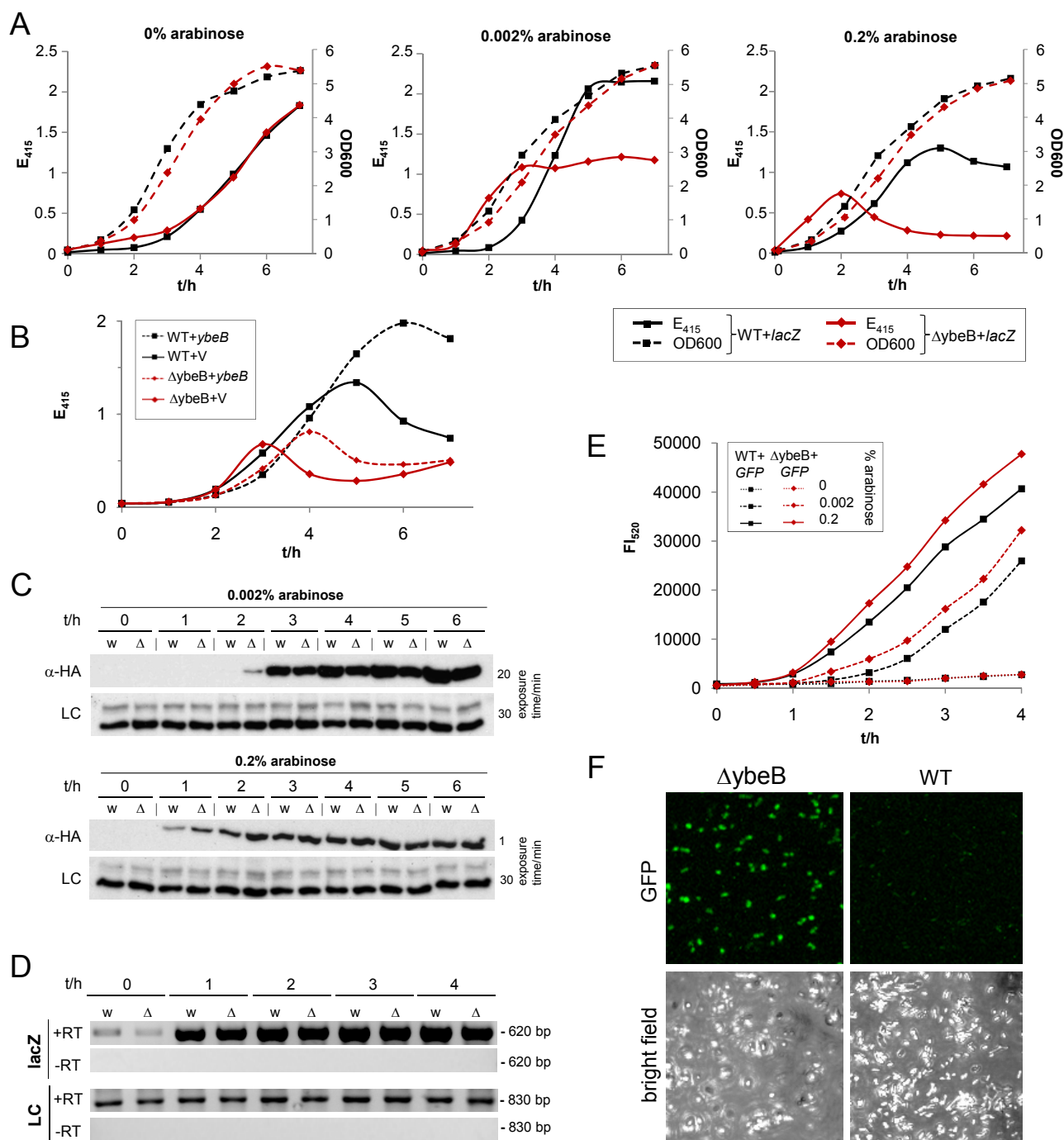


Fig. 43 The loss of YbeB results in a more rapid onset of protein translation

(A) A β -galactosidase reporter gene was induced with arabinose in the WT and the $\Delta ybeB$ gene deletion strain and its activity measured as ONPG turnover for varying inducer concentrations. Solid lines represents ONPG turnover, dashed lines OD600 reference growth curves determined in a separate experiment under identical conditions but without ONPG. The assay was repeated several times independently with the same result. In addition, a control gene deletion mutant (*AhrpA*) was assayed and showed ONPG turnover identical to the wild-type, further supporting that a YbeB specific effect is observed (data not shown). (B) Rescue assay: the curves show ONPG turnover assays equally done as in (A). *lacZ* reporter was induced with 0.2% L-arabinose. YbeB was constitutively expressed from pCA24N (Kitagawa et al., 2005). “V” represents the control measurements (strains transformed with pCA24N empty vector). Note, that YbeB-overexpression causes in the WT as well in the $\Delta ybeB$ strain a delayed signal. (C) β -galactosidase levels, measured by Western blotting using a rabbit α -HA antibody. The *lacZ* reporter ORF was 3'-terminally tagged with a hemagglutinine tag. LC: loading control. The stripped PVDF membrane was hybridized with polyclonal goat α -GST antibody that binds to an unknown *E. coli* protein with a MWG of approx.72 kDa, here used as loading control. (D) *lacZ* mRNA level as determined

by RT-PCR. Total RNA was extracted from samples that had been induced with 0.002% L-arabinose and were obtained under the same experimental conditions as in the reference growth curves from (A). 5 μ l PCR products were separated on a 2% agarose gel and stained with ethidium bromide. (-RT) Negative controls are shown (DNaseI treated samples w/o reverse transcriptase). As loading control (LC) *E. coli* alkaline phosphatase *phoA* mRNA was assayed. (E) Translation of a GFP reporter expressed from plasmid pBAD-GFP (Albano et al., 1998) results also in a premature protein level in *YbeB*. In this assay the GFP reporter protein production is measured by its GFP fluorescence in growing *E. coli* cultures. A maximal *YbeB*/WT GFP fluorescence signal ratio is observed 1 ½ hours post induction with 0.002% arabinose. This experiment was repeated several times independently with the same result. (F) This difference can also be directly observed by fluorescence microscopy: while *YbeB* shows already a bright GFP fluorescence after 1 ½ h post induction with 0.002% L-arabinose the wild-type GFP signal is still weak. Images show cells 400 x magnified.

3.2.4 *In silico*-protein docking of YbeB/L14 reveals that YbeB interferes with 50S-30S assembly on the 3D-level

Because the crystal structures of both YbeB and L14 are known (2.2.6), as is the fact that L14 is associated with the 50S subunit (Jiang et al., 2007), I docked YbeB to the area of L14 that is accessible in the context of the 50S particle. This was done to prove on the 3D-level which consequence this interaction could have on the ribosomal subunit assembly. A completely unconstrained docking of L14 and YbeB produced a list of putative conformations (“models”) sorted by the docking scoring potential (see Methods). Model #17 was the first one where YbeB had no backbone clashes with other parts of 50S.

Remarkably, that model also involved the representatives from the two sets of highly conserved L14 residues that have been shown as critical for the interaction based on the direct mutagenesis experiments (Fig. 41). Specifically, the interface contains residue K114 (shown to destroy the interaction when mutated alone) and residues T97, R98 (the mutagenesis has only been done for a group mutation 97-100 and has been shown to interrupt the interaction). T97 and R98 also belong to the bridge B8 with the 16S RNA.

As it can be seen from the Fig. 44, binding of YbeB to 50S at this position would overlap substantially with 16S RNA contacts of L14 in the assembled 70S unit. Thus, the predicted binding of YbeB would prevent the assembly of 70S from 50S and 30S, supporting the hypothesis that YbeB functions as a negative regulator of ribosomal translation.

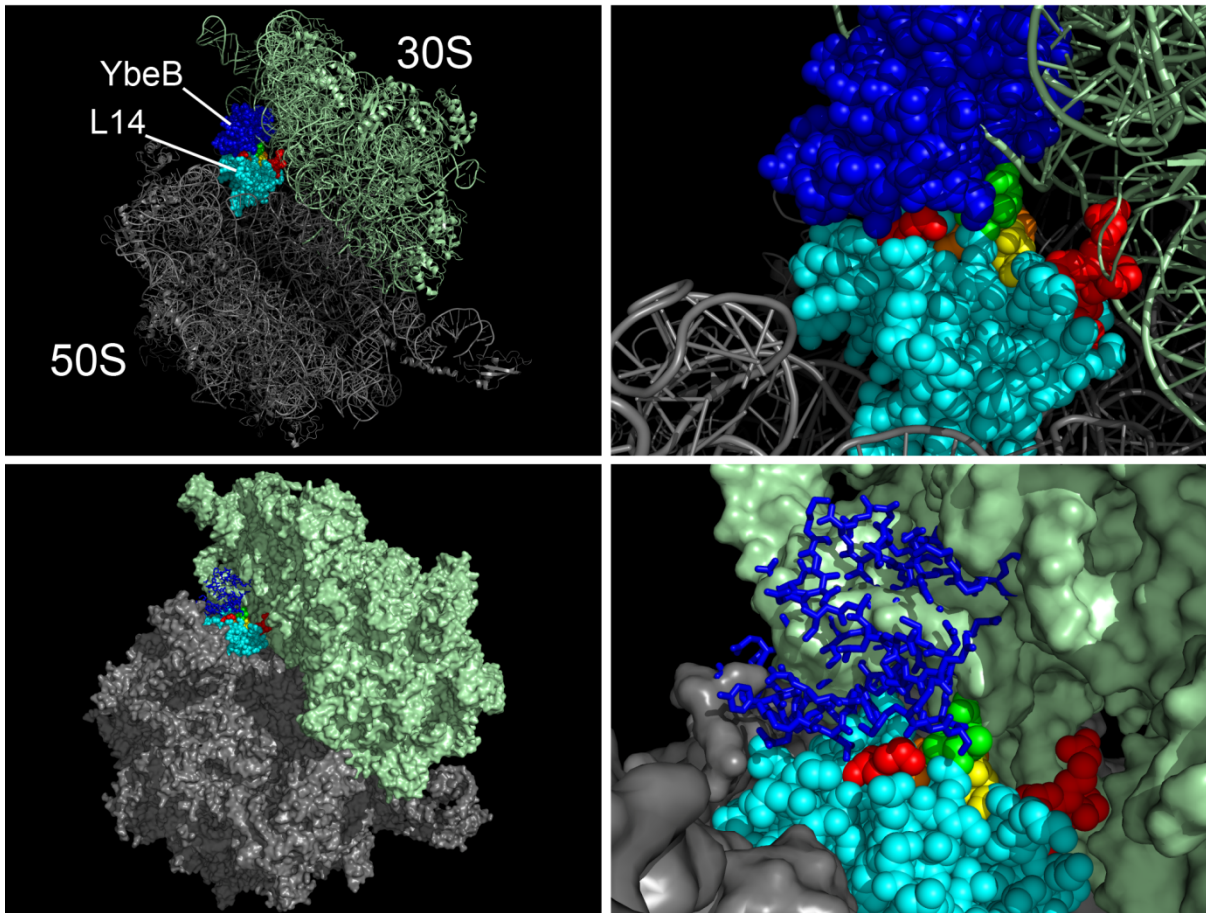


Fig. 44 YbeB interferes with 50S-30S assembly on the 3D-level

A docking model of YbeB against L14 on the 50S subunit of *E. coli*. Full views on 70S (left) and close-ups (right) in different presentations. Critical residues that mediate YbeB interaction on L14 or are involved in contacting 16S rRNA are stained according the color code from Fig. 41. When YbeB is docked onto L14 on the 50S subunit, the 30S subunit cannot bind to 50S anymore in the L14 region. This suggests that binding of YbeB is transient and may be regulated by an unknown mechanism. *In silico* docking was done in collaboration with Andrey Tovchigrechko, JCVI, Rockville, MD, USA. The coordinate files are attached to the e-supplement.

4 Discussion

4.1 Intra-viral PPIs of Dp-1 and Cp-1 and interactions with their host *S. pneumoniae*

4.1.1 PPI data quality – the problem with false-negatives and false-positives

Each PPI detection method is prone to false-negative and false-positive interactions (Braun et al., 2009; Venkatesan et al., 2009). Since no PPI dataset determined by high-throughput methods is ever perfect, it is of general interest to determine or estimate the number of PPIs that failed to be detected or that might be false. This can be done experimentally, by bioinformatical analysis, e.g., comparison of conserved interactions from various datasets (Titz et al., 2008), or by comparing literature curated PPI datasets (Rajagopala et al., 2009).

False-positives: I verified the phage-host PPI dataset by LuMPIS pull down assays and could reproduce 75.5% by using this alternative method. The remaining interactions were below the threshold indicating that $\sim \frac{1}{4}$ of the PPIs for the phage-host and intra-viral dataset could be false-positives, at least in this system. However, this assumption takes into account the non-reproducible PPIs only. As shown by several studies, it is not surprising that PPIs detected by a primary method pass the successful detection by an alternative method (Braun et al., 2009; Venkatesan et al., 2009). This might be due to the application of different fusion tags, expression systems, or the experimental situation (*in vivo* vs. *in vitro*). However, if a PPI is reproducible by different methods, this supports its validity and makes it attractive for further experimental investigation.

False-positives may be PPIs that take place *in vivo* but without having any physiological effect. Alternatively, certain bait or prey proteins may allow the yeast cells to grow without an underlying interaction. In my case, I cannot estimate the total number of false-positive PPIs detected in this study. Usually large PPI datasets can be filtered to exclude sticky proteins that exhibit an abnormally high PPI number. Using such a filtering strategy was not reasonable for my datasets since the phage genomes and thus the interaction spectrum is too small to define a clear threshold.

Recent attempts estimated the number of Y2H false-positives by using a negative gold-standard set of 92 literature-curated human PPI pairs (Braun et al., 2009). This negative set was recently tested also with the Y2H expression vector systems that have been used in this work (Chen et al., 2010). Chen and colleagues determined a false-positive rate of 6.5% for the pGBKT7g/GADT7g vector system and 5.4% for the pDEST32/22 system. If both vector systems are combined, an overall false-positive rate of 10.9% can be assumed for the intra-viral PPI network which is clearly lower than the estimated false-positive rate determined by the LuMPIS assays.

False-negatives: Determination of PPIs that are missed by Y2H or other assays is not a trivial task since these PPIs are “invisible”. False-negatives can be caused by applying recombinant proteins (steric hindrance caused by the tags) or by the absence of protein modifications (e.g., proteolysis for structural proteins, discussed below).

Attempts were also done to estimate this invisible PPI fraction by using a positive gold-standard set of 92 literature-curated human interaction pairs (Braun et al., 2009). It has also been tested with the expression vector systems that were used in this study (Chen et al., 2010). A total of 48.9% PPIs were not detectable by the pGBKT7g/pGADT7g system and 66.3% failed the detection once tested with the pDEST32/22 system. When both vector systems were used in parallel, the false-negative rate was lowered to 37%. It is unclear why these two expression vector systems lead to large non-overlapping PPI patterns. For instance, I detected in the intra-viral screens of Dp-1 and Cp-1 an overlapping fraction of 9.6% and 17.6%, respectively (Fig. 22A, Fig. 24A). As already mentioned, it is suggested that differences in the plasmid copy number or the fusion protein linker regions cause such differences (Rajagopala et al., 2009). However, 37% false-negatives can be estimated for the intra-viral PPI datasets since I screened the phage proteins with both vector systems under the same conditions. This confirms that the application of alternative systems can clearly help to enhance the screen sensitivity.

In the case of the phage host-PPIs this number of putative false-negatives is presumably higher since (i) only the pDEST32/22 system was applied, (ii) the *S. pneumoniae* array covered only 76.3% of its proteome, (iii) the pooling-strategy caused another number of intrinsic false-negatives of approx. 43%, and (iv) the fusions were screened only in one tag-direction (as phage-DBD → host-AD fusions). Only about ~1% of all interactions can be found when bait and prey fusions are inverted (Titz et al., 2008). Considering all these factors, I estimate the false-negative rate of the phage-host PPI dataset to be >90% with an absolute number of ~440 missing PPIs. However, this immense number might be too high considering the fact that approx. 20% of the host proteome had to interact with phage proteins (and this would result in an abnormal high in average node degree for phage proteins). In that case the false-negative rate might be presumably lower than calculated.

Finally, I summarized putative false-negatives that failed the detection but which I expect to bind *in vivo* based on homologous and analogous PPIs known from other phages (Tab. 33) (highlighted also in Fig. 45C and Fig. 46A). For instance, some expected PPIs were absent among structural Dp-1 and Cp-1 proteins. For instance, I observed no interaction involving the Dp-1 portal protein gp38. In many other phages, portal proteins undergo a homomeric interaction by building dodecameric rings (Doan and Dokland, 2007; Driedonks et al., 1981; Kochan et al., 1984). In coliphage T4 the portal protein was found interacting with the major capsid protein and the large terminase subunit (Fokine et al., 2004; Malys et al., 2002). In *Bacillus* phage SPP1 the portal protein connects the capsid with the tail (Lhuillier et al., 2009). A possible reason for the absence of detected interaction is that some phage proteins undergo post-translational cleavage as demonstrated for T4 major and minor capsid proteins and Cp-1 major head protein (Laemmli, 1970; Martin et al., 1998b). Such cleavage processes occur in

DISCUSSION

a sequential manner during the stepwise coordination of the virion assembly *in vivo* but cannot be simulated in pairwise interaction tests.

On the other hand, I expected the Dp-1 sigma subunits to interact with the host RNA polymerase and the Dp-1 DNA polymerase III system to bind to host DNA polymerase III core subunits since these components are not encoded by the Dp-1 genome. Although such PPIs were not identified, it might be necessary to recruit these factors to successfully drive Dp-1 transcription and DNA replication.

Tab. 33 Putative false-negatives that failed detection in the screens

The table lists published PPIs from other phages (PPI is expected to occur in Dp-1 or Cp-1), direct references for Cp-1 and Dp-1, and a logical interpretation of the dataset. Methods were included that do not demonstrate direct physical binding (e.g., electron microscopy (EM)) but suggest that a PPI must occur between proteins.

PPI	Comment	Method	Reference
Cp-1 terminal protein (TP) with Cp-1 DNA polymerase	DNA pol. makes use of TP-dependent protein priming mechanism while binding to TP. Demonstrated with ϕ 29 homologs.	Co-purification	(Martin et al., 1996a)
Cp-1 MHP protease with MHP	No direct physical PPI was demonstrated. Leads to cleavage of MHP precursor.	Co-expression	(Martin et al., 1998b)
Cp-1 and Dp-1 MHP (homomerization)	Information considered from T7 and ϕ 29	Yeast Two-Hybrid, EM	(Bartel et al., 1996; Tao et al., 1998)
Cp-1 lower collar protein with antireceptor	Expected (discussed in the reference); ϕ 29	-	(Martin et al., 1996b; Tao et al., 1998)
Cp-1 connector with MHP	Information considered from ϕ 29	EM	(Tao et al., 1998)
Cp-1 lower collar protein with connector	Information considered from ϕ 29	EM	(Tao et al., 1998)
Cp-1 terminase (encapsidation ATPase) with connector	Suggested. No direct binding was demonstrated in ϕ 29.	DNA packaging assays	(Grimes and Anderson, 1997; Guo et al., 1986)
Dp-1 large terminase subunit (homomerization)	Information considered from T4	Peptide display	(Malys et al., 2002)
Dp-1 portal protein (homomerization)	Information considered from SPP1	EM	(Dube et al., 1993)
Dp-1 tape measure protein with major tail protein	Information considered from SPP1	EM	(Plisson et al., 2007)
Dp-1 portal protein with major head protein	Information considered from T4	EM	(Fokine et al., 2004)
Dp-1 portal protein with large terminase subunit	Information considered from T4	Peptide display	(Malys et al., 2002)
A Dp-1 sigma subunit with a host RNA polymerase core subunit	Is expected since Dp-1 does not encode for an own RNA polymerase but for sigma factors.	Interpretation	-
A Dp-1 DNA pol. III subunit with the host DNA pol. III core enzyme.	Is expected since Dp-1 does not encode for own DNA polymerase III core components.	Interpretation	-

4.1.2 Virion models: functional protein linkage maps of virion associated proteins of Dp-1 and Cp-1 implicate expected and unexpected proteins to function in virion assembly - a comparative view

Partial and whole virion 3D-structures have been solved successfully and provide detailed information about the composition and arrangements of structural proteins of phages such as T4, SSP1, ϕ 29, N4, and P22 (Chang et al., 2006; Choi et al., 2008; Dube et al., 1993; Fokine et al., 2004; Leiman et al., 2003; Plisson et al., 2007; Tao et al., 1998; Xiang et al., 2006). These studies were mainly done by using cryo-electron microscopy in combination with protein crystallography under static conditions. Since virion assembly is a highly dynamic process and no structural information is available for the Cp-1 and Dp-1 particles, I used a data integration strategy to develop virion models for these two phages: these models have the advantage that the applied Y2H system is adequate to identify also transient PPIs that could play a role during virion morphogenesis and are normally not detectable by other methods. I used the identified Y2H PPIs, homology-based protein annotation, information from literature, and mass spectrometry results for Dp-1 proteins to propose functional virion linkage maps that highlight which expected and unexpected proteins could play a role in this process. The models are especially helpful to get novel hints about the functions of unknown proteins.

A Dp-1 virion model

Structural proteins and putative maturation factors: In order to develop a model of the Dp-1 virion, I integrated genome annotation, binary interactions, and LC-MS/MS data of whole Dp-1 particles (unpublished LC-MS/MS data was kindly provided by Mourad Sabri, Université Laval, Québec) (Fig. 45). LC-MS/MS identified 8 proteins in purified Dp-1 particles, primarily known structural proteins and one hypothetical protein (gp42) (red proteins in Fig. 45). Another three proteins (gp40, gp41, and gp53) could be linked by functional annotation as well as by binary protein-protein interactions and might function as morphogenetic factors such as chaperones dissociating from the mature virion since they were not detected by LC-MS/MS. For example, the major capsid protein gp43 interacts with gp41 which is a homolog of the minor structural protein gp20 of *Staphylococcus* phage ϕ ETA. Gp41 could be the Dp-1 capsid scaffolding protein since it interacts specifically with gp43, but does not appear to be a major structural component as determined by LC-MS/MS. Finally, 10 hypothetical gene-products from the structural gene clusters including such whose annotation do not clearly link them to the virion process could be linked to the virion by Y2H interactions (blue proteins in Fig. 45AB). Also, these candidates were not detected by LC-MS/MS indicating that their role in virion assembly is presumably transient. Several known or predicted structural proteins were not detected by LC-MS/MS but found to interact in Y2H screens as homomers such as gp56. Another hypothetical protein, gp33, is encoded by the reverse transcribed module and links head and tail proteins. Finally, gp42, gp45, and gp49 were detected by LC-MS/MS but binary PPIs were not found that link them directly with other structural

components. As a consequence, by combining mass spectrometry data with Y2H data, I could physically associate 21 proteins that are encoded by the morphological/structural module including several hypothetical proteins.

Unexpected interactions with other virion components: Seven out of eight proteins identified by LC-MS/MS were expected structural proteins since they are homologous to known virion proteins. In addition I found 12 non-structural related gene products that interact with the structural core (Fig. 45D), e.g., unexpected PPIs between enzymes involved in queuosine metabolism (QueD, QueE, QueF, QueT) with the proteins that are presumably involved in virion assembly. One possible explanation is that these interactions may function *in vivo* as feedback signals for protein translation. Proteins such as gp44, gp33, or gp55 could block the Que enzymes when a critical level of structural phage proteins is reached, thus saving energy spent on translation and queuosine biosynthesis. Ten additional interactions were also detected that involved the putative RNA polymerase sigma factor gp69.

As mentioned already (3.1.3), there is an additional active cross-talk among structural components and replication proteins. For example, the DNA polymerase I (gp71), DNA polymerase III subunit β (gp10), DNA ligase (gp16), and DNA primase (gp68) interacted with proteins likely involved in DNA packaging (Fig. 45D). The transfer of DNA into the pro-capsid must be coordinated with replication and the observed interactions associate terminase assembly, DNA packaging, and other processes. Holliday structure intermediates that occur during T4 DNA replication have a negative effect on DNA packaging but can be resolved by a T4-encoded endonuclease that binds the portal protein (Kemper and Brown, 1976; Luftig et al., 1971). The fact that DNA replication/repair enzymes like T4 DNA ligase are needed *in vivo* for efficient packaging suggests a close local association of DNA packaging and DNA replication/repair processes (Zachary and Black, 1981). In Dp-1 the interaction of gp23 with gp16 (DNA ligase) indicates that the ligase could close DNA nicks during packaging. Recruitment of Dp-1 gp71 (DNA polymerase I) via gp22/gp23 interaction could also initiate DNA repair or primase primer displacement in parallel to packaging. Interestingly, gp23 of Dp-1 interacts in addition with the DNA polymerase III β -clamp subunit (gp10), while in T4 the large terminase subunit was shown to interact with the T4 β -clamp subunit (Malys et al., 2002). This suggests a role of the β -clamp/DNA polymerase III in DNA packaging while distant phages developed analogous interactions to target the β -clamp.

DNA packaging and terminase: In many tailed phages which use a headful packaging mechanism, efficient DNA encapsidation requires an active portal protein embedded in the procapsid portal vertex as well as an interacting terminase complex (Chai et al., 1994; Droge and Tavares, 2000; Isidro et al., 2004a; Isidro et al., 2004b). Although I could not detect any interaction of the Dp-1 terminase (gp37) with the portal protein, I discovered several interactions (Fig. 45B) that link terminase-associated proteins to capsid components: gp39, gp23, and gp33 bind to gp43 (major capsid protein), gp22 to gp41 (minor capsid component), and gp33 to gp44 (predicted lipoprotein). These interactions could

mediate the terminase assembly in a cooperative manner with the portal vertex. The terminase associated gp23 contains a partial low level similarity to large terminase subunits from other phages and seems to act as a hub by binding virion associated proteins (gp49, gp22, gp48, gp39). Gp23 thus could be important for recruiting additional protein factors to the terminase complex that stimulate the packaging process. It is interesting that the gp23 sequence does not contain a HTH motif which is a key feature for the small terminase subunit in phage λ and important for cooperative DNA-binding and DNA packaging (de Beer et al., 2002). Thus, gp23 cannot be viewed as a canonical small terminase subunit, but it could rather function as an adaptor protein. However, the hypothetical protein gp39 contains a DNA-binding motif (DNA polymerase alpha zinc finger, (Pfam PF08996)) and associates via gp36 with the large terminase subunit. Thus Gp39 could be a key component of the terminase complex by binding the DNA substrate. Notably, in genomes of other long tailed pneumococcal phages like EJ-1 and MM1 a canonical small terminase subunit also appears to be missing (Obregon et al., 2003; Romero et al., 2004). All these phages may thus use untypical packaging machines.

How do the lipids become an integral part of the head? It was previously reported that mature Dp-1 particles contain lipids that originate from the host cell (Lopez et al., 1977). This rare combination has not been reported so far for any other phages of the Caudovirales order (Ackermann, 2006) but membranes were found in *Pseudoalteromonas* phage PM2 (Corticoviridae family), *Pseudomonas* phage ϕ 6 (Cystoviridae), and for enterobacterial phages of the PRD1 group (Tectiviridae) (Bamford et al., 1981; Franklin et al., 1969; Vidaver et al., 1973). In the non-tailed dsRNA phage ϕ 6 the lipids surround the capsid and contain additional envelope proteins and lipoproteins (Bamford et al., 1976; Sinclair et al., 1975). Members of Corticoviridae contain a lipid vesicle between the outer and inner layer of their capsid, while Tectiviridae contain a lipid bilayer inside their capsid. The exact location of the lipids in Dp-1 is unknown (Bamford et al., 1990; Lopez et al., 1977). However, Lopez and colleagues supposed the lipids to be arranged as a double-layer membrane that surrounds the capsid.

As described above, the hypothetical protein gp44 of Dp-1 carries an N-terminal signal peptide and is predicted to be a lipoprotein by its homology to a lipoprotein domain (PROSITE PS51257). I found it to interact with the major capsid protein gp43. It is tempting to speculate that gp44 is inserted via its N-terminal signal peptide into the host cell membrane during intracellular phage replication. Its interaction with the major capsid protein may indicate that the pro-head assembly is localized to the membrane. Interestingly, phage T4 pro-heads assembly is also localized at the inner side of the cytoplasmic membrane, at which a membrane-spanning initiator complex forms consisting of T4 portal protein and T4 gp40. After the release of the pro-head into the cytoplasm, gp40 is not interacting with assembled capsids anymore (Hsiao and Black, 1978). An analogous situation might be present in Dp-1. The lipoprotein gp44 was not found by LC-MS/MS, and a membrane-dependent pro-capsid assembly could be initiated by the gp43/gp44 interaction rather than by the portal protein. Gp44 could be the key component and could therefore explain how the host membrane becomes an integral

part of the mature virion. It is possible that the pro-head assembly at the membrane site might result in a membrane bulb structure due to membrane ex- or invagination. This may be observable by electron microscopy. It is noteworthy that gp44 additionally binds to the host protein SP1575 (Fig. 45E), an essential conserved hypothetical protein that contains a replication initiation and membrane attachment protein (DnaB) signature (Pfam PF07261). DnaB from *B. subtilis* is essential for both replication initiation and membrane attachment of the *ori* region of the chromosome and plasmid pUB110 (Hoshino et al., 1987). Thus, SP1575 might be an additional membrane-associated factor but which is provided by the host and might be involved in building a complex together with gp44/gp43/gp41 supporting the membrane-associated capsid assembly.

It has been suggested that such a membrane uncommon in siphophages provides resistance for Dp-1 virions against H₂O₂ which is produced by the host (Duane et al., 1993).

Tail: LC-MS/MS analysis confirmed gp52 (tail length tape measure protein), gp54 (antireceptor), and gp55 (minor structural protein) as structural proteins. The latter one carries a Pfam HMM signature (DUF859) which is found in 44 protein sequences, mainly in *Streptococcus* phage proteins. DUF859 proteins are annotated as “structural” or “minor tail proteins”. Protein gp55 is probably the major component of the Dp-1 tail and its homomeric interaction confirms its ability to polymerize. In addition, gp55 interacts with the receptor-binding protein gp54 (Fig. 45B). No binary contact was found with the tail length tape measure protein that acts as ruler of the tail length (Katsura and Hendrix, 1984), possibly because it requires cooperative binding to assembled gp55 and/or other chaperone proteins.

Structural components that interact with host proteins: As demonstrated in Fig. 45D many Dp-1 proteins involved in DNA replication/recombination bind to structural components and imply that virion morphogenesis/DNA packaging and DNA replication are physically co-located or co-regulated. It is noteworthy that also some host proteins involved in various DNA repair pathways bind to the proposed morphogenesis factors (Fig. 45E). A putative role of gp44/DnaB interaction was already suggested above. The terminase-associated protein gp39 binds to the host RuvB Holliday junction helicase and could function in recognition of Holliday intermediates on the synthesized Dp-1 DNA, thus leading to its disaggregation and efficient DNA packaging. On the other hand, it might be possible that these structural components have an extra function beyond morphogenesis and could affect the DNA repair pathways to block DNA repair of the host. Note, that coliphage T4 packaging operates quite independent of host components thus suggesting terminase phage protein interactions may predominate (Rao and Black, 2005). In contrast, efficient λ DNA packaging was shown to be dependent on DNA-binding host factors IHF and HU (Catalano et al., 1995; Mendelson et al., 1991).

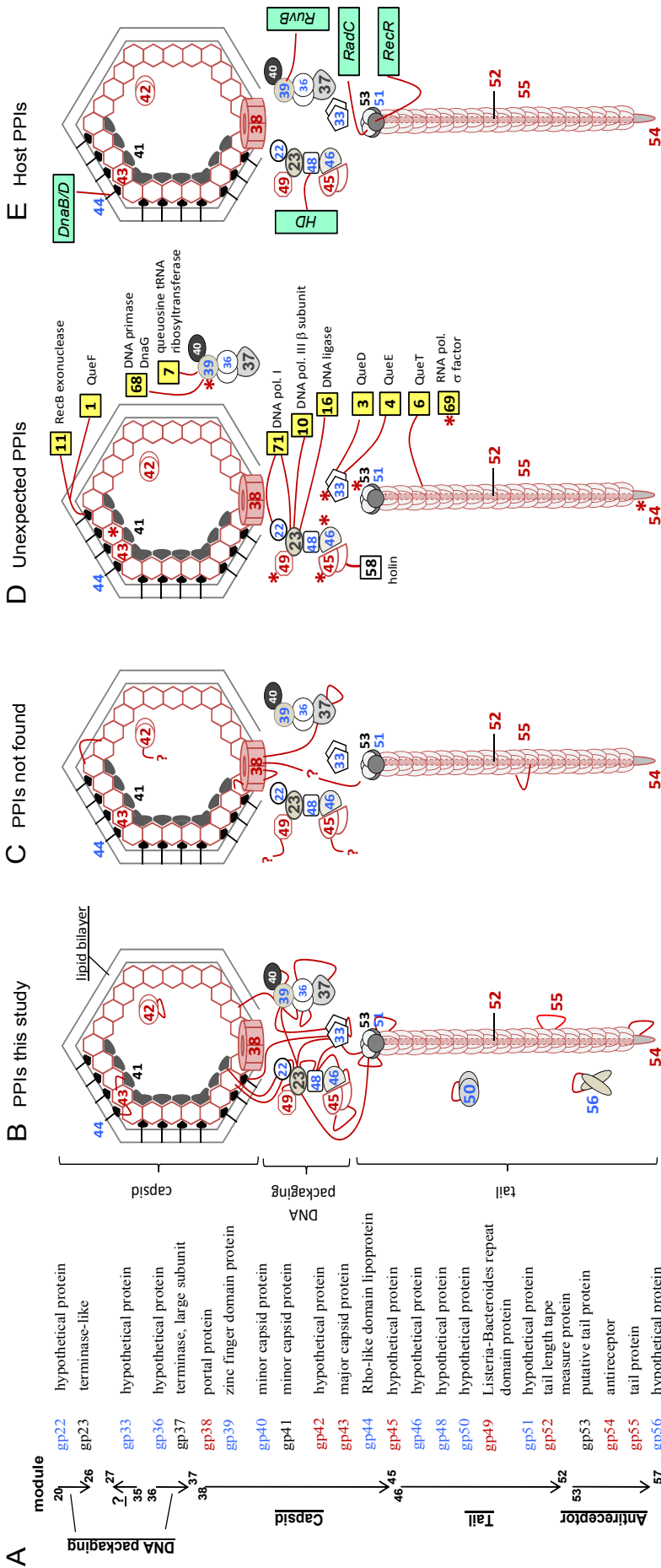


Fig. 45 A Dp-1 virion model based on homology predictions, binary protein-protein interactions, and virion LC-MS/MS data

(A) A list of all proteins that were considered for the model given in (B). Only proteins were included that (i) are predicted to function in morphogenesis/virion structure by homology annotation, (ii) that were confirmed by LC-MS/MS (in red), (iii) proteins encoded by the structural/morphogenesis gene cluster that bind with proteins described in (i) and (ii) with an interaction distance of 1 (direct interaction) plus homomeric interactions, (iv) hypothetical protein gp33 since it binds predominantly with head, tail, and terminase components. Proteins of unknown or little known function are highlighted in blue. (B) Core model of Dp-1 virion (schematic representation). Interacting proteins identified by Y2H screens are connected by red lines. Numbers reflect the systematic Dp-1 protein IDs: Proteins identified by LC-MS/MS (red), hypotheticals and of little known function (blue), proteins with exact known function based on homology predictions (font in black or grey). Protein symbols that touch each other but are not connected via red lines were not found to interact in the Y2H screens but were found previously in other phages (see main text and Tab. 33). Note, that these interactions do not need to take place in Dp-1 and are thus merely hypotheses. Compare to C. (C) Interactions of structural Dp-1 proteins not found in this study, but based on homologous and analogous interaction information from other bacteriophages. Putative false-negative interactions are symbolized by red lines, also of these proteins that show only a homomeric interaction or that were confirmed by LC-MS/MS but for which I couldn't find a direct interaction link with other structural components (indicated “?”). Refer also to Tab. 33. (D) Unexpected binary interactions with functionally annotated proteins from other gene expression clusters. Unexpected binding partners are encircled by yellow squares. Interacting proteins are connected by red lines. Interaction partners of RNA polymerase sigma factor gp69 are marked by a red asterisks. (E) Dp-1 PPIs with host proteins that function in DNA replication/repair. Host proteins are indicated as green squares and are connected by red lines with the interacting phage proteins. (HD) SPI746 is a conserved hypothetical protein containing a HD phosphohydrolase signature.

A Cp-1 virion model

As for Dp-1, I created for Cp-1 a virion linkage map to highlight which structural proteins and morphogenesis factors might participate in virion assembly. As already shown in the results section (3.1.3), most of detected intra-viral PPIs of Cp-1 were found among structural proteins underlining the stringent association of the structural proteins on the genomic and interaction level.

Nine different structural Cp-1 proteins have been originally identified by SDS-PAGE analysis (Ronda et al., 1981). Since a genome sequence was not available at the time the exact structural gene products were unknown. The Cp-1 structural gene module encodes for 13 proteins (Martin et al., 1996b). However, not all of these proteins are static virion features but might be morphogenesis factors which need to be absent after the maturation process from the mature virion. I used this information to model a Cp-1 virion protein-linkage map and integrated the detected binary PPIs. Since no further experimental information is available about the Cp-1 particle structure, I used the 3D-virion information from *B. subtilis* podophage ϕ 29 (Xiang et al., 2006) which is related to Cp-1.

Notably, all proteins encoded by the 13 ORFs of the structural module could be included in the model (Fig. 46A). They highlight which PPIs take place in the virion and provide a model for the poorly understood virion particle. Three of these proteins are known to play a role during morphogenesis and function in pro-capsid assembly (gp8), post-translational cleavage of the major head protein (gp13), and DNA-packaging (gp20) (Martin et al., 1996b; Martin et al., 1998b).

Most of the PPIs were expected. For instance, the major head protein (gp9, MHP) binds with the scaffolding protein (gp8) and the latter with connector protein (gp10). Also PPIs among tail and collar components were expected (Fig. 46B). Some expected interactions were not detectable in the Y2H screens and probably represent false-negatives. These are the homomerization PPI of the the MHP, the missing link of the terminase to the connector, a connector-base plate interaction, an interaction of the tail proteins gp17 or gp19 with the antireceptor gp18, and the MHP endoprotease (gp13) with the MHP (Fig. 46B). In the latter case it was shown that gp13 is a maturation factor which is responsible for cleaving an N-terminal peptide from MHP (Martin et al., 1998b). Failure of the detection of this PPI could be due to protease-substrate interactions that are not examinable by Y2H at all. This could also explain the missing self-interaction of the MHP because it is processed and only full-length ORF constructs were tested in the screen experiments. However, I detected a PPI of gp13 with the scaffolding protein (gp8). This interaction could indicate that cleavage of the MHP by gp13 might take place during the pro-head assembly and that gp13 uses the scaffolding protein as an adaptor to coordinate the cleavage of the MHP in a cooperative manner.

Hypothetical protein gp14 shows no homology to any other protein sequences in various databases. I propose it as head fiber protein since it interacts exclusively with the MHP.

The tail proteins N and C (gp17 and 19) were shown to have different sequence homology regions with the tail protein of ϕ 29 phage. While gp17 is homologous to the N-terminal part of the ϕ 29 tail

protein, gp19 aligns with the C-terminal part (Martin et al., 1996b). Interestingly, in Cp-1 the gp17 and gp19 tail proteins interact while in ϕ 29 the homologs were fused into one polypeptide chain during evolution. This phenomenon is also known as the Rosetta Stone principle (Marcotte et al., 1999) and underlines that in ϕ 29 the strong physical association of these two proteins lead presumably to the ORF fusion.

Finally, I highlighted unexpected interactions among structural proteins with host proteins and virion unrelated Cp-1 proteins in Fig. 46C. For instance, a single PPI was identified between the scaffolding protein gp8 and hypothetical protein gp6 that also binds the co-expressed DNA polymerase (gp5). In addition, some structural proteins undergo PPIs with host proteins whose annotations do not imply them directly to be relevant for Cp-1 morphogenesis. It is more likely that the interacting phage proteins evolved dual functions, a structural one and e.g., a regulatory one, while additionally affecting host proteins. In the case of the connector (gp10) that interacts with the essential glutamate-racemase MurI this interaction might block MurI and thus cell wall biosynthesis (this PPI is discussed below).

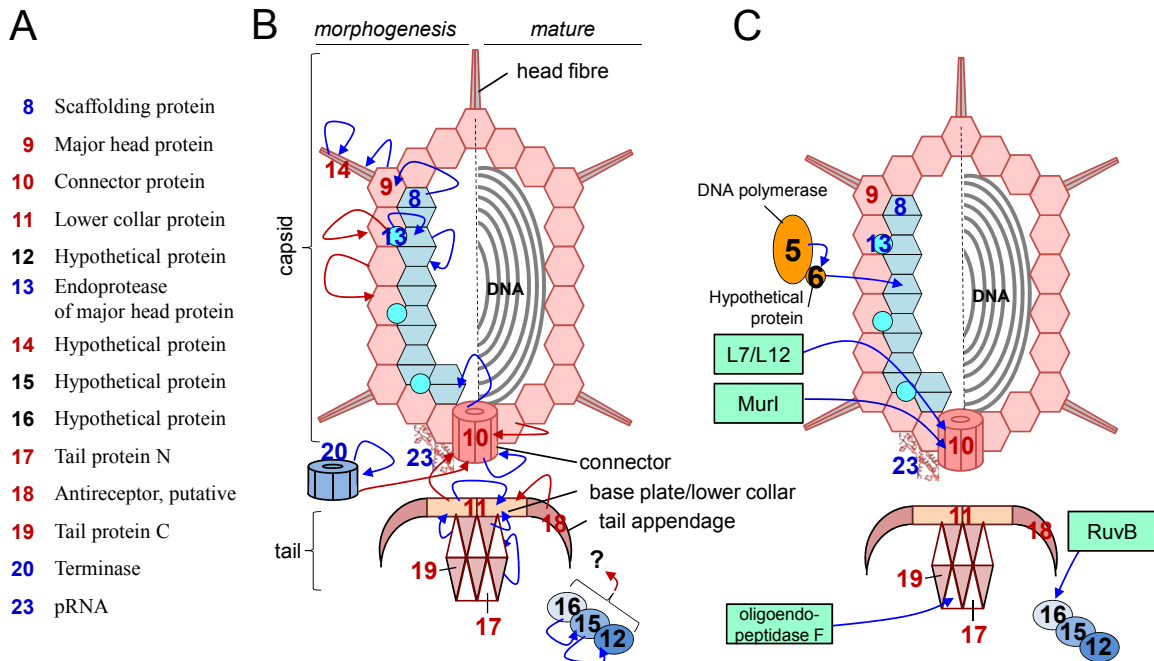


Fig. 46 A Cp-1 virion PPI linkage map

Schematic representation of the Cp-1 virion. Cp-1 EM virion structure information (Fig. 5) and 3D-information of ϕ 29 was combined in this model (Tao et al., 1998). **(A)** A list of all proteins for which binary PPIs were detected in this study, that were included in the model, and that are encoded by the structural gene module (gp8 to gp20). Orf23 encodes for a putative RNA involved in packaging (pRNA) (Martin et al., 1996b). Structural components are labeled in red font, morphogenesis factors in blue font, and unknown components in black font (valid also for B and C). **(B)** Interactions among structural components and maturation factors projected onto a schematic Cp-1 particle. Identified interactions are indicated by blue edges (arrows) that connect the corresponding proteins. Red edges (arrows) connect proteins that are known to undergo binding based on information from phage ϕ 29 (Tao et al., 1998; Xiang et al., 2006) and (Martin et al., 1998b) but failed the detection in the Y2H screens. The left part of Fig. B indicates proteins and PPIs that presumably play a role during pro-capsid morphogenesis. The right part represents the model of the Cp-1 mature capsid. **(C)** Unexpected PPIs of structural proteins with host proteins and other Cp-1 proteins that were identified.

4.1.3 How can a bacteriophage hijack its host? A comparative view

Bacteriophages are the most common biological entities on earth. Although their genomic features reveal them to function as compact functional entities, it is clear that each phage had to adapt to its specific host to maintain its efficient reproduction while optimizing, modifying, or blocking certain host pathways. A survey of published bacteriophage-host interactions reveals that the targeted host proteins are inhibited, activated or redirected to another function by specific phage proteins. In this discussion section I will propose putative mechanisms how phages can hijack the host by interpreting their genomic features and the detected phage-host PPIs of Cp-1 and Dp-1 while comparing them to known strategies from other bacteriophages.

Phage proteins can optimize the host cell

Optimizing protein translation: It is thought that phages can optimize their reproduction environment by simply expressing their proteins in the host cell. As shown in the results section (3.1.1), the Dp-1 genome encodes for a wide set of queuosine (Q) biosynthesis enzymes. Since Q is known to optimize codon/anti-codon base pairing of tRNAs including Asp, Asn, His, and Tyr (Iwata-Reuyl, 2003), Dp-1 could improve the protein translation process through modifying the host tRNAs by its Q enzymes, thus gaining translation efficiency and specificity. As shown already above (3.1.2), other phages also encode for such Q enzymes but notably Dp-1 seems to have the most complete set. Other phages provide alternative strategies to optimize protein translation. For instance, T4 and T5 are known to introduce their own tRNAs (Scherberg and Weiss, 1972). The phage tRNAs were shown to prefer alternative codons and are able to use a degenerated genetic code. For instance, in the case of T4 the tRNA specific for isoleucine prefers the AUA codon while *E. coli* Ile-tRNA binds with AUU (Scherberg and Weiss, 1972). Scherberg and Weiss demonstrated that the amino acid incorporation into polypeptides is approximately two fold higher when certain T4 tRNAs/mRNAs are used compared to corresponding host tRNAs. This highlights that Dp-1 might also use a degenerated genetic code in certain codon cases and optimizes protein translation by modifying the host tRNAs by its Q enzymes while other phages encode code-optimized tRNAs.

Optimizing nucleotide levels: Dp-1 encodes for a dUTPase (gp14, deoxyuridine triphosphatase), an enzyme that catalyzes the hydrolysis of pyrophosphate from dUTP, maintaining a low intracellular concentration of dUTP so that uracil cannot be incorporated into DNA (Shlomain and Kornberg, 1978). This makes sense since gp14 is presumably co-expressed in a Dp-1 gene cluster encoding for a wide set of DNA replication/repair-related proteins (Fig. 17). Consequently, the intra-cellular levels of dUTP could be lowered during Dp-1 DNA replication and the frequency of mutagenic events through dUTP incorporation into the Dp-1 DNA could be avoided. dUTPases are also encoded by various other phages, e.g., by *E. coli* siphophage T5 (Wang et al., 2005). In fact, it is known that other bacteriophages can introduce various nucleotide-related enzymes to optimize nucleotide levels during

DNA replication. In case of myophage T4, the genome encodes for a dCTPase (gp56), thymidine kinase (Tk), dNMP kinase (gp1), dCMP deaminase (Cd), and dTMP (thymidylate) synthase (Miller et al., 2003) indicating that T4 can globally modify nucleotide levels. However, T4 seems to be an extreme example and most other phages are known to encode only few or none nucleotide-related enzymes. Moreover, other bacteriophages are able to optimize dNTP levels by encoding their own ribonucleotide (NDP) reductases (discussed below).

Genetic studies of T7-host and more recently also of λ -host interactions revealed that several host enzymes involved in nucleotide synthesis pathways are important during the growth of these phages (Maynard et al., 2010; Qimron et al., 2006). Overexpression of *E. coli* uridine-kinase (Udk) and dGTP triphosphatase (Dgt) affected T7 growth showing that high levels of UMP/CMP and low levels of dGTP or their downstream products have adverse effects on phage reproduction. Absence of the CMP/dCMP kinase (Cmk) causes a T7 growth defect, thus indicating that higher levels of CDP/dCTP are of advantage. Similarly, it was shown for λ that the host CMP/dCMP kinase (Cmk) and thymidylate synthetase (ThyA) is required for efficient phage reproduction. In all these cases the effects suppose that the mentioned phages cannot compensate or neutralize adverse effects caused by these nucleotide-related host enzymes, e.g., by introducing self-encoded enzymes and thus, are dependent on natural levels of these host enzymes and the nucleotides they produce.

Inducing host virulence: For a wide range of mainly temperate bacteriophages it is known that they modify the host cell so that they gain or increase their virulence during the lysogenic state (Wagner and Waldor, 2002). The phages benefit from the altered bacterial host since the gained virulence helps their host to adapt to its environment and thus increases its survival. In general, phages introduce factors that stimulate the bacterial invasion, adhesion, and colonization, the main criteria that make a bacterium virulent to humans. For instance, the λ -encoded *lom* gene promotes host adhesion to buccal epithelial cells (Barondess and Beckwith, 1990), *Streptococcus pyogenes* phage H4489A introduces a hyaluronidase and stimulates bacterial host invasion (Hynes and Ferretti, 1989), and *Vibrio cholera* phage CTX ϕ laterally transfers a cholera toxin gene and stimulates toxin production (Waldor and Mekalanos, 1996).

In the case of Dp-1 gp49 I identified by homology predictions an interesting protein that might be involved in improving the *S. pneumoniae* adhesion. Although an N-terminal signal peptide is not predictable, Gp49's C-terminus is homologous to a wide range of internalins (A, F, and G) from *Listeria monocytogenes*. This sequence region contains the Pfam signature "Listeria-Bacteroides repeat domain" (Fig. 19). Adhesion of *S. pneumoniae* cells to nasopharyngeal and oral mucosa cells is usually mediated by adhesins (RrgA, RrgB and RrgC) that are located on the adhesion pili (Nelson et al., 2007). For *S. pneumoniae* no internalins are known. However, in nine other *Streptococcus* species proteins are present that contain the Pfam Listeria-Bacteroides repeat domain (Finn et al., 2010). In *Listeria* internalins play a critical role in invasion and internalization of mammalian cells via

contacting cadherins (Hamon et al., 2006). *Streptococci* are not internalized, thus suggesting that internalin-related proteins rather play a role in adhesion or invasion. The presence of this internalin-related domain in gp49 indicates that the host cell adhesion to epithelial cells might be improved by Dp-1 or altered so that *S. pneumoniae* can colonize a different epithelial tissue than mediated by the endogenous *S. pneumoniae* adhesins.

Do virulent (lytic) bacteriophages kill the host prior to the lytic step?

Virulent phages as Cp-1 and Dp-1 do not integrate their genomes into the host genome, in contrast to temperate phages. Thus, there is no need for them to take care of the host cell because they kill it anyway in the final lytic step. Because of this final “egoistic” behavior, it is supposable that the phages could be able to reprogram the host physiology prior to lysis to such a drastic extent that it is virtually “killed in advance”. Since phages, as soon as they entered the lytic pathway, have to produce large amounts of DNA and structural proteins, they presumably send the host cell into a hyperenergetic state through mobilizing large amounts of energy equivalents. In parallel they could block certain host pathways that are essential for the host but not for the phage.

In the results section I could show that Cp-1 and Dp-1 proteins bind frequently with many essential host proteins. For instance, the Cp-1 connector protein (gp10) binds with the essential host glutamate racemase (MurI). This homodimeric enzyme catalyzes the racemization of L-glutamate to D-glutamate and the latter amino acid is an essential component of the cell wall peptide cross-linkers (Doublet et al., 1993; Doublet et al., 1992). Since the connector protein is known to function as structural component of the virion (Martin et al., 1996b), its interaction with MurI suggests that it has evolved additional functions. The PPI with MurI could result in inhibition of the enzymatic function of the host enzyme resulting in lower levels of D-glutamate. This would help the phage to lyse its host and thus increase its reproductive success. Since the cell wall maintenance/biosynthesis might not be essential for Cp-1 reproduction at least in the late phase, namely then when gp10 is maximally expressed (Martin et al., 1996b), Cp-1 could save energy from the host which was normally “wasted” on peptidoglycane synthesis while supporting indirectly the final lytic step.

Interestingly, it was shown for the ssDNA coliphage ϕ X174 that its lytic protein E has an inhibitory effect on MraY function, an essential host protein which catalyzes the formation of the first lipid-linked intermediate in cell wall biosynthesis (Bernhardt et al., 2000; Zheng et al., 2008). E is a small membrane protein that has no catalytic activity. Its influence on MraY leads to a penicillin-like effect resulting in lysis, thereby preventing the construction of novel peptide cross-links during cell growth. Although dsDNA phages use the substantially different holin-lysine strategy, the detected PPI of Cp-1 gp10 with MurI indicates that also dsDNA phages might inhibit cell wall biosynthesis and thus combine the ϕ X174 penicillin-like lysis strategy with their highly efficient lytic enzymes.

For the lytic T4 and T-even phages it was shown that they degrade the host DNA by endonucleases (Parson and Snustad, 1975). T4 endonucleases can discriminate T4 DNA because

hydroxymethylcytosine rather than cytosine is used in the T4 DNA. This underlines that at least in this phage, the host is killed prior to the lytic step since destruction of the host genome consequently results in the death of the host cell. Thus, it is supposable that also other essential host pathways can be blocked, e.g., in the case of Cp-1 the cell wall biogenesis by the gp10/MurI interaction.

Moreover, it was shown for λ that its protein P blocks host DNA replication through affecting host DnaA (see below) (Datta et al., 2005). Blocking of the host DNA replication seems to play also a general role for other phages. For instance, it was demonstrated that the host sliding clamp DnaN of *Staphylococcus aureus* is blocked through binding with two proteins from phage Twort and G1. Consequently, the sliding clamp-DNA loading event is inhibited and the host replication is arrested (Belley et al., 2006). Hence, these phages can redirect the pool of free dNTPs to use it especially for phage genome synthesis.

Phages fill up their “genomic gaps” through recruitment of host protein

An alternative virus strategy was to recruit host proteins to certain phage-related pathways. This strategy could be applied when proteins are not encoded by the phage itself but whose physical association is necessary to assure the proper function of a molecular system.

DNA replication: As already shown in the results section (Fig. 21) Dp-1 encodes for a wide set of DNA polymerase III subunits except the core subunits α , ϵ , and θ and replication initiator DnaA. Thus, it is supposable that the Dp-1 clamp loader complex has to utilize these proteins from the host to create a functional replisome. Unfortunately, I could not detect any interaction that supports this hypothesis. While most other dsDNA phages use complete sets of DNA polymerase (Weigel and Seitz, 2006), some phages like λ seem to be completely dependent on the presence of a full set of replisomal host proteins. In the case of λ it was shown that λ protein B recruits *E. coli* DNA polymerase III through DnaB interaction to the λ -replication origin (Mallory et al., 1990) while protein B is surprisingly capable to inhibit the host genome replication by blocking the DnaA function (Datta et al., 2005).

In contrast, T4 introduces a full set of 10 DNA replication proteins which makes it virtually independent of the host replication system (Nossal, 1992). The T7 replisome is simple. It includes only a small number of proteins encoded by T7 itself (DNA polymerase gp5, DNA helicase/primase gp4, and SsB gp2.5) (Lee et al., 2006) but also recruits the host thioredoxin which interacts directly with the DNA polymerase gp5 as an accessory protein to bestow processivity on the polymerizing reaction (Tabor et al., 1987).

In the Y2H screens I detected some interesting interactions among Dp-1 proteins with host proteins that are involved in DNA replication/repair. For instance, hypothetical protein gp32 binds to the essential *S. pneumoniae* SsB which is not encoded by Dp-1 but is present in many other phage genomes. Since SsB is important for DNA-related processes (such as phage and host DNA replication and repair; it stabilizes ssDNA strands), this interaction indicates that Dp-1 might recruit this host

protein by protein binding (a blockage of the host SsB makes no sense at that point since Dp-1 does not introduce an own SsB). Although *orf32* is not localized in a DNA-replication operon, its interactions with the Dp-1 primase (DnaG), DNA polymerase I (gp71), and DNA ligase (gp16) portend a role in Dp-1 DNA replication.

On one hand, these examples underline that phages employ very different DNA replication systems. On the other hand, phages are dependent on the host replisome and for some phages it is even known that they block host DNA replication (see above). While some phages require a full set of replisomal host proteins, others recruit specific factors by protein interactions and in turn others introduce partial or full sets of replicational proteins into the host, thus amplifying their genomes more or less autonomously.

DNA packaging: As already mentioned above, the T4 DNA-packaging process is known to function relatively independent of host proteins (Rao and Black, 2005). In contrast, λ recruits the DNA-binding proteins HU and IHF from the host that stimulates the DNA packaging process (Mendelson et al., 1991). The integration host factor (IHF) is a site-specific DNA binding protein that introduces a 180° bend into duplex DNA. *In vitro* DNA packaging is lowered by 4-fold in the absence of IHF (Rice et al., 1996; Yang and Catalano, 2003). It was shown to bind to the *cos*-region (the DNA site where the terminase finally cleaves concatemeric λ DNA) supporting the cooperative DNA-binding of the small terminase subunit (Rice et al., 1996). However, direct protein-protein contacts with terminase components are not described for IHF.

In the case of Dp-1 I was able to detect a surprisingly large number of PPIs with DNA replication proteins as well as *S. pneumoniae* proteins. This indicates that Dp-1 might use a wide set of its own as well as host accessory proteins that are recruited to the terminase complex and thus might function in DNA-packaging (Fig. 45).

Transcription: The idea that Dp-1 and Cp-1 mobilize the *S. pneumoniae* RNA polymerase core (subunits α , β , β') was not confirmed by PPIs but might be the logical consequence since Dp-1 and Cp-1 do not introduce their own RNA polymerase homologs except for three sigma factors in the case of Dp-1.

Other phages introduce complete RNA polymerase systems. For instance, T7 uses initially the multimeric host RNA polymerase for early transcription which is replaced in the middle and late phase by the highly efficient monomeric T7 RNA polymerase. Moreover, it was shown that in the middle and late phase the host RNA polymerase is inhibited by T7 gp2 that binds the β' subunit and thus triggers an effect similar to rifampicin through blocking promoter clearance (Nechaev and Severinov, 1999). However, T7 transcription initiation in the early phase is dependent on the host $\sigma 70$ factor whereas T7 RNA polymerase is specific for its own promoters, a conserved 23 bp sequence (Dunn and Studier, 1983; Nechaev and Severinov, 1999). Involvement of other host σ factors than $\sigma 70$ in phage

gene expression is only known for the siphophage YuA of *Pseudomonas aeruginosa*. It was shown that YuA late gene expression is dependent on host σ_{54} that usually regulates nitrogen-related host operons (Ceysens et al., 2008).

By contrast, λ phage consequently uses the host RNA polymerase but regulates its transcription, e.g., decision of the lytic and lysogenic pathway, mainly through its own repressors and transcription antitermination (Dodd et al., 2005). For instance, λ antitermination protein N was shown to bind to host NusA and thereby recruits NusABE to regulate gene expression of early and middle λ genes (Das and Wolska, 1984; Schauer et al., 1987).

Phage $\phi 29$ uses the *B. subtilis* RNA polymerase but regulates its early and late transcription by two own transcriptional activators (p4 and p6) cooperatively (Camacho and Salas, 2001). P4 recruits the host RNA polymerase to the promoters by interaction with the RNA polymerase α subunit C-terminal domain (Mencia et al., 1996). Since Cp-1 is closely related to $\phi 29$ it could make use of a similar transcriptional strategy although sequence homologs of $\phi 29$ p4 and p6 are not present and might be substituted by Cp-1 specific proteins. Late and middle Dp-1 transcription initiation could be managed by its own σ factors (gp24, gp62, and gp69) while recruiting the multimeric host RNA polymerase.

I detected various PPIs among Cp-1 and Dp-1 proteins with host repressors. It is uncertain that these repressors regulate Cp-1 or Dp-1 gene expression because normally (as in the case of λ and other temperate phages) phages use their own repressors to regulate their gene expression.

Global reprogramming of the host's gene expression by attacking regulatory host proteins

A simple but efficient strategy to manipulate the host system was to attack it on the level of its gene expression. This could result in consequent reprogramming of whole host pathways since they could be easily switched on or off by just targeting single host regulatory proteins. Repression and expression of transcription in bacteria is mostly regulated by repressor proteins that block transcription through binding to certain operator sites, thus inhibiting transcription initiation. In many cases small molecules bind to the repressors and thus regulate DNA-binding and transcription (Fig. 47A). For instance, in the case of the well investigated *E. coli* lactose operon the lac-repressor LacI blocks the expression of genes needed for metabolizing this disaccharide in absence of the catabolite allolactose but presence of the inductor leads to the dissociation of the repressor and expression of the lac-operon genes (Oehler et al., 1990; Wanner et al., 1977). Although expression initiation is also dependent on transcriptional activators, a possible mechanism to reprogram the host's gene expression is to attack such repressor proteins by phage antirepressor proteins that consequently results in an inductor-like state, e.g., by mimicking the inductor or inhibiting the repressors homomerization (Fig. 47C).

Only few reports describe how phage repressors have a direct influence on negative regulation on host gene expression. Chen and colleagues reported that λ cI repressors shut down the expression of the *pckA* gene during lysogeny of several temperate phages (Chen et al., 2005). *pckA*, encoding for

phosphoenol carboxykinase is the first gene involved in gluconeogenesis. Its downregulation suggests an adaptive phage response to growth on energy-poor environments and could lead to an increased lysogen fitness (Chen et al., 2005) (Fig. 47B).

However, this latter case provides evidence for the principle that temperate phages can downregulate the host gene expression by self-encoded repressors. Surprisingly, it was never reported that phages indeed can block host repressors through binding with phage encoded antirepressors. In my work I identified certain interactions among Cp-1 and Dp-1 proteins with host repressor proteins: Cp-1 proteins bind two host repressors (NrdR, FcsR) and Dp-1 proteins bind four host repressors (FcsR, ScrR, SP1050, CodY) (Fig. 47C). These proteins could have a regulatory effect on the host repressors and might especially function as anti-repressors. As shown in Fig. 29E, regulatory host proteins are relatively bound more frequently by phage proteins than by other host proteins. This makes sense on the basis that small molecules act mainly as natural anti-repressors/inductors and thus there is no need for host repressors to undergo regulatory interactions with other host proteins. This regulatory effect could be mimicked similarly by the interacting phage proteins, e.g., preventing binding of the repressor to the DNA through inhibiting their homomerization, binding to their DNA-binding domain, or mimicking the inductor interaction. Consequently, the host gene expression must be activated and the operon gene expression could be switched on. In Fig. 47C I summarized all relevant PPIs that have been identified in this work and that support the hypothesis.

Sugar metabolism: Interactions with repressors of sugar metabolism such as FcsR (SP2168) and ScrR (SP1725) could result in higher levels of the corresponding metabolic gene products while stimulating the metabolism of sugars like saccharose and fucose. Finally, this may provide higher levels of energy equivalents during phage reproduction. However, this would depend on the presence of the corresponding carbohydrates and might only be relevant if *S. pneumoniae* grows, e.g., under saccharose-rich growth conditions. The *S. pneumoniae fcsR* gene is located upstream of the fucose-related operon which contains 10 genes. Expression induction is triggered by fucose and galactose and the operon's gene products represent fucose metabolizing enzymes and PTS components (phosphotransferase phosphoenolpyruvate sugar transport system) (Chan et al., 2003). The role of fucose in pneumococcal metabolism is unclear since *S. pneumoniae* is unable to grow either in L- or D-fucose when provided as sole carbon source in minimal medium. However, it was reported that L-fucose is a minor component of the capsular polysaccharides (Jedrzejewski, 2004).

Reprogramming whole regulons: Another interesting interaction was found with the Dp-1 DNA polymerase III δ' subunit (HolB, gp18) and the host CodY (SP1584), a transcriptional pleiotropic repressor. *S. pneumoniae* CodY was shown to regulate expression of more than 60 genes belonging to the CodY-regulon (Hendriksen et al., 2008). While other bacteria like *B. subtilis* sense the GTP level via CodY and thereby regulate growth transition from the logarithmic to stationary phase (Molle et al., 2003), *S. pneumoniae* CodY was shown to sense only the nutritional state via branched chain amino

acids (BCAAs) like Ile, Leu, and Val. BCAAs bind to the N-terminal dimerization domain of CodY that enhances its DNA-binding, thus blocking gene expression. It was shown that *S. pneumoniae* CodY suppresses in presence of BCAAs the expression of genes that are mainly involved in amino acid metabolism. Furthermore, the *S. pneumoniae* CodY regulates gene expression important for cell adherence to lungs and nasopharyngeal cells (Hendriksen et al., 2008). Binding of the Dp-1 δ' subunit could consequently result in blocking of CodY function and thus expression activation of the whole CodY-regulon could be managed (consider also Fig. 48). The δ' subunit is primarily expected to function in Dp-1 DNA replication but its PPI with CodY suggests an extra-replicative and regulatory function.

How do viruses become independent from the host dNTP level? dsDNA viruses have to face a main problem during replication of their genome, namely that a reasonable pool of dNTPs must be available to ensure the efficient DNA amplification. dNTP levels of the host are well regulated but might not be perfect for the virus.

For the host cells it is important to maintain a constant DNA/cell mass ratio and a complicated system of control and feedback mechanisms are applied to sense the initiation and rate of DNA replication (Herrick and Sclavi, 2007). Initiation is mainly controlled through DnaA (Kaguni, 2006) whereas the elongation rate is mainly regulated by the ribonucleotide reductase (RNR, also known as ribonucleoside diphosphate reductase), a heterodimeric enzyme that converts NDPs to dNDPs (Nordlund and Reichard, 2006). It is the rate-limiting enzyme and ensures a balanced pool of dNTPs during replication since too high levels increase the mutation frequency and a too low levels would result in the stall of replication forks (Herrick and Sclavi, 2007). Its activity itself is controlled by a substrate/product feedback mechanism (Nordlund and Reichard, 2006). Furthermore, in bacteria RNR activity is indirectly controlled by its gene expression which reaches a peak around the initiation time point but decreases to basic levels afterwards (Sun and Fuchs, 1992). Hereby the bacterial NrdR repressor regulates the expression of RNR operons together with DnaA (Grinberg et al., 2006; Herrick and Sclavi, 2007). NrdR consist of a conserved N-terminal DNA-binding domain as well a C-terminal ATP-cone dimerization domain. It was shown to sense nucleotide levels since its ATP-cone domain binds dATP/ATP (Grinberg et al., 2006). A high level of ATP results in its dissociation and a high level of dATP to its association to its repressor binding site. By this means, the expression of the RNR genes are gained or repressed (Grinberg et al., 2009). Hence, NrdR would provide an excellent target for bacterial viruses to increase dNTP levels.

A wide range of phages as well as mammalian viruses solve the dNTP problem by introducing complete RNR systems into the host, like phage T4, T5, and herpesviruses (Berglund et al., 1969; Eriksson and Berglund, 1974; Fossum et al., 2009; Miller et al., 2003). This provides the advantage that they can optimize the dNTP level independently. However, various phage genomes do not encode for RNR homologs including Cp-1 and Dp-1.

I already proposed for Dp-1 a replication model where it had to recruit host DNA pol. III subunits since it does not encode for the core subunits as well as DnaA (remember also the DnaA boxes in the predicted replication origin). Hence, the native host dNTP level might be enough for efficient Dp-1 genome synthesis because its own replication machinery is presumably coupled with the host replication machinery anyway. In contrast, Cp-1 replicates via a simple strategy through linear genome replication (Martin et al., 1996a) that is independent of host replisomal proteins. Surprisingly, I identified an interaction of the hypothetical Cp-1 protein gp6 with the host NrdR repressor. I predict that this interaction could consequently result in gaining expression of *S. pneumoniae* RNR genes, thus increasing the dNTP concentration which supports the Cp-1 genome replication. If confirmed, this strategy would represent a novel principle for phages that do not encode their own RNR. Moreover, gp6 is the only protein in my dataset that binds with the Cp-1 DNA polymerase gp5. I predict that gp6 could function as gp5-sensing protein – as soon as gp5 stalls during replication due to a low dNTP level, gp6 could trigger its effect on NrdR as antirepressor with the consequence of an increasing RNR and thus dNTP level.

Isolating the host from environmental signals: Other targets to modulate host gene expression are its two-component systems (TCS). In bacteria these signal transduction systems are involved in transduction of various environmental signals. The response is mainly differential gene expression. The principle of TCS is illustrated in Fig. 47D. I detected one PPI that indicates that phages could also be able to neutralize environmental signal transduction. Dp-1 Gp44 binds to *S. pneumoniae* LyTr (SP1915), a protein that contains a DNA-binding domain which is unique to TCS response regulator proteins of the AlgR/AgrA/LytR family (Nikolskaya and Galperin, 2002). Although for some *S. pneumoniae* response regulators the exact role has been determined, e.g., involvement in toxin production and competence (Knutsen et al., 2004), the function for SP1915 is unclear. However, the LyTr family members seem to be mainly involved in gene expression regulation of virulence-related genes (Fig. 47D) (Nikolskaya and Galperin, 2002). Consequently, the interaction with gp44 could result in transduction disruption and down-regulation of virulence related operons (Fig. 47D). The existing literature does not describe any single case whereupon phages interfere with two-component signaling. But their disruption could provide an efficient strategy for phages to prevent its host from responding to certain environmental signals that could interfere with phage reproduction.

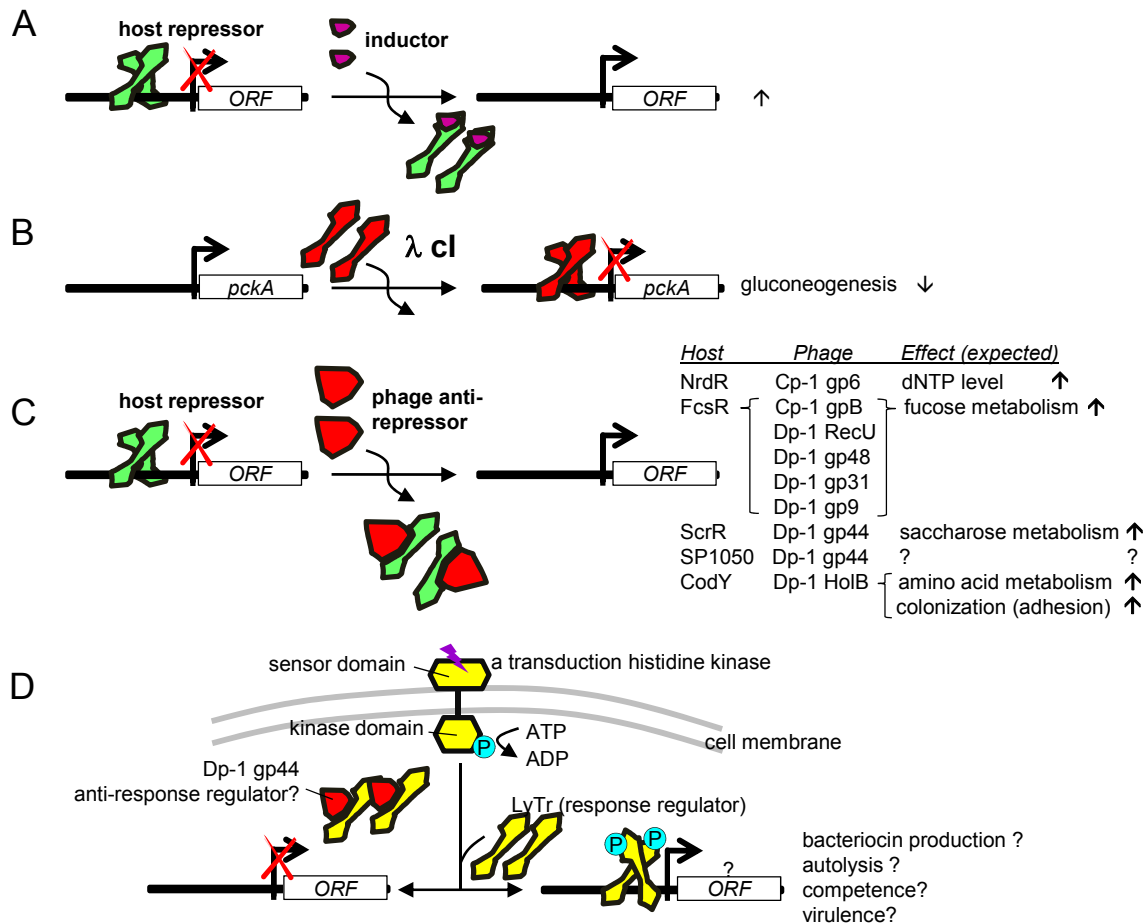


Fig. 47 Putative effects on host gene expression mediated by phage-host PPIs

(A) Host repressor-mediated host gene regulation. Binding of an inducer leads to the dissociation of the host repressor. Note, that small molecules can also function as regulators, that lead to the association of host repressors to their DNA-binding site, e.g., as described in the main text for the CodY repressor. (B) Negative regulation of the host *pckA* gene expression by λ *cl* as suggested by (Chen et al., 2005). (C) Suggested effects of putative Cp-1 and Dp-1 anti-repressors on host gene expression. Putative repressor-anti-repressor systems are listed to the right and the suggested gene expression effects are indicated. These binary PPIs among Cp-1 and Dp-1 proteins with *S. pneumoniae* repressor proteins were detected in this work. (D) Principle of a two-component system and proposed effect of Dp-1 gp44-LyTr interaction: a sensory transduction histidine kinase receives an environmental stimulus (chemical or physical) via its extracellular/periplasmic sensor domain. This leads to its autophosphorylation (P) of histidyl residues. The phosphoryl group then is transferred onto a response regulator that is activated. The response regulator acts mainly as transcriptional activator (or alternatively as repressor; not illustrated) causing differential gene expression as response on environmental changes. Gp44 could act as anti-response regulator through binding with LyTr, thereby neutralizing the signal transduction. Illustrations are simplified.

glyceraldehyde 3-phosphate dehydrogenase (Gap, SP2012) interactions. The latter one would result in modification of glycolysis. However, the proposed effect of phage proteins on transcriptional regulators would clearly result in a more dramatic reprogramming of the metabolic pathways since more metabolic enzymes and uptake systems are under transcriptional control (Fig. 47 and Fig. 48). Surprisingly, phage PPIs with metabolic enzymes are only known in one case. Roucourt and colleagues recently identified an interaction among an early protein of *Pseudomonas aeruginosa* infecting podophage ϕ KMV with the malate synthase G which is involved in pyruvate metabolism (Roucourt et al., 2009). The authors expect the interaction to have a regulatory effect on the enzyme's activity.

The physiological meaning of the Cp-1 lysozyme-host uridine-cytidine kinase interaction

Originally, I hypothesized that the identified interaction of Cp-1 lysozyme (Cpl1) with the *S. pneumoniae* uridine-cytidine kinase (Udk) results in the inhibition of the enzymatic function of Udk. Since the Udk products CMP and UMP and their downstream products are important for mRNA synthesis and host capsule synthesis (as indicated for Udk (SP1208) in Fig. 48) the inhibitory effect of Cpl1 on Udk could result in controlling or lowering the levels of pyrimidine nucleotides and thus mRNA and capsule synthesis. Maximal Cpl1 protein levels are expected in the very late infection phase, namely then when the lytic step takes place. Although low expression of *cpl1* is expected to occur also at earlier periods, low levels of this protein might have no impact on Udk, depending on the protein level of Udk itself. It was shown in this study by several independent methods that Udk binds with Cpl1 (Y2H, LuMPIS, Y2H domain mapping, gel filtration, and protein cross-linking) (3.1.5). However, I could not clearly demonstrate that Cpl1 has neither an effect on the natural homotetrameric Udk structure nor on its enzymatic activity under *in vitro* conditions.

Nevertheless, these findings do not exclude the possibility of Cpl1 to bind to Udk *in vivo*. Here, the PPI could alternatively function to retain Cpl1 by Udk while making it inaccessible to the holin-mediated extracellular transport and to lyse the host cell. Thus, lysis could be regulated through this PPI in dependence on the Udk protein level. Therefore, the actual protein level of Udk would function as a Cpl1-specific "protein buffer system". Quimron and colleagues found that T7 growth is restricted as soon as *E. coli* Udk is overexpressed (Quimron et al., 2006). Although no experimental evidence is available yet that demonstrates direct binding of T7 lysozyme with *E. coli* Udk, the inhibitory effect of Udk at high levels could be explained by its proposed regulatory effect on the lysozyme. In any case, the genetic link for T7 and the direct PPI between the Cp-1 lysozyme and the *S. pneumoniae* Udk suggests a conserved role of Udk in siphophage reproduction.

4.2 Functional and conserved association of the YbeB protein family with the large ribosomal subunit and protein translation

The evolutionary distribution of YbeB orthologs link it to the eubacterial origin

The YbeB protein family has a broad evolutionary distribution and is present in eukaryotes and Eubacteria while it is absent in Archaea. Transcription and translation in Archaea is more similar to that of eukaryotes, e.g., the usage of methionine as initiator amino acid in contrast to N-formyl-methionine, the higher complexity of translation factor systems, and components such as rRNAs are of eukaryotic origin (Allers and Mevarech, 2005; Dennis, 1997). This indicates that YbeB is associated with an eubacterial process. Consequently, YbeB orthologs in Eukaryota might rather function in eubacterial derivatives such as plastids and mitochondria than in the cytosolic compartment: In fact, eukaryotic YbeB homologs localize to chloroplasts in *Zea mays* and mitochondria in baker's yeast (Han and Martienssen, 1995; Reinders et al., 2006). Moreover, I found out that the human YbeB orthologs C7orf30 is targeted exclusively to mitochondria as well. These results indicate also on the localization level that YbeB orthologs are linked to the eubacterial origin.

L14 recruits YbeB to an important 50S-30S interface region on the large ribosomal subunit

High-throughput screens have identified many interaction partners of YbeB orthologs in various organisms (see supplementary Tab. 37 for a detailed list). In Tab. 34 I summarized all available interactions that link YbeB orthologs with L14 and/or ribosomal subunits. Previous studies have found YbeB orthologs to be associated with the large ribosomal. This strongly suggests that the YbeB function is linked in a conserved manner with the large subunit.

Jiang and colleagues (Jiang et al., 2007) showed that *E. coli* YbeB co-migrates exclusively with the 50S particle and is absent from 70S, polysomes, and the small ribosomal subunit fraction. In addition, co-migration with 50S progenitors has been ruled out, indicating that YbeB does not play a role in 50S biogenesis. However, previous studies of ribosomal protein complexes could not resolve the key binding site for YbeB, e.g., if there are many docking sites, specific docking sites or even where these putative binding sites are localized on the ribosomal surface.

In *Campylobacter jejuni* YbeB has been recently identified by Parrish and colleagues to bind to L14, among other proteins (Parrish et al., 2007). Moreover, I identified many interaction partners of *T. pallidum* YbeB including L14 (Titz et al., 2008). This overlap motivated me to search systematically for conserved interactions of *E. coli* YbeB - the interaction with L14 turned out to be the only conserved one. No other interaction originally identified in *T. pallidum* and *C. jejuni* turned out to be conserved across multiple species, indicating that they are either species specific or unspecific altogether.

DISCUSSION

L14 is localized on the 50S subunit facing the 30S subunit and critical in contacting the latter. Besides L14, only L2, L5, and L19 are located on the inter-subunit surface of 50S, forming additional protein-16S rRNA bridges with the small subunit (Gao et al., 2003). During transition from an initiation-like

Tab. 34 Interactions of YbeB orthologs with L14 and the large ribosomal subunit

All known interactions including co-purifications with ribosomes from this and other studies. L14 interactions from binary interaction assays are shaded in light grey. Co-purifications with complexes are highlighted in dark grey.

Species	Ortholog (locus tag)		Description interaction partner	Method	Reference
	YbeB	L14/50S			
<i>Escherichia coli</i>	b0637	50S subunit	50S large ribosomal subunit	iTRAQ, pull down	(Jiang et al., 2007)
		b3310	50S ribosomal subunit protein L14	Y2H, pull down	this work
				LC-MS	(Butland et al., 2005)
<i>Treponema pallidum</i>	TP0738	TP0199	50S ribosomal protein L14	Y2H	(Titz et al., 2008), this work
<i>Campylobacter jejuni</i>	Cj1405	Cj1697c	50S ribosomal protein L14	Y2H	(Parrish et al., 2007)
<i>S. pneumoniae</i> TIGR4	SP1744	SP0219	50S ribosomal protein L14	Y2H	this work
<i>Synechocystis</i> PCC 6803	slr1886	sll1806	50S ribosomal protein L14	Y2H	this work
<i>Homo sapiens</i>	C7orf30	L14 _{mt}	39S ribosomal protein L14, mitochondrial	BiFC, pull down	this work
<i>Zea mays</i>	lj (Iojap)	rp14	50S ribosomal protein L14, chloroplastic	Pull down	this work
		50S subunit	50S chloroplast ribosomal subunit	IP	(Han and Martienssen, 1995)
<i>Saccharomyces cerevisiae</i>	YMR098C	54S subunit	Mitochondrial large subunit	MALDI-TOF MS	(Gavin et al., 2006)

to EF-G·GTP bound state, the 50S and 30S subunit undergo a ratchet-like motion relative to each other. While the rRNAs show only little conformational changes, L2, L5, and L14 undergo large local movements and trigger the ratchet-like motion. Especially L14 is highly dynamic and rotates by 14° but maintains the contact to 16S rRNA via the B5 and B8 bridge (Gao et al., 2003). However, little is known about the structural dynamics of the ribosome during 50S-30S assembly. I could show that YbeB binds to critical residues of L14 that are involved in bridge B8. Because of these findings, YbeB may function as a specific L14 capping protein that prevents assembly of 50S and 30S subunits in a regulated fashion.

Such a regulatory, rather than a structural role, is also suggested by the stoichiometry of YbeB which is only present at one copy per 10 assembled ribosomes (Ishihama et al., 2008) which may be similar to the number of free ribosomal subunits. By contrast, Jing and colleagues demonstrated that most of YbeB protein migrates at the top of a density gradient, thus representing free protein. However, YbeB was also shown to clearly associate with the 50S particle (Jiang et al., 2007). This suggests that either there is an excess of YbeB compared to ribosomes or that YbeB does not stoichiometrically bind to 50S subunits. In any case, it is likely that YbeB binding to 50S subunits must be regulated, given its role in translation, and yet unknown regulatory mechanisms may determine its stoichiometry.

Although I examined the interactions by binary detection methods, there is no doubt that YbeB acts on L14 on the 50S particle in contrast to the possibility to bind only with free L14. This is explainable by the findings of Jiang et al., especially since YbeB has been co-purified with ribosomal protein complexes and 50S particles (Tab. 34). Moreover, ribosomal proteins might be quickly incorporated into rRNA during ribosome biogenesis and removed from the free pool. It is supposable that YbeB and L14 are also capable to bind *in vivo* binarily under “real”, non-experimental conditions in absence of the ribosome since my results indicate that a binary contact definitely occurs. Extra-ribosomal functions have been described for certain bacterial ribosomal proteins: S1 functions in RNA polymerase regulation and S10, S4, L3, L4, and L13 function in transcriptional antitermination (Greive et al., 2005; Sukhodolets and Garges, 2003; Torres et al., 2001). Moreover, S8 was shown to retro-regulate the stability of *spc* operon mRNA encoding for more than 10 ribosomal proteins. S8 takes influence on the mRNA stability by binding to the L5 gene, the third gene in this operon, and thus blocks translation of it (Dean et al., 1981). The *rplN* gene encoding for L14 is localized as the first ORF in the *spc* operon. It was demonstrated that *rplN* plays a role in retroregulation and stability of *spc* as well (Liang et al., 1999), but a direct link with L14 protein itself is missing. If L14 would play a role in retroregulation of *spc*, freely available YbeB protein could function as sensor of the pool of free L14. If an access of free L14 was available, L14 could take negative influence on *spc* stability. But since YbeB orthologs associate also in eukaryotes with the large ribosomal subunit and transcriptional regulation is remarkably different to bacteria, an universal extra-ribosomal YbeB function is not plausible. Also the interaction epitope mapping experiments imply YbeB function with the ribosomal particle since YbeB binds to the surface region of L14 that faces the 30S interface. A more general chaperone function of YbeB can be excluded since the *ybeB* gene deletion mutant exhibits no chaperone-related phenotype.

YbeB – physiological function

The exact role and molecular function of YbeB in translation remains mysterious and might be only indirect due to altered assembly of ribosomal subunits. Functional studies are also impaired by the lack of easily detected phenotypes in YbeB mutants. The knock-out has no deleterious effect under optimal growth conditions nor does treatment of the mutant with various antibiotics that interfere with translation affect this process. Similarly, no sensitivity to osmotic, heat, or cold stress is detectable in the mutant. In the competitive time-course assay, however, the *ybeB* KO strain was clearly outcompeted by wild-type cells after several generations. This clearly shows that *ybeB* mutants have a selective disadvantage. By contrast, a homozygous nuclear mutation in the *Zea mays* ortholog *ij* (Iojap) gene leads to irregular albino striped patterns on maize leaves (Fig. 8), germless seeds, and albino seedlings caused by a failure of proplastids to differentiate into chloroplasts (Jenkins, 1924; Rhoades, 1943). The defective plastids contain few internal membranes and no ribosomes while the plastids maintain a normal genome (Jenkins, 1924; Shumway and Weier, 1967; Thompson et al.,

1983). Besides loss of ribosomal particles, affected cells were also free from plastid-encoded polypeptides while transcripts were detectable (although their amount and processing was altered) (Han et al., 1993). This already indicated that translation in plastids breaks down while transcripts are still present. However, as soon as the ribosomes are lost in *Iojap*-deficient plastids they cannot be resynthesized (Zubko and Day, 1998). It remains unclear how the *ij* mutation leads to the breakdown of the translational machinery. Han and Martienssen were able to demonstrate the binding of *Iojap* to the plastidal large ribosomal subunit and hypothesized *Iojap* to be a regulatory factor of plastid translation (Han and Martienssen, 1995). The failure to synthesize ribosomal proteins might explain the loss of ribosomes and imply that in maize the *YbeB* homolog is essential, at least in maintenance of the plastid translation system.

In contrast, in yeast mutations in the nuclear gene of the mitochondrial *YbeB* ortholog ATP25 (YMR098C) lead to pleiotropic phenotypes, e.g., decreased acid resistance, decrease of glycogen level, heat sensitivity, but also decreased competitive fitness (Deutschbauer et al., 2005; Mira et al., 2009; Sinha et al., 2008; Wilson et al., 2002). Moreover, Zeng and colleagues detected a growth defect under respiratory conditions in presence of glycerol/ethanol coupled with high-temperature conditions (Zeng et al., 2008). In yeast the situation is more complicated than in bacteria since mitochondrial ATP25 is cleaved into two halves. The C-terminal half was shown to specifically stabilize the mRNA of mitochondrial F_1F_0 -ATPase subunit C. This sequence region shows only homology towards protein sequences in few budding and filamentous fungi (Zeng et al., 2008). However, restoration of Atp9p (OLI1) subunit needs also presence of the N-terminal half of ATP25 that carries the DUF143 signature (Zeng et al., 2008). Zeng et al. proposed the N-terminal half to function as factor that helps to assemble the Atp9p ring of F_1F_0 -ATPase since Atp9p was still freshly synthesized in a mutant, expressing only the C-terminal half of ATP25. But an assembly of Atp9p ring was not observable. Notably, the ATP25 null mutation in yeast does not lead to the breakdown of translational machinery in mitochondria as in the case of *Iojap*-affected maize plastids. *In vivo* labeling revealed that ATP25 mutants are capable to translate all mitochondrial gene products except Atp9p (Zeng et al., 2008). I tried to find experimental evidence also for ATP25 to bind to mitochondrial L14 of yeast. However, the baits and preys turned out to be toxic when transfected into yeast and thus could not be examined by Y2H (data not shown). Nevertheless, ATP25 was found by Gavin and colleagues as an “attachment” of the mitochondrial large ribosomal subunit, even though such “attachments” were considered to be rather loosely associated as opposed to “core” or “modular” components (Gavin et al., 2006). This also indicates that ATP25 could play a general role in mitochondrial translation although the direct L14-binding evidence is still missing. While the pleiotropic phenotypes can be explained by the role of both ATP25 halves in maintenance of the F_1F_0 -ATPase system, the role of the N-terminal half of ATP25 in translation is rather circumstantial. The *E. coli* *YbeB* mutant did not exhibit a growth deficiency under respiratory conditions when grown in minimal medium containing glucose as carbon source. This makes it unlikely that *YbeB* is responsible for maintenance of the

ATPase subunit C in *E. coli*. Furthermore, organisms which do not encode for YbeB orthologs but for F₁F₀-ATPase subunit C like *Mycoplasma*, *Buchnera*, or Archaea (COG0636) could indicate that the relevance of this specific functional link is a non-conserved finding and a specific property of yeast ATP25.

Kazuta et al. recently examined the effects of nearly all *E. coli* ORFs on translation *in vitro* (Kazuta et al., 2008). They found YbeB to have a slightly positive effect (rank 217 of 344 beneficial factors) although the effect was quite mild. They analyzed samples after 3 h post induction of a GFP reporter *in vitro* by a single experimental snapshot. In my experiments β -galactosidase reporter protein levels had already adopted nearly to wild-type levels at this point (Fig. 43C). This delayed measurement could explain the contradictory findings, given that I already found peak effects after 1 to 2 h.

Finally, the hypothesis that YbeB functions as a negative modulator in protein translation is supported by the *in silico*-protein docking results (Fig. 44): as soon as YbeB is bound to L14 on a 50S particle, it interferes sterically with the contacts that form between L14 and 16S rRNA in the assembled 70S complex. Consequently, YbeB interferes with the assembly of the functional ribosome.

5 Conclusions and outlook

Cp-1 and Dp-1

This work provides a fully and comprehensively annotated genome of the lytic siphophage Dp-1. The results help to understand its proteomic composition and the arrangement of the Dp-1 genome. However, this was a first step and further experiments will be necessary to understand the exact function/regulation of hypothetical proteins and the predicted operons. The genome annotation revealed that Dp-1 seems to be capable to optimize its reproduction environment, e.g., through introducing queuosine enzymes that presumably optimize protein translation.

The protein-interaction maps of Cp-1 and Dp-1 represent the most comprehensive networks for phages so far. They are helpful to link hypothetical proteins to certain cellular processes, e.g., as demonstrated in great detail for the virion morphogenesis of Cp-1 and Dp-1. Combination of homology-based gene product annotation with Y2H data and mass spectrometry data help to get a better understanding how these virions are assembled and which gene products could participate in this process. This first level analysis is helpful to get novel ideas about the function of individual phage proteins and whole virion-related pathways.

The identified PPIs are helpful to get a broad overview which host proteins are presumably attacked by phage proteins and thus which host pathways could be important for Cp-1 and Dp-1 reproduction.

Proteome/genome-wide analyses detect unexpected phenomena since tests are done comprehensively and without hypothesis-driven bias. Thus, they often raise more questions than they can answer. Such experiments provide a lot of experimental data and are mainly helpful to generate hypotheses. The major challenge will be to verify them by detailed experiments.

The next step will be to check the physiological relevance of the phage-host PPIs *in vivo*. This will be done by testing host gene knock-outs strains for phage reproduction efficiency as it was already demonstrated for model phage T7 (Qimron et al., 2006). However, this will be not a trivial task since *S. pneumoniae* and Cp-1 and Dp-1 are not as easy to manipulate as *E. coli* and its phages. It would be also interesting to know which *S. pneumoniae* genes are up- or down-regulated during phage reproduction by using transcriptomic analyses. This could confirm the proposed phage-interaction effects on host repressor proteins on the gene expression level, e.g., the predicted upregulation of RNR-related operons in the case of the Cp-1 gp6/NrdR PPI and the CodY regulon in the case of Dp-1 gp19/CodY PPI. Furthermore, biochemical assays will be done to investigate the effects of phage proteins on host protein functions *in vitro*. Of special interest is the Holliday junction helicase RuvB that turned out to bind to several Dp-1 proteins and a Cp-1 protein. Effects of the interacting phage proteins onto RuvB/RuvA helicase activity will be tested by competitive branch migration assays of Holliday intermediates *in vitro*.

YbeB

The results of this work could demonstrate for the first time that YbeB of *E. coli* functions in protein translation *in vivo* and that its exact docking region on the ribosome is on L14. However, the exact function of YbeB and its orthologs in translation remains unclear. In different systems this protein group seems to affect translation but to a different extent. In *E. coli* protein levels prematurely increase in the absence of YbeB. Nonetheless, these effects are not dramatic. They could explain why there is no general growth defect since the speed of protein translation is usually not critical. Nevertheless, these effects are detectable and imply YbeB in translation as a negative regulator *in vivo*.

Taken all information, I propose YbeB's general role in translation as an accessory protein that specifically caps L14. YbeB might stabilize free 50S particles or prevent them from unproductive assembly with 30S particles. Thus, YbeB might have a dynamic effect on translation initiation or ribosome recycling. The existing literature does not describe that any ribosome-associated proteins function as capping proteins that hide other critical 50S-30S interface proteins, e.g., L2 and L5 or ribosomal RNA surfaces. A related question is how the dissociation of YbeB from 50S is regulated. Ribosome gradients of $\Delta ybeB$ of *E. coli* do not exhibit differences of 50S, 30S, 70S, or polysome compositions (Jiang et al., 2007). They were examined under continuous growth conditions. It would be helpful to examine ribosome profiles of $\Delta ybeB$ under changing conditions since I could show that the induction of reporter genes lead to a faster translational output in *ybeB* mutant cells.

Nothing was known about where YbeB binds on the 50S subunit prior to this work. My results now link YbeB and all its orthologs specifically to L14. This information is very helpful to identify its exact, molecular function since it limits its mode of action on the ribosome. Open questions are, e.g., which signal leads to the dissociation or association of YbeB from the 50S subunit, are there any factors else involved, does YbeB also interact with rRNA and thus can modify it? This will be tough to answer since the YbeB protein family is unique on the structural and sequence level. For other ribosome associated factors their function was easier to resolve since they often contain a GTPase or transferase signature implying their enzymatic mode of action. This is not the case for YbeB. YbeB could mediate additional rRNA or even tRNA contacts. The latter could lead during initiation to dissociation of YbeB.

Nonetheless, my results show that the data from high-throughput experiments are very valuable once followed up with experiments and combined with any existing information from other datasets. They help to get detailed ideas about the function of conserved hypothetical proteins, even if they are tough to analyze. A great finding of this work was that even the human YbeB-homolog, C7orf30, undergoes the interaction with mitochondrial L14. This highlights that we still can transfer conserved findings from bacterial research to human biology.

References

- Ackermann, H.-W. (2006). Classification of Bacteriophages: Oxford University Press.
- Ahmed, N. K. and Welch, A. D. (1979). Some properties of uridine-cytidine kinase from a human malignant lymphoma. *Cancer Res* **39**, 3102-6.
- Albano, C. R., Randers-Eichhorn, L., Bentley, W. E. and Rao, G. (1998). Green fluorescent protein as a real time quantitative reporter of heterologous protein production. *Biotechnol Prog* **14**, 351-4.
- Albrich, W. C., Monnet, D. L. and Harbarth, S. (2004). Antibiotic selection pressure and resistance in *Streptococcus pneumoniae* and *Streptococcus pyogenes*. *Emerg Infect Dis* **10**, 514-7.
- Allers, T. and Mevarech, M. (2005). Archaeal genetics - the third way. *Nat Rev Genet* **6**, 58-73.
- Almaghrabi, M. and Clokie, M. (2010). The isolation of *Streptococcus pneumoniae* lytic phages and their in vivo and in vitro interactions. In *Poster presented on Pasteur Meeting Viruses of Microbes 2010*, (ed).
- AlonsoDeVelasco, E., Verheul, A. F., Verhoef, J. and Snippe, H. (1995). *Streptococcus pneumoniae*: virulence factors, pathogenesis, and vaccines. *Microbiol Rev* **59**, 591-603.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-10.
- Altschul, S. F. and Koonin, E. V. (1998). Iterated profile searches with PSI-BLAST--a tool for discovery in protein databases. *Trends Biochem Sci* **23**, 444-7.
- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. et al. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* **32**, D115-9.
- Aravind, L. and Koonin, E. V. (2000). Eukaryote-specific domains in translation initiation factors: implications for translation regulation and evolution of the translation system. *Genome Res* **10**, 1172-84.
- Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H. C., Hirai, A. et al. (2006). Large-scale identification of protein-protein interaction of *Escherichia coli* K-12. *Genome Res* **16**, 686-91.
- Ashelford, K. E., Day, M. J. and Fry, J. C. (2003). Elevated abundance of bacteriophage infecting bacteria in soil. *Appl Environ Microbiol* **69**, 285-9.
- Avery, O. T., Macleod, C. M. and McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from *Pneumococcus* Type Iii. *J Exp Med* **79**, 137-58.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008.
- Bamford, D. H., Bamford, J. K., Towse, S. A. and Thomas, G. J., Jr. (1990). Structural study of the lipid-containing bacteriophage PRD1 and its capsid and DNA components by laser Raman spectroscopy. *Biochemistry* **29**, 5982-7.
- Bamford, D. H., Palva, E. T. and Lounatmaa, H. (1976). Ultrastructure and Life Cycle of the Lipid-containing Bacteriophage phi6
J. Gen. Virol. **32**, 249-259.
- Bamford, D. H., Rouhiainen, L., Takkinen, K. and Soderlund, H. (1981). Comparison of the lipid-containing bacteriophages PRD1, PR3, PR4, PR5 and L17. *J Gen Virol* **57**, 365-73.
- Baptista, C., Santos, M. A. and Sao-Jose, C. (2008). Phage SPP1 reversible adsorption to *Bacillus subtilis* cell wall teichoic acids accelerates virus recognition of membrane receptor YueB. *J Bacteriol* **190**, 4989-96.
- Barondess, J. J. and Beckwith, J. (1990). A bacterial virulence determinant encoded by lysogenic coliphage lambda. *Nature* **346**, 871-4.
- Barrios-Rodiles, M., Brown, K. R., Ozdamar, B., Bose, R., Liu, Z., Donovan, R. S., Shinjo, F., Liu, Y., Dembowy, J., Taylor, I. W. et al. (2005). High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* **307**, 1621-5.
- Bartel, P. L., Roecklein, J. A., SenGupta, D. and Fields, S. (1996). A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat Genet* **12**, 72-7.
- Belley, A., Callejo, M., Arhin, F., Dehbi, M., Fadhil, I., Liu, J., McKay, G., Srikumar, R., Bauda, P., Bergeron, D. et al. (2006). Competition of bacteriophage polypeptides with native replicase proteins for binding to the DNA sliding clamp reveals a novel mechanism for DNA replication arrest in *Staphylococcus aureus*. *Mol Microbiol* **62**, 1132-43.
- Benelli, D., Maone, E. and Londei, P. (2003). Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation. *Mol Microbiol* **50**, 635-43.
- Berg, O., Borsheim, K. Y., Bratbak, G. and Heldal, M. (1889). High abundance of viruses found in aquatic environments. *Nature* **340**, 467-468

- Berglund, O., Karlstrom, O. and Reichard, P.** (1969). A new ribonucleotide reductase system after infection with phage T4. *Proc Natl Acad Sci U S A* **62**, 829-35.
- Bernhardt, T. G., Roof, W. D. and Young, R.** (2000). Genetic evidence that the bacteriophage phi X174 lysis protein inhibits cell wall synthesis. *Proc Natl Acad Sci U S A* **97**, 4297-302.
- Bernheimer, H. P.** (1979). Lysogenic pneumococci and their bacteriophages. *J Bacteriol* **138**, 618-24.
- Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F. et al.** (1997). The complete genome sequence of Escherichia coli K-12. *Science* **277**, 1453-62.
- Borodovsky, M., Mills, R., Besemer, J. and Lomsadze, A.** (2003). Prokaryotic gene prediction using GeneMark and GeneMark.hmm. *Curr Protoc Bioinformatics Chapter 4*, Unit4 5.
- Bothwell, A. L., Stark, B. C. and Altman, S.** (1976). Ribonuclease P substrate specificity: cleavage of a bacteriophage phi80-induced RNA. *Proc Natl Acad Sci U S A* **73**, 1912-6.
- Boxem, M., Maliga, Z., Klitgord, N., Li, N., Lemmens, I., Mana, M., de Lichtervelde, L., Mul, J. D., van de Peut, D., Devos, M. et al.** (2008). A protein domain-based interactome network for *C. elegans* early embryogenesis. *Cell* **134**, 534-45.
- Bradley, D. E.** (1969). Ultrastructure of bacteriophages and bacteriocins *Bacteriol. Revs* **31**, 230-314
- Braun, P., Tasan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., Yu, H., Sahalie, J. M., Murray, R. R., Roncari, L., de Smet, A. S. et al.** (2009). An experimentally derived confidence score for binary protein-protein interactions. *Nat Methods* **6**, 91-7.
- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P. and Rohwer, F.** (2003). Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* **185**, 6220-3.
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F. and Rohwer, F.** (2002). Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A* **99**, 14250-5.
- Bronzwaer, S. L., Cars, O., Buchholz, U., Molstad, S., Goetsch, W., Veldhuijzen, I. K., Kool, J. L., Sprenger, M. J. and Degener, J. E.** (2002). A European study on the relationship between antimicrobial use and antimicrobial resistance. *Emerg Infect Dis* **8**, 278-82.
- 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. et al.** (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058-73.
- Busso, D., Delagoutte-Busso, B. and Moras, D.** (2005). Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*. *Anal Biochem* **343**, 313-21.
- Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N. et al.** (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531-7.
- Byrne, M. and Taylor, W. C.** (1996). Analysis of Mutator-induced mutations in the *iojap* gene of maize. *Mol Gen Genet* **252**, 216-20.
- Cagney, G., Uetz, P. and Fields, S.** (2001). Two-hybrid analysis of the *Saccharomyces cerevisiae* 26S proteasome. *Physiol Genomics* **7**, 27-34.
- Camacho, A. and Salas, M.** (2001). Mechanism for the switch of phi29 DNA early to late transcription by regulatory protein p4 and histone-like protein p6. *EMBO J* **20**, 6060-70.
- Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. and Darst, S. A.** (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell* **9**, 527-39.
- Campbell, T. L. and Brown, E. D.** (2008). Genetic interaction screens with ordered overexpression and deletion clone sets implicate the *Escherichia coli* GTPase YjeQ in late ribosome biogenesis. *J Bacteriol* **190**, 2537-45.
- Catalano, C. E., Cue, D. and Feiss, M.** (1995). Virus DNA packaging: the strategy used by phage lambda. *Mol Microbiol* **16**, 1075-86.
- Ceyssens, P. J., Mesyanzhinov, V., Sykilinda, N., Briers, Y., Roucourt, B., Lavigne, R., Robben, J., Domashin, A., Miroshnikov, K., Volckaert, G. et al.** (2008). The genome and structural proteome of YuA, a new *Pseudomonas aeruginosa* phage resembling M6. *J Bacteriol* **190**, 1429-35.
- Chai, S., Kruff, V. and Alonso, J. C.** (1994). Analysis of the *Bacillus subtilis* bacteriophages SPP1 and SF6 gene 1 product: a protein involved in the initiation of headful packaging. *Virology* **202**, 930-9.
- Chan, P. F., O'Dwyer, K. M., Palmer, L. M., Ambrad, J. D., Ingraham, K. A., So, C., Lonetto, M. A., Biswas, S., Rosenberg, M., Holmes, D. J. et al.** (2003). Characterization of a novel fucose-regulated promoter (P_{fcsK}) suitable for gene essentiality and antibacterial mode-of-action studies in *Streptococcus pneumoniae*. *J Bacteriol* **185**, 2051-8.
- Chang, J., Weigele, P., King, J., Chiu, W. and Jiang, W.** (2006). Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. *Structure* **14**, 1073-82.

- Chelbi-Alix, M. K., Expert-Bezancon, A., Hayes, F., Alix, J. H. and Branlant, C.** (1981). Properties of ribosomes and ribosomal RNAs synthesized by *Escherichia coli* grown in the presence of ethionine. Normal maturation of ribosomal RNA in the absence of methylation. *Eur J Biochem* **115**, 627-34.
- Chen, Y., Golding, I., Sawai, S., Guo, L. and Cox, E. C.** (2005). Population fitness and the regulation of *Escherichia coli* genes by bacterial viruses. *PLoS Biol* **3**, e229.
- Chen, Y. C., Rajagopala, S. V., Stellberger, T. and Uetz, P.** (2010). Benchmarking protein interaction methods by exhaustive yeast two-hybrid testing. *Nature Methods*, In Press.
- Cherepanov, P. P. and Wackernagel, W.** (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9-14.
- Choi, K. H., McPartland, J., Kaganman, I., Bowman, V. D., Rothman-Denes, L. B. and Rossmann, M. G.** (2008). Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *J Mol Biol* **378**, 726-36.
- Cochran, P. K. and Paul, J. H.** (1998). Seasonal abundance of lysogenic bacteria in a subtropical estuary. *Appl Environ Microbiol* **64**, 2308-12.
- Colson, C., Lhoest, J. and Urlings, C.** (1979). Genetics of ribosomal protein methylation in *Escherichia coli*. III. Map position of two genes, *prmA* and *prmB*, governing methylation of proteins L11 and L3. *Mol Gen Genet* **169**, 245-50.
- Cramer, A., Whitehorn, E. A., Tate, E. and Stemmer, W. P.** (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* **14**, 315-9.
- Crooks, G. E., Hon, G., Chandonia, J. M. and Brenner, S. E.** (2004). WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-90.
- d'Herelle, F.** (1917). Sur un microbe invisible antagoniste des bacilles dysentériques. *C. R. Acad. Sci. Paris* **165**, 373-75.
- d'Herelle, F.** (1926a). The Bacteriophage and Its Behavior. *Baltimore, MD: Williams & Wilkins.*, 490-97.
- d'Herelle, F.** (1926b). The Bacteriophage and Its Behavior. *Baltimore, MD: Williams & Wilkins.*, 497-503.
- d'Herelle, F., Malone, R. H. and Lahiri, M.** (1929). Studies Upon Asiatic Cholera. *Indian Med. Res. Mem.* **14**, 1-161.
- Das, A. and Wolska, K.** (1984). Transcription antitermination in vitro by lambda N gene product: requirement for a phage nut site and the products of host *nusA*, *nusB*, and *nusE* genes. *Cell* **38**, 165-73.
- Datsenko, K. A. and Wanner, B. L.** (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-5.
- Datta, I., Sau, S., Sil, A. K. and Mandal, N. C.** (2005). The bacteriophage lambda DNA replication protein P inhibits the *oriC* DNA- and ATP-binding functions of the DNA replication initiator protein DnaA of *Escherichia coli*. *J Biochem Mol Biol* **38**, 97-103.
- Daubin, V. and Ochman, H.** (2004). Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. *Genome Res* **14**, 1036-42.
- Davidson, T., Beck, E., Ganapathy, A., Montgomery, R., Zafar, N., Yang, Q., Madupu, R., Goetz, P., Galinsky, K., White, O. et al.** (2010). The comprehensive microbial resource. *Nucleic Acids Res* **38**, D340-5.
- Dawes, J.** (1975). Characterisation of the bacteriophage T4 receptor site. *Nature* **256**, 127-8.
- Day, R. S., 3rd.** (1977). UV-induced alleviation of K-specific restriction of bacteriophage lambda. *J Virol* **21**, 1249-51.
- de Beer, T., Fang, J., Ortega, M., Yang, Q., Maes, L., Duffy, C., Berton, N., Sippy, J., Overduin, M., Feiss, M. et al.** (2002). Insights into specific DNA recognition during the assembly of a viral genome packaging machine. *Mol Cell* **9**, 981-91.
- Dean, D., Yates, J. L. and Nomura, M.** (1981). *Escherichia coli* ribosomal protein S8 feedback regulates part of *spc* operon. *Nature* **289**, 89-91.
- Decatur, W. A. and Fournier, M. J.** (2002). rRNA modifications and ribosome function. *Trends Biochem Sci* **27**, 344-51.
- Dennis, P. P.** (1997). Ancient ciphers: translation in Archaea. *Cell* **89**, 1007-10.
- Deutschbauer, A. M., Jaramillo, D. F., Proctor, M., Kumm, J., Hillenmeyer, M. E., Davis, R. W., Nislow, C. and Giaever, G.** (2005). Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* **169**, 1915-25.
- Diaz, E., Lopez, R. and Garcia, J. L.** (1992). EJ-1, a temperate bacteriophage of *Streptococcus pneumoniae* with a Myoviridae morphotype. *J Bacteriol* **174**, 5516-25.
- Diefenbacher, M., Sekula, S., Heilbock, C., Maier, J. V., Litfin, M., van Dam, H., Castellazzi, M., Herrlich, P. and Kassel, O.** (2008). Restriction to Fos family members of Trip6-dependent coactivation and glucocorticoid receptor-dependent trans-repression of activator protein-1. *Mol Endocrinol* **22**, 1767-80.
- DiMauro, A. J., Lin, D., Guo, S., Karr, D. B., Tanner, J. J. and Guo, P.** (2007). Crystallization of Phi29 spindle-shaped nano-bar anti-receptor with glycosidase domain. *J Nanosci Nanotechnol* **7**, 2616-22.

- Doan, D. N. and Dokland, T.** (2007). The gpQ portal protein of bacteriophage P2 forms dodecameric connectors in crystals. *J Struct Biol* **157**, 432-6.
- Dodd, I. B., Shearwin, K. E. and Egan, J. B.** (2005). Revisited gene regulation in bacteriophage lambda. *Curr Opin Genet Dev* **15**, 145-52.
- Dorson, J. W., Deutsch, W. A. and Moses, R. E.** (1978). Role of DNA polymerases in excision repair in *Escherichia coli*. *J Biol Chem* **253**, 660-4.
- Dosztanyi, Z., Chen, J., Dunker, A. K., Simon, I. and Tompa, P.** (2006). Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res* **5**, 2985-95.
- Doublet, P., van Heijenoort, J., Bohin, J. P. and Mengin-Lecreulx, D.** (1993). The murI gene of *Escherichia coli* is an essential gene that encodes a glutamate racemase activity. *J Bacteriol* **175**, 2970-9.
- Doublet, P., van Heijenoort, J. and Mengin-Lecreulx, D.** (1992). Identification of the *Escherichia coli* murI gene, which is required for the biosynthesis of D-glutamic acid, a specific component of bacterial peptidoglycan. *J Bacteriol* **174**, 5772-9.
- Driedonks, R. A., Engel, A., tenHeggeler, B. and van, D.** (1981). Gene 20 product of bacteriophage T4 its purification and structure. *J Mol Biol* **152**, 641-62.
- Droge, A. and Tavares, P.** (2000). In vitro packaging of DNA of the *Bacillus subtilis* bacteriophage SPP1. *J Mol Biol* **296**, 103-15.
- Duane, P. G., Rubins, J. B., Weisel, H. R. and Janoff, E. N.** (1993). Identification of hydrogen peroxide as a *Streptococcus pneumoniae* toxin for rat alveolar epithelial cells. *Infect Immun* **61**, 4392-7.
- Dube, P., Tavares, P., Lurz, R. and van Heel, M.** (1993). The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *EMBO J* **12**, 1303-9.
- Dunn, J. J. and Studier, F. W.** (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol* **166**, 477-535.
- Edwards, R. A. and Rohwer, F.** (2005). Viral metagenomics. *Nat Rev Microbiol* **3**, 504-10.
- Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H.** (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**, 953-71.
- Eriksson, S. and Berglund, O.** (1974). Bacteriophage-induced ribonucleotide reductase systems. T5- and T6-specific ribonucleotide reductase and thioredoxin. *Eur J Biochem* **46**, 271-8.
- Estojak, J., Brent, R. and Golemis, E. A.** (1995). Correlation of two-hybrid affinity data with in vitro measurements. *Mol Cell Biol* **15**, 5820-9.
- Faruque, S. M. and Mekalanos, J. J.** (2003). Pathogenicity islands and phages in *Vibrio cholerae* evolution. *Trends Microbiol* **11**, 505-10.
- Fields, S. and Song, O.** (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-6.
- Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K. et al.** (2010). The Pfam protein families database. *Nucleic Acids Res* **38**, D211-22.
- Fischetti, V. A.** (2010). Bacteriophage endolysins: A novel anti-infective to control Gram-positive pathogens. *Int J Med Microbiol*.
- Fokine, A., Chipman, P. R., Leiman, P. G., Mesyanzhinov, V. V., Rao, V. B. and Rossmann, M. G.** (2004). Molecular architecture of the prolate head of bacteriophage T4. *Proc Natl Acad Sci U S A* **101**, 6003-8.
- Fossum, E., Friedel, C. C., Rajagopala, S. V., Titz, B., Baiker, A., Schmidt, T., Kraus, T., Stellberger, T., Rutenberg, C., Suthram, S. et al.** (2009). Evolutionarily conserved herpesviral protein interaction networks. *PLoS Pathog* **5**, e1000570.
- Franklin, R. M., Salditt, M. and Silbert, J. A.** (1969). Structure and synthesis of a lipid-containing bacteriophage : I. Growth of bacteriophage PM2 and alterations in nucleic acid metabolism in the infected cell. *Virology* **38**, 627-640.
- Fraser, C. M., Eisen, J., Fleischmann, R. D., Ketchum, K. A. and Peterson, S.** (2000). Comparative genomics and understanding of microbial biology. *Emerg Infect Dis* **6**, 505-12.
- Fuhrman, J. A. and Noble, R. T.** (1995). Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnology and Oceanography* **40**, 1236-1242
- Galperin, M. Y.** (2001). Conserved 'hypothetical' proteins: new hints and new puzzles. *Comp Funct Genomics* **2**, 14-8.
- Galperin, M. Y. and Koonin, E. V.** (2004). 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res* **32**, 5452-63.
- Gan, X., Kitakawa, M., Yoshino, K., Oshiro, N., Yonezawa, K. and Isono, K.** (2002). Tag-mediated isolation of yeast mitochondrial ribosome and mass spectrometric identification of its new components. *Eur J Biochem* **269**, 5203-14.
- Gao, F. and Zhang, C. T.** (2008). Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. *BMC Bioinformatics* **9**, 79.

- Gao, H., Sengupta, J., Valle, M., Korostelev, A., Eswar, N., Stagg, S. M., Van Roey, P., Agrawal, R. K., Harvey, S. C., Sali, A. et al. (2003). Study of the structural dynamics of the E coli 70S ribosome using real-space refinement. *Cell* **113**, 789-801.
- García, E., Ronda, C. and Lopez, R. (1979). Bacteriophages of *Streptococcus pneumoniae*. Physicochemical properties of bacteriophage Dp-4 and its transfecting DNA. *Eur J Biochem* **101**, 59-64.
- García, J. L., García, E., Arraras, A., García, P., Ronda, C. and Lopez, R. (1987). Cloning, purification, and biochemical characterization of the pneumococcal bacteriophage Cp-1 lysin. *J Virol* **61**, 2573-80.
- García, P., García, J. L., López, R. and García, E. (2005). *Pneumococcal phages*. Washington, D.C: ASM Press.
- García, P., Martín, A. C. and Lopez, R. (1997). Bacteriophages of *Streptococcus pneumoniae*: a molecular approach. *Microb Drug Resist* **3**, 165-76.
- García, P., Mendez, E., García, E., Ronda, C. and Lopez, R. (1984). Biochemical characterization of a murein hydrolase induced by bacteriophage Dp-1 in *Streptococcus pneumoniae*: comparative study between bacteriophage-associated lysin and the host amidase. *J Bacteriol* **159**, 793-6.
- García, P., Paz Gonzalez, M., García, E., García, J. L. and Lopez, R. (1999). The molecular characterization of the first autolytic lysozyme of *Streptococcus pneumoniae* reveals evolutionary mobile domains. *Mol Microbiol* **33**, 128-38.
- Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B. et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631-6.
- Gindreau, E., Lopez, R. and García, P. (2000). MM1, a temperate bacteriophage of the type 23F Spanish/USA multiresistant epidemic clone of *Streptococcus pneumoniae*: structural analysis of the site-specific integration system. *J Virol* **74**, 7803-13.
- Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E. et al. (2003). A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727-36.
- Gray, M. W. (1989). The evolutionary origins of organelles. *Trends Genet* **5**, 294-9.
- Gray, M. W. (1992). The endosymbiont hypothesis revisited. *Int Rev Cytol* **141**, 233-357.
- Greive, S. J., Lins, A. F. and von Hippel, P. H. (2005). Assembly of an RNA-protein complex. Binding of NusB and NusE (S10) proteins to boxA RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in *Escherichia coli*. *J Biol Chem* **280**, 36397-408.
- Griffith, F. (1928). The Significance of Pneumococcal Types. *J Hyg (Lond)* **27**, 113-59.
- Grimes, S. and Anderson, D. (1997). The bacteriophage phi29 packaging proteins supercoil the DNA ends. *J Mol Biol* **266**, 901-14.
- Grinberg, I., Shteinberg, T., Gorovitz, B., Aharonowitz, Y., Cohen, G. and Borovok, I. (2006). The *Streptomyces* NrdR transcriptional regulator is a Zn ribbon/ATP cone protein that binds to the promoter regions of class Ia and class II ribonucleotide reductase operons. *J Bacteriol* **188**, 7635-44.
- Grinberg, I., Shteinberg, T., Hassan, A. Q., Aharonowitz, Y., Borovok, I. and Cohen, G. (2009). Functional analysis of the *Streptomyces* coelicolor NrdR ATP-cone domain: role in nucleotide binding, oligomerization, and DNA interactions. *J Bacteriol* **191**, 1169-79.
- Guo, P., Grimes, S. and Anderson, D. (1986). A defined system for in vitro packaging of DNA-gp3 of the *Bacillus subtilis* bacteriophage phi 29. *Proc Natl Acad Sci U S A* **83**, 3505-9.
- Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, 4121-30.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **20**, 426-427.
- Halter, C. P., Peeters, N. M. and Hanson, M. R. (2004). RNA editing in ribosome-less plastids of iojap maize. *Curr Genet* **45**, 331-7.
- Hamon, M., Bierne, H. and Cossart, P. (2006). *Listeria monocytogenes*: a multifaceted model. *Nat Rev Microbiol* **4**, 423-34.
- Han, C.-d., Coe, E. H., Jr. and Martienssen, R. A. (1992). Molecular cloning and characterization of iojap (ij), a pattern striping gene of maize. *EMBO J* **11**, 4037-46.
- Han, C.-d. and Martienssen, R. A. (1995). The Iojap (Ij) protein is associated with 50S chloroplast ribosomal subunits. *MNL* **69**, 32-34.
- Han, C.-d., Patrie, W., Polacco, M. and Coe, E. H. (1993). Aberrations in plastid transcripts and deficiency of plastid DNA in striped and albino mutants in maize. *Planta* **191**, 552-563.
- Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. and Elledge, S. J. (1993). The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805-16.
- Hendriksen, W. T., Bootsma, H. J., Esteveao, S., Hoogenboezem, T., de Jong, A., de Groot, R., Kuipers, O. P. and Hermans, P. W. (2008). CodY of *Streptococcus pneumoniae*: link between nutritional gene regulation and colonization. *J Bacteriol* **190**, 590-601.

- Hermoso, J. A., Monterroso, B., Albert, A., Galan, B., Ahrazem, O., Garcia, P., Martinez-Ripoll, M., Garcia, J. L. and Menendez, M. (2003). Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. *Structure* **11**, 1239-49.
- Herrick, J. and Sclavi, B. (2007). Ribonucleotide reductase and the regulation of DNA replication: an old story and an ancient heritage. *Mol Microbiol* **63**, 22-34.
- Herskowitz, I. and Hagen, D. (1980). The lysis-lysogeny decision of phage lambda: explicit programming and responsiveness. *Annu Rev Genet* **14**, 399-445.
- Hilton, J. L., Kearney, P. C. and Ames, B. N. (1965). Mode of action of the herbicide, 3-amino-1,2,4-triazole(amtrole): inhibition of an enzyme of histidine biosynthesis. *Arch Biochem Biophys* **112**, 544-7.
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J. and Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res* **35**, W585-7.
- Hoshino, T., McKenzie, T., Schmidt, S., Tanaka, T. and Sueoka, N. (1987). Nucleotide sequence of *Bacillus subtilis* dnaB: a gene essential for DNA replication initiation and membrane attachment. *Proc Natl Acad Sci U S A* **84**, 653-7.
- Hoskins, J., Alborn, W. E., Jr., Arnold, J., Blaszcak, L. C., Burgett, S., DeHoff, B. S., Estrem, S. T., Fritz, L., Fu, D. J., Fuller, W. et al. (2001). Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J Bacteriol* **183**, 5709-17.
- Hsiao, C. L. and Black, L. W. (1978). Head morphogenesis of bacteriophage T4. II. The role of gene 40 in initiating prehead assembly. *Virology* **91**, 15-25.
- Hu, C. D., Chinenov, Y. and Kerppola, T. K. (2002). Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell* **9**, 789-98.
- Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L. et al. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**, D211-5.
- Hynes, W. L. and Ferretti, J. J. (1989). Sequence analysis and expression in *Escherichia coli* of the hyaluronidase gene of *Streptococcus pyogenes* bacteriophage H4489A. *Infect Immun* **57**, 533-9.
- Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F. U., Kerner, M. J. and Frishman, D. (2008). Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* **9**, 102.
- Isidro, A., Henriques, A. O. and Tavares, P. (2004a). The portal protein plays essential roles at different steps of the SPP1 DNA packaging process. *Virology* **322**, 253-63.
- Isidro, A., Santos, M. A., Henriques, A. O. and Tavares, P. (2004b). The high-resolution functional map of bacteriophage SPP1 portal protein. *Mol Microbiol* **51**, 949-62.
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* **98**, 4569-74.
- Iwata-Reuyl, D. (2003). Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA. *Bioorg Chem* **31**, 24-43.
- James, P., Halladay, J. and Craig, E. A. (1996). Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144**, 1425-36.
- Jedrzejewski, M. J. (2004). Extracellular virulence factors of *Streptococcus pneumoniae*. *Front Biosci* **9**, 991-914.
- Jenkins, M. T. (1924). Heritable characters of maize. XX. Iojap-striping, a chlorophyll defect. *J. Hered.* **15**, 467-472.
- Jenner, L. B., Demeshkina, N., Yusupova, G. and Yusupov, M. (2010). Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat Struct Mol Biol* **17**, 555-60.
- Jiang, M., Datta, K., Walker, A., Strahler, J., Bagamasbad, P., Andrews, P. C. and Maddock, J. R. (2006). The *Escherichia coli* GTPase CgtAE is involved in late steps of large ribosome assembly. *J Bacteriol* **188**, 6757-70.
- Jiang, M., Sullivan, S. M., Walker, A. K., Strahler, J. R., Andrews, P. C. and Maddock, J. R. (2007). Identification of novel *Escherichia coli* ribosome-associated proteins using isobaric tags and multidimensional protein identification techniques. *J Bacteriol* **189**, 3434-44.
- Jin, F., Avramova, L., Huang, J. and Hazbun, T. (2007). A yeast two-hybrid smart-pool-array system for protein-interaction mapping. *Nat Methods* **4**, 405-7.
- Jin, F., Hazbun, T., Michaud, G. A., Salcius, M., Predki, P. F., Fields, S. and Huang, J. (2006). A pooling-deconvolution strategy for biological network elucidation. *Nat Methods* **3**, 183-9.
- Johnsson, N. and Varshavsky, A. (1994). Ubiquitin-assisted dissection of protein transport across membranes. *EMBO J* **13**, 2686-98.
- Jones, D. T., Shirley, M., Wu, X. and Keis, S. (2000). Bacteriophage infections in the industrial acetone butanol (AB) fermentation process. *J Mol Microbiol Biotechnol* **2**, 21-6.
- Jothi, R., Manikandakumar, K., Ganesan, K. and Parthasarathy, S. (2008). On the analysis of the virulence nature of TIGR4 and R6 strains of *Streptococcus pneumoniae* using genome comparison tools *Journal of Chemical Sciences* **119**, 559-563.

- Kaczanowska, M. and Ryden-Aulin, M.** (2007). Ribosome biogenesis and the translation process in *Escherichia coli*. *Microbiol Mol Biol Rev* **71**, 477-94.
- Kaguni, J. M.** (2006). DnaA: controlling the initiation of bacterial DNA replication and more. *Annu Rev Microbiol* **60**, 351-75.
- Kang, W. K., Icho, T., Isono, S., Kitakawa, M. and Isono, K.** (1989). Characterization of the gene rimK responsible for the addition of glutamic acid residues to the C-terminus of ribosomal protein S6 in *Escherichia coli* K12. *Mol Gen Genet* **217**, 281-8.
- Katsura, I. and Hendrix, R. W.** (1984). Length determination in bacteriophage lambda tails. *Cell* **39**, 691-8.
- Kazuta, Y., Adachi, J., Matsuura, T., Ono, N., Mori, H. and Yomo, T.** (2008). Comprehensive analysis of the effects of *Escherichia coli* ORFs on protein translation reaction. *Mol Cell Proteomics* **7**, 1530-40.
- Kemper, B. and Brown, D. T.** (1976). Function of gene 49 of bacteriophage T4. II. Analysis of intracellular development and the structure of very fast-sedimenting DNA. *J Virol* **18**, 1000-15.
- Kibbe, W. A.** (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* **35**, W43-6.
- Kirby, E. P., Jacob, F. and Goldthwait, D. A.** (1967). Prophage induction and filament formation in a mutant strain of *Escherichia coli*. *Proc Natl Acad Sci U S A* **58**, 1903-10.
- Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H. and Mori, H.** (2005). Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res* **12**, 291-9.
- Knutsen, E., Ween, O. and Havarstein, L. S.** (2004). Two separate quorum-sensing systems upregulate transcription of the same ABC transporter in *Streptococcus pneumoniae*. *J Bacteriol* **186**, 3078-85.
- Kochan, J., Carrascosa, J. L. and Murialdo, H.** (1984). Bacteriophage lambda preconnectors. Purification and structure. *J Mol Biol* **174**, 433-47.
- Kreuzer, K. N.** (2000). Recombination-dependent DNA replication in phage T4. *Trends Biochem Sci* **25**, 165-73.
- Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P. et al.** (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637-43.
- Krzyzosiak, W., Denman, R., Nurse, K., Hellmann, W., Boublik, M., Gehrke, C. W., Agris, P. F. and Ofengand, J.** (1987). In vitro synthesis of 16S ribosomal RNA containing single base changes and assembly into a functional 30S ribosome. *Biochemistry* **26**, 2353-64.
- Laemmli, U. K.** (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5.
- Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. and Ben-Tal, N.** (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* **33**, W299-302.
- Landy, A.** (1989). Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem* **58**, 913-49.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R. et al.** (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-8.
- Larsen, T. S. and Krogh, A.** (2003). EasyGene--a prokaryotic gene finder that ranks ORFs by statistical significance. *BMC Bioinformatics* **4**, 21.
- Lee, J. B., Hite, R. K., Hamdan, S. M., Xie, X. S., Richardson, C. C. and van Oijen, A. M.** (2006). DNA primase acts as a molecular brake in DNA replication. *Nature* **439**, 621-4.
- Lee, M. S., Dougherty, B. A., Madeo, A. C. and Morrison, D. A.** (1999). Construction and analysis of a library for random insertional mutagenesis in *Streptococcus pneumoniae*: use for recovery of mutants defective in genetic transformation and for identification of essential genes. *Appl Environ Microbiol* **65**, 1883-90.
- Leiman, P. G., Kanamaru, S., Mesyanzhinov, V. V., Arisaka, F. and Rossmann, M. G.** (2003). Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **60**, 2356-70.
- Leplae, R., Lima-Mendez, G. and Toussaint, A.** (2010). ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res* **38**, D57-61.
- Lesnik, E. A., Sampath, R., Levene, H. B., Henderson, T. J., McNeil, J. A. and Ecker, D. J.** (2001). Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res* **29**, 3583-94.
- Letunic, I. and Bork, P.** (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127-8.
- Letunic, I., Doerks, T. and Bork, P.** (2009). SMART 6: recent updates and new developments. *Nucleic Acids Res* **37**, D229-32.
- Lhuillier, S., Gallopin, M., Gilquin, B., Brasiles, S., Lancelot, N., Letellier, G., Gilles, M., Dethan, G., Orlova, E. V., Couprie, J. et al.** (2009). Structure of bacteriophage SPPI head-to-tail connection reveals mechanism for viral DNA gating. *Proc Natl Acad Sci U S A* **106**, 8507-12.

- Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T. et al. (2004). A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540-3.
- Liang, S. T., Ehrenberg, M., Dennis, P. and Bremer, H. (1999). Decay of rplN and lacZ mRNA in *Escherichia coli*. *J Mol Biol* **288**, 521-38.
- Lindahl, L. (1975). Intermediates and time kinetics of the in vivo assembly of *Escherichia coli* ribosomes. *J Mol Biol* **92**, 15-37.
- Little, J. W. (1984). Autodigestion of lexA and phage lambda repressors. *Proc Natl Acad Sci U S A* **81**, 1375-9.
- Lock, R. A., Hansman, D. and Paton, J. C. (1992). Comparative efficacy of autolysin and pneumolysin as immunogens protecting mice against infection by *Streptococcus pneumoniae*. *Microb Pathog* **12**, 137-43.
- Loeffler, J. M., Djurkovic, S. and Fischetti, V. A. (2003). Phage lytic enzyme Cpl-1 as a novel antimicrobial for pneumococcal bacteremia. *Infect Immun* **71**, 6199-204.
- Loeffler, J. M., Nelson, D. and Fischetti, V. A. (2001). Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. *Science* **294**, 2170-2.
- Lof, D., Schillen, K., Jonsson, B. and Evilevitch, A. (2007). Forces controlling the rate of DNA ejection from phage lambda. *J Mol Biol* **368**, 55-65.
- Londei, P. (2005). Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol Rev* **29**, 185-200.
- Lopez, R., Garcia, E., Garcia, P., Ronda, C. and Tomasz, A. (1982). Choline-containing bacteriophage receptors in *Streptococcus pneumoniae*. *J Bacteriol* **151**, 1581-90.
- Lopez, R., Ronda, C., Garcia, P., Escarmis, C. and Garcia, E. (1984). Restriction cleavage maps of the DNAs of *Streptococcus pneumoniae* bacteriophages containing protein covalently bound to their 5' ends. *Mol Gen Genet* **197**, 67-74.
- Lopez, R., Ronda, C., Tomasz, A. and Portoles, A. (1977). Properties of "diplophage": a lipid-containing bacteriophage. *J Virol* **24**, 201-10.
- Lu, Z., Breidt, F., Plengvidhya, V. and Fleming, H. P. (2003). Bacteriophage ecology in commercial sauerkraut fermentations. *Appl Environ Microbiol* **69**, 3192-202.
- Lucchini, S., Desiere, F. and Brussow, H. (1998). The structural gene module in *Streptococcus thermophilus* bacteriophage phi Sfi11 shows a hierarchy of relatedness to Siphoviridae from a wide range of bacterial hosts. *Virology* **246**, 63-73.
- Luftig, R. B., Wood, W. B. and Okinaka, R. (1971). Bacteriophage T4 head morphogenesis. On the nature of gene 49-defective heads and their role as intermediates. *J Mol Biol* **57**, 555-73.
- Lukashin, A. V. and Borodovsky, M. (1998). GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* **26**, 1107-15.
- Mallory, J. B., Alfano, C. and McMacken, R. (1990). Host virus interactions in the initiation of bacteriophage lambda DNA replication. Recruitment of *Escherichia coli* DnaB helicase by lambda P replication protein. *J Biol Chem* **265**, 13297-307.
- Malys, N., Chang, D. Y., Baumann, R. G., Xie, D. and Black, L. W. (2002). A bipartite bacteriophage T4 SOC and HOC randomized peptide display library: detection and analysis of phage T4 terminase (gp17) and late sigma factor (gp55) interaction. *J Mol Biol* **319**, 289-304.
- Marchler-Bauer, A. and Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32**, W327-31.
- Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O. and Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science* **285**, 751-3.
- Markmann-Mulisch, U. and Subramanian, A. R. (1988). Nucleotide sequence and linkage map position of the genes for ribosomal proteins L14 and S8 in the maize chloroplast genome. *Eur J Biochem* **170**, 507-14.
- Markson, G., Kiel, C., Hyde, R., Brown, S., Charalabous, P., Bremm, A., Semple, J., Woodsmith, J., Duley, S., Salehi-Ashtiani, K. et al. (2009). Analysis of the human E2 ubiquitin conjugating enzyme protein interaction network. *Genome Res* **19**, 1905-11.
- Martin, A. C., Blanco, L., Garcia, P., Salas, M. and Mendez, J. (1996a). In vitro protein-primed initiation of pneumococcal phage Cp-1 DNA replication occurs at the third 3' nucleotide of the linear template: a stepwise sliding-back mechanism. *J Mol Biol* **260**, 369-77.
- Martin, A. C., Lopez, R. and Garcia, P. (1996b). Analysis of the complete nucleotide sequence and functional organization of the genome of *Streptococcus pneumoniae* bacteriophage Cp-1. *J Virol* **70**, 3678-87.
- Martin, A. C., Lopez, R. and Garcia, P. (1998a). Functional analysis of the two-gene lysis system of the pneumococcal phage Cp-1 in homologous and heterologous host cells. *J Bacteriol* **180**, 210-7.
- Martin, A. C., Lopez, R. and Garcia, P. (1998b). Pneumococcal bacteriophage Cp-1 encodes its own protease essential for phage maturation. *J Virol* **72**, 3491-4.
- Martinez-Jimenez, M. I., Alonso, J. C. and Ayora, S. (2005). *Bacillus subtilis* bacteriophage SPP1-encoded gene 34.1 product is a recombination-dependent DNA replication protein. *J Mol Biol* **351**, 1007-19.

- Marzi, S., Fechter, P., Chevalier, C., Romby, P. and Geissmann, T.** (2008). RNA switches regulate initiation of translation in bacteria (*). *Biol Chem*.
- Maslov, S. and Sneppen, K.** (2002). Specificity and stability in topology of protein networks. *Science* **296**, 910-3.
- Maynard, N. D., Birch, E. W., Sanghvi, J. C., Chen, L., Gutschow, M. V. and Covert, M. W.** (2010). A forward-genetic screen and dynamic analysis of lambda phage host-dependencies reveals an extensive interaction network and a new anti-viral strategy. *PLoS Genet* **6**, e1001017.
- McCafferty, J., Griffiths, A. D., Winter, G. and Chiswell, D. J.** (1990). Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552-4.
- McDonnell, M., Lain, R. and Tomasz, A.** (1975). "Diplophage": a bacteriophage of *Diplococcus pneumoniae*. *Virology* **63**, 577-82.
- Meier, F., Suter, B., Grosjean, H., Keith, G. and Kubli, E.** (1985). Queuosine modification of the wobble base in tRNA^{His} influences 'in vivo' decoding properties. *EMBO J* **4**, 823-7.
- Mencia, M., Monsalve, M., Rojo, F. and Salas, M.** (1996). Transcription activation by phage phi29 protein p4 is mediated by interaction with the alpha subunit of *Bacillus subtilis* RNA polymerase. *Proc Natl Acad Sci U S A* **93**, 6616-20.
- Mendelson, I., Gottesman, M. and Oppenheim, A. B.** (1991). HU and integration host factor function as auxiliary proteins in cleavage of phage lambda cohesive ends by terminase. *J Bacteriol* **173**, 1670-6.
- Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juarez, K., Contreras-Moreira, B. et al.** (2009). Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS One* **4**, e7526.
- Messing, J., Crea, R. and Seeburg, P. H.** (1981). A system for shotgun DNA sequencing. *Nucleic Acids Res* **9**, 309-21.
- Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. and Ruger, W.** (2003). Bacteriophage T4 genome. *Microbiol Mol Biol Rev* **67**, 86-156, table of contents.
- Mira, N. P., Lourenco, A. B., Fernandes, A. R., Becker, J. D. and Sa-Correia, I.** (2009). The RIM101 pathway has a role in *Saccharomyces cerevisiae* adaptive response and resistance to propionic acid and other weak acids. *FEMS Yeast Res* **9**, 202-16.
- Molle, V., Nakaura, Y., Shivers, R. P., Yamaguchi, H., Losick, R., Fujita, Y. and Sonenshein, A. L.** (2003). Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J Bacteriol* **185**, 1911-22.
- Morris, R. C. and Elliott, M. S.** (2001). Queuosine modification of tRNA: a case for convergent evolution. *Mol Genet Metab* **74**, 147-59.
- Mosig, G., Gewin, J., Luder, A., Colowick, N. and Vo, D.** (2001). Two recombination-dependent DNA replication pathways of bacteriophage T4, and their roles in mutagenesis and horizontal gene transfer. *Proc Natl Acad Sci U S A* **98**, 8306-11.
- Mulder, N. and Apweiler, R.** (2007). InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol Biol* **396**, 59-70.
- Murphy, K. C.** (1998). Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol* **180**, 2063-71.
- Nakai, K. and Horton, P.** (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* **24**, 34-6.
- Nechaev, S. and Severinov, K.** (1999). Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T7 gene 2 protein. *J Mol Biol* **289**, 815-26.
- Neidhardt, F. C., Bloch, P. L. and Smith, D. F.** (1974). Culture medium for enterobacteria. *J Bacteriol* **119**, 736-47.
- Nelson, A. L., Ries, J., Bagnoli, F., Dahlberg, S., Falker, S., Rounioja, S., Tschop, J., Morfeldt, E., Ferlenghi, I., Hilleringmann, M. et al.** (2007). RrgA is a pilus-associated adhesin in *Streptococcus pneumoniae*. *Mol Microbiol* **66**, 329-40.
- Nikolskaya, A. N. and Galperin, M. Y.** (2002). A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res* **30**, 2453-9.
- Nordlund, P. and Reichard, P.** (2006). Ribonucleotide reductases. *Annu Rev Biochem* **75**, 681-706.
- Nossal, N. G.** (1992). Protein-protein interactions at a DNA replication fork: bacteriophage T4 as a model. *FASEB J* **6**, 871-8.
- O'Brien, K. L., Wolfson, L. J., Watt, J. P., Henkle, E., Deloria-Knoll, M., McCall, N., Lee, E., Mulholland, K., Levine, O. S. and Cherian, T.** (2009). Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet* **374**, 893-902.
- Obregon, V., Garcia, J. L., Garcia, E., Lopez, R. and Garcia, P.** (2003). Genome organization and molecular analysis of the temperate bacteriophage MM1 of *Streptococcus pneumoniae*. *J Bacteriol* **185**, 2362-8.

- Ochman, H., Lawrence, J. G. and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304.
- Oehler, S., Eismann, E. R., Kramer, H. and Muller-Hill, B. (1990). The three operators of the lac operon cooperate in repression. *EMBO J* **9**, 973-9.
- Ogawa, T. and Okazaki, T. (1984). Function of RNase H in DNA replication revealed by RNase H defective mutants of Escherichia coli. *Mol Gen Genet* **193**, 231-7.
- Orren, D. K., Selby, C. P., Hearst, J. E. and Sancar, A. (1992). Post-incision steps of nucleotide excision repair in Escherichia coli. Disassembly of the UvrBC-DNA complex by helicase II and DNA polymerase I. *J Biol Chem* **267**, 780-8.
- Parks, R. E. and Agarwal, R. P. (1973). 9 Nucleoside Diphosphokinases. In *The Enzymes*, P.D. Boyer, ed. (New York: Academic Press), 307-334.
- Parrish, J. R., Yu, J., Liu, G., Hines, J. A., Chan, J. E., Mangiola, B. A., Zhang, H., Pacifico, S., Fotouhi, F., DiRita, V. J. et al. (2007). A proteome-wide protein interaction map for Campylobacter jejuni. *Genome Biol* **8**, R130.
- Parson, K. A. and Snustad, D. P. (1975). Host DNA degradation after infection of Escherichia coli with bacteriophage T4: dependence of the alternate pathway of degradation which occurs in the absence of both T4 endonuclease II and nuclear disruption on T4 endonuclease IV. *J Virol* **15**, 221-4.
- Pearson, W. (2004). Finding protein and nucleotide similarities with FASTA. *Curr Protoc Bioinformatics Chapter 3*, Unit3 9.
- Pelletier, J., Gros, P. and Dubow, M. (2000). Development of novel anti-microbial agents based on bacteriophage genomics, vol. U.S. patent W00032825A2 (ed).
- Perez-Dorado, I., Campillo, N. E., Monterroso, B., Heseck, D., Lee, M., Paez, J. A., Garcia, P., Martinez-Ripoll, M., Garcia, J. L., Mobashery, S. et al. (2007). Elucidation of the molecular recognition of bacterial cell wall by modular pneumococcal phage endolysin CPL-1. *J Biol Chem* **282**, 24990-9.
- Plisson, C., White, H. E., Auzat, I., Zafarani, A., Sao-Jose, C., Lhuillier, S., Tavares, P. and Orlova, E. V. (2007). Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. *EMBO J* **26**, 3720-8.
- Porter, R. D. and Guild, W. R. (1976). Characterization of some pneumococcal bacteriophages. *J Virol* **19**, 659-67.
- Qimron, U., Marintcheva, B., Tabor, S. and Richardson, C. C. (2006). Genomewide screens for Escherichia coli genes affecting growth of T7 bacteriophage. *Proc Natl Acad Sci U S A* **103**, 19039-44.
- Rain, J. C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V. et al. (2001). The protein-protein interaction map of Helicobacter pylori. *Nature* **409**, 211-5.
- Rajagopala, S. V., Hughes, K. T. and Uetz, P. (2009). Benchmarking yeast two-hybrid systems using the interactions of bacterial motility proteins. *Proteomics* **9**, 5296-302.
- Randall-Hazelbauer, L. and Schwartz, M. (1973). Isolation of the bacteriophage lambda receptor from Escherichia coli. *J Bacteriol* **116**, 1436-46.
- Rao, V. B. and Black, L. W. (2005). DNA Packaging in Bacteriophage T4. Georgetown: Landes Bioscience.
- Raquet, X., Eckert, J. H., Muller, S. and Johnsson, N. (2001). Detection of altered protein conformations in living cells. *J Mol Biol* **305**, 927-38.
- Rasouly, A., Schonbrun, M., Shenhar, Y. and Ron, E. Z. (2009). YbeY, a heat shock protein involved in translation in Escherichia coli. *J Bacteriol* **191**, 2649-55.
- Rebar, E. J. and Pabo, C. O. (1994). Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671-3.
- Reinders, J., Zahedi, R. P., Pfanner, N., Meisinger, C. and Sickmann, A. (2006). Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics. *J Proteome Res* **5**, 1543-54.
- Rhoades, M. M. (1943). Genic Induction of an Inherited Cytoplasmic Difference. *Proc Natl Acad Sci U S A* **29**, 327-9.
- Rice, P. A., Yang, S., Mizuuchi, K. and Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* **87**, 1295-306.
- Roberts, J. W. and Roberts, C. W. (1975). Proteolytic cleavage of bacteriophage lambda repressor in induction. *Proc Natl Acad Sci U S A* **72**, 147-51.
- Roberts, J. W., Roberts, C. W. and Craig, N. L. (1978). Escherichia coli recA gene product inactivates phage lambda repressor. *Proc Natl Acad Sci U S A* **75**, 4714-8.
- Roder, I. V., Choi, K. R., Reischl, M., Petersen, Y., Diefenbacher, M. E., Zaccolo, M., Pozzan, T. and Rudolf, R. (2010). Myosin Va cooperates with PKA RIalpha to mediate maintenance of the endplate in vivo. *Proc Natl Acad Sci U S A* **107**, 2031-6.
- Rodnina, M. V., Beringer, M. and Wintermeyer, W. (2007). How ribosomes make peptide bonds. *Trends Biochem Sci* **32**, 20-6.

- Romero, A., Lopez, R., Lurz, R. and Garcia, P. (1990). Temperate bacteriophages of *Streptococcus pneumoniae* that contain protein covalently linked to the 5' ends of their DNA. *J Virol* **64**, 5149-55.
- Romero, P., Croucher, N. J., Hiller, N. L., Hu, F. Z., Ehrlich, G. D., Bentley, S. D., Garcia, E. and Mitchell, T. J. (2009a). Comparative genomic analysis of ten *Streptococcus pneumoniae* temperate bacteriophages. *J Bacteriol* **191**, 4854-62.
- Romero, P., Garcia, E. and Mitchell, T. J. (2009b). Development of a prophage typing system and analysis of prophage carriage in *Streptococcus pneumoniae*. *Appl Environ Microbiol* **75**, 1642-9.
- Romero, P., Lopez, R. and Garcia, E. (2004). Genomic organization and molecular analysis of the inducible prophage EJ-1, a mosaic myovirus from an atypical pneumococcus. *Virology* **322**, 239-52.
- Ronda, C., Garcia, J. L. and Lopez, R. (1989). Infection of *Streptococcus oralis* NCTC 11427 by pneumococcal phages. *FEMS Microbiol Lett* **53**, 187-92.
- Ronda, C., Lopez, R. and Garcia, E. (1981). Isolation and characterization of a new bacteriophage, Cp-1, infecting *Streptococcus pneumoniae*. *J Virol* **40**, 551-9.
- Rossmann, M. G., Mesyanzhinov, V. V., Arisaka, F. and Leiman, P. G. (2004). The bacteriophage T4 DNA injection machine. *Curr Opin Struct Biol* **14**, 171-80.
- Roucourt, B., Lecoutere, E., Chibeu, A., Hertveldt, K., Volckaert, G. and Lavigne, R. (2009). A procedure for systematic identification of bacteriophage-host interactions of *P. aeruginosa* phages. *Virology* **387**, 50-8.
- Roy, A., Kucukural, A. and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-38.
- Roy, M. K., Singh, B., Ray, B. K. and Apirion, D. (1983). Maturation of 5-S rRNA: ribonuclease E cleavages and their dependence on precursor sequences. *Eur J Biochem* **131**, 119-27.
- Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N. et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173-8.
- Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M. and Smith, M. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265**, 687-95.
- Sanger, F. and Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**, 441-8.
- Sao-Jose, C., Baptista, C. and Santos, M. A. (2004). *Bacillus subtilis* operon encoding a membrane receptor for bacteriophage SPP1. *J Bacteriol* **186**, 8337-46.
- Schattner, P., Brooks, A. N. and Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* **33**, W686-9.
- Schauer, A. T., Carver, D. L., Bigelow, B., Baron, L. S. and Friedman, D. I. (1987). lambda N antitermination system: functional analysis of phage interactions with the host NusA protein. *J Mol Biol* **194**, 679-90.
- Scherberg, N. H. and Weiss, S. B. (1972). T4 transfer RNAs: codon recognition and translational properties. *Proc Natl Acad Sci U S A* **69**, 1114-8.
- Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M. and Cate, J. H. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827-34.
- Schwartz, H., Alvares, C. P., White, M. B. and Fields, S. (1998). Mutation detection by a two-hybrid assay. *Hum Mol Genet* **7**, 1029-32.
- SenGupta, D. J., Zhang, B., Kraemer, B., Pochart, P., Fields, S. and Wickens, M. (1996). A three-hybrid system to detect RNA-protein interactions in vivo. *Proc Natl Acad Sci U S A* **93**, 8496-501.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-504.
- Sharma, R. C. and Smith, K. C. (1987). Role of DNA polymerase I in postreplication repair: a reexamination with *Escherichia coli* delta polA. *J Bacteriol* **169**, 4559-64.
- Sheehan, M. M., Garcia, J. L., Lopez, R. and Garcia, P. (1997). The lytic enzyme of the pneumococcal phage Dp-1: a chimeric lysin of intergeneric origin. *Mol Microbiol* **25**, 717-25.
- Shinagawa, H., Mizuuchi, K. and Emmerson, P. T. (1977). Induction of prophage lambda by gamma-rays, mitomycin C and tif; repressor cleavage studied by immunoprecipitation. *Mol Gen Genet* **155**, 87-91.
- Shlomai, J. and Kornberg, A. (1978). Deoxyuridine triphosphatase of *Escherichia coli*. Purification, properties, and use as a reagent to reduce uracil incorporation into DNA. *J Biol Chem* **253**, 3305-12.
- Shumway, L. K. and Weier, T. E. (1967). The Chloroplast Structure of Iojap Maize. *American Journal of Botany* **54**, 773-780.
- Sinclair, J. F., Tzagoloff, A., Levine, D. and Mindich, L. (1975). Proteins of bacteriophage phi6. *J Virol* **16**, 685-95.

- Sinha, H., David, L., Pascon, R. C., Clauder-Munster, S., Krishnakumar, S., Nguyen, M., Shi, G., Dean, J., Davis, R. W., Oefner, P. J. et al. (2008). Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. *Genetics* **180**, 1661-70.
- Skordalakes, E. and Berger, J. M. (2003). Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* **114**, 135-46.
- Sonenberg, N. and Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-45.
- Song, J. H., Ko, K. S., Lee, J. Y., Baek, J. Y., Oh, W. S., Yoon, H. S., Jeong, J. Y. and Chun, J. (2005). Identification of essential genes in *Streptococcus pneumoniae* by allelic replacement mutagenesis. *Mol Cells* **19**, 365-74.
- Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S. et al. (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell* **122**, 957-68.
- Studier, F. W. and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-30.
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185**, 60-89.
- Sukhodolets, M. V. and Garges, S. (2003). Interaction of *Escherichia coli* RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. *Biochemistry* **42**, 8022-34.
- Summers, W. C. (1993). Plague and cholera in India: the bacteriophage inquiry of 1928–1936. *J. Hist. Med. All. Sci.* **48**, 275-301.
- Sun, L. and Fuchs, J. A. (1992). *Escherichia coli* ribonucleotide reductase expression is cell cycle regulated. *Mol Biol Cell* **3**, 1095-105.
- Sun, T. P. and Webster, R. E. (1986). *fii*, a bacterial locus required for filamentous phage infection and its relation to colicin-tolerant *tolA* and *tolB*. *J Bacteriol* **165**, 107-15.
- Suzuki, N. N., Koizumi, K., Fukushima, M., Matsuda, A. and Inagaki, F. (2004). Structural basis for the specificity, catalysis, and regulation of human uridine-cytidine kinase. *Structure* **12**, 751-64.
- Tabor, S., Huber, H. E. and Richardson, C. C. (1987). *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *J Biol Chem* **262**, 16212-23.
- Tao, Y., Olson, N. H., Xu, W., Anderson, D. L., Rossmann, M. G. and Baker, T. S. (1998). Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* **95**, 431-7.
- Tepljakov, A., Obmolova, G., Chu, S. Y., Toedt, J., Eisenstein, E., Howard, A. J. and Gilliland, G. L. (2003). Crystal structure of the YchF protein reveals binding sites for GTP and nucleic acid. *J Bacteriol* **185**, 4031-7.
- Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J. et al. (2001). Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498-506.
- Thanassi, J. A., Hartman-Neumann, S. L., Dougherty, T. J., Dougherty, B. A. and Pucci, M. J. (2002). Identification of 113 conserved essential genes using a high-throughput gene disruption system in *Streptococcus pneumoniae*. *Nucleic Acids Res* **30**, 3152-62.
- Thompson, D., Walbot, V. and Coe, E. H. (1983). Plastid Development in Iojap- and Chloroplast Mutator-Affected Maize Plants. *American Journal of Botany* **70**, 940-950.
- Thunell, R. K. and Sandine, W. E. (1981). Phage-Insensitive, Multiple-Strain Starter Approach to Cheddar Cheese Making. *Journal of Dairy Science* **64**, 2270-2277.
- Tiraby, J. G., Tiraby, E. and Fox, M. S. (1975). Pneumococcal bacteriophages. *Virology* **68**, 566-9.
- Titz, B., Hauser, R., Engelbrecher, A. and Uetz, P. (2007). The *Escherichia coli* protein YjjG is a housecleaning nucleotidase in vivo. *FEMS Microbiol Lett* **270**, 49-57.
- Titz, B., Rajagopala, S. V., Ester, C., Hauser, R. and Uetz, P. (2006a). Novel conserved assembly factor of the bacterial flagellum. *J Bacteriol* **188**, 7700-6.
- Titz, B., Rajagopala, S. V., Goll, J., Hauser, R., McKevitt, M. T., Palzkill, T. and Uetz, P. (2008). The binary protein interactome of *Treponema pallidum*—the syphilis spirochete. *PLoS One* **3**, e2292.
- Titz, B., Thomas, S., Rajagopala, S. V., Chiba, T., Ito, T. and Uetz, P. (2006b). Transcriptional activators in yeast. *Nucleic Acids Res* **34**, 955-67.
- Torrents, E., Grinberg, I., Gorovitz-Harris, B., Lundstrom, H., Borovok, I., Aharonowitz, Y., Sjoberg, B. M. and Cohen, G. (2007). NrdR controls differential expression of the *Escherichia coli* ribonucleotide reductase genes. *J Bacteriol* **189**, 5012-21.
- Torres, M., Condon, C., Balada, J. M., Squires, C. and Squires, C. L. (2001). Ribosomal protein S4 is a transcription factor with properties remarkably similar to NusA, a protein involved in both non-ribosomal and ribosomal RNA antitermination. *EMBO J* **20**, 3811-20.

- Tovchigrechko, A. and Vakser, I. A.** (2006). GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res* **34**, W310-4.
- Uchiyama, I., Higuchi, T. and Kawai, M.** (2010). MGD update 2010: toward a comprehensive resource for exploring microbial genome diversity. *Nucleic Acids Res* **38**, D361-5.
- Uetz, P., Dong, Y. A., Zeretzke, C., Atzler, C., Baiker, A., Berger, B., Rajagopala, S. V., Roupelieva, M., Rose, D., Fossum, E. et al.** (2006). Herpesviral protein networks and their interaction with the human proteome. *Science* **311**, 239-42.
- Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P. et al.** (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623-7.
- van Opijnen, T., Bodi, K. L. and Camilli, A.** (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6**, 767-72.
- van Wezenbeek, P. M., Hulsebos, T. J. and Schoenmakers, J. G.** (1980). Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. *Gene* **11**, 129-48.
- Venkatesan, K., Rual, J. F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K. I. et al.** (2009). An empirical framework for binary interactome mapping. *Nat Methods* **6**, 83-90.
- Venter, J. C. Adams, M. D. Myers, E. W. Li, P. W. Mural, R. J. Sutton, G. G. Smith, H. O. Yandell, M. Evans, C. A. Holt, R. A. et al.** (2001). The sequence of the human genome. *Science* **291**, 1304-51.
- Vidal, M. and Endoh, H.** (1999). Prospects for drug screening using the reverse two-hybrid system. *Trends Biotechnol* **17**, 374-81.
- Vidal, M. and Legrain, P.** (1999). Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res* **27**, 919-29.
- Vidaver, A. K., Koski, R. K. and Van Etten, J. L.** (1973). Bacteriophage phi6: a Lipid-Containing Virus of *Pseudomonas phaseolicola*. *J Virol* **11**, 799-805.
- Vizoso Pinto, M. G., Villegas, J. M., Peter, J., Haase, R., Haas, J., Lotz, A. S., Muntau, A. C. and Baiker, A.** (2009). LuMPIS--a modified luminescence-based mammalian interactome mapping pull-down assay for the investigation of protein-protein interactions encoded by GC-low ORFs. *Proteomics* **9**, 5303-8.
- Vollert, C. S. and Uetz, P.** (2004). The phox homology (PX) domain protein interaction network in yeast. *Mol Cell Proteomics* **3**, 1053-64.
- von Mering, C., Jensen, L. J., Kuhn, M., Chaffron, S., Doerks, T., Kruger, B., Snel, B. and Bork, P.** (2007). STRING 7--recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* **35**, D358-62.
- Wagner, P. L., Acheson, D. W. and Waldor, M. K.** (1999). Isogenic lysogens of diverse shiga toxin 2-encoding bacteriophages produce markedly different amounts of shiga toxin. *Infect Immun* **67**, 6710-4.
- Wagner, P. L., Livny, J., Neely, M. N., Acheson, D. W., Friedman, D. I. and Waldor, M. K.** (2002). Bacteriophage control of Shiga toxin 1 production and release by *Escherichia coli*. *Mol Microbiol* **44**, 957-70.
- Wagner, P. L. and Waldor, M. K.** (2002). Bacteriophage control of bacterial virulence. *Infect Immun* **70**, 3985-93.
- Waldor, M. K. and Mekalanos, J. J.** (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910-4.
- Wang, I. N., Smith, D. L. and Young, R.** (2000). Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol* **54**, 799-825.
- Wang, J., Jiang, Y., Vincent, M., Sun, Y., Yu, H., Bao, Q., Kong, H. and Hu, S.** (2005). Complete genome sequence of bacteriophage T5. *Virology* **332**, 45-65.
- Wanner, B. L., Kodaira, R. and Neidhardt, F. C.** (1977). Physiological regulation of a decontrolled lac operon. *J Bacteriol* **130**, 212-22.
- Washburn, R. S., Marra, A., Bryant, A. P., Rosenberg, M. and Gentry, D. R.** (2001). rho is not essential for viability or virulence in *Staphylococcus aureus*. *Antimicrob Agents Chemother* **45**, 1099-103.
- Weigel, C. and Seitz, H.** (2006). Bacteriophage replication modules. *FEMS Microbiol Rev* **30**, 321-81.
- Weiss, B. and Richardson, C. C.** (1967). Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from *Escherichia coli* infected with T4 bacteriophage. *Proc Natl Acad Sci U S A* **57**, 1021-8.
- Whitman, W. B., Coleman, D. C. and Wiebe, W. J.** (1998). Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95**, 6578-83.
- Wilhelm, S. W. and Suttle, C. A.** (1999). Viruses and Nutrient Cycles in the Sea. *BioScience* **49**, 781-788.
- Williamson, J. R.** (2003). After the ribosome structures: how are the subunits assembled? *RNA* **9**, 165-7.
- Wilson, W. A., Wang, Z. and Roach, P. J.** (2002). Systematic identification of the genes affecting glycogen storage in the yeast *Saccharomyces cerevisiae*: implication of the vacuole as a determinant of glycogen level. *Mol Cell Proteomics* **1**, 232-42.

- Witzenrath, M., Schmeck, B., Doehn, J. M., Tschernig, T., Zahlten, J., Loeffler, J. M., Zemlin, M., Muller, H., Gutbier, B., Schutte, H. et al. (2009). Systemic use of the endolysin Cpl-1 rescues mice with fatal pneumococcal pneumonia. *Crit Care Med* **37**, 642-9.
- Wommack, K. E. and Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**, 69-114.
- Xiang, Y., Morais, M. C., Battisti, A. J., Grimes, S., Jardine, P. J., Anderson, D. L. and Rossmann, M. G. (2006). Structural changes of bacteriophage phi29 upon DNA packaging and release. *EMBO J* **25**, 5229-39.
- Xin, X., Rual, J. F., Hirozane-Kishikawa, T., Hill, D. E., Vidal, M., Boone, C. and Thierry-Mieg, N. (2009). Shifted Transversal Design smart-pooling for high coverage interactome mapping. *Genome Res* **19**, 1262-9.
- Yang, Q. and Catalano, C. E. (2003). Biochemical characterization of bacteriophage lambda genome packaging in vitro. *Virology* **305**, 276-87.
- Yoon, J. R., Laible, P. D., Gu, M., Scott, H. N. and Collart, F. R. (2002). Express primer tool for high-throughput gene cloning and expression. *Biotechniques* **33**, 1328-33.
- Yu, F. and Mizushima, S. (1982). Roles of lipopolysaccharide and outer membrane protein OmpC of Escherichia coli K-12 in the receptor function for bacteriophage T4. *J Bacteriol* **151**, 718-22.
- Yu, H., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N. et al. (2008). High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104-10.
- Zachary, A. and Black, L. W. (1981). DNA ligase is required for encapsidation of bacteriophage T4 DNA. *J Mol Biol* **149**, 641-58.
- Zalacain, M., Biswas, S., Ingraham, K. A., Ambrad, J., Bryant, A., Chalker, A. F., Iordanescu, S., Fan, J., Fan, F., Lunsford, R. D. et al. (2003). A global approach to identify novel broad-spectrum antibacterial targets among proteins of unknown function. *J Mol Microbiol Biotechnol* **6**, 109-26.
- Zeng, X., Barros, M. H., Shulman, T. and Tzagoloff, A. (2008). ATP25, a new nuclear gene of Saccharomyces cerevisiae required for expression and assembly of the Atp9p subunit of mitochondrial ATPase. *Mol Biol Cell* **19**, 1366-77.
- Zheng, Y., Struck, D. K., Bernhardt, T. G. and Young, R. (2008). Genetic analysis of MraY inhibition by the phiX174 protein E. *Genetics* **180**, 1459-66.
- Zhong, J., Zhang, H., Stanyon, C. A., Tromp, G. and Finley, R. L., Jr. (2003). A strategy for constructing large protein interaction maps using the yeast two-hybrid system: regulated expression arrays and two-phase mating. *Genome Res* **13**, 2691-9.
- Zubko, M. K. and Day, A. (1998). Stable albinism induced without mutagenesis: a model for ribosome-free plastid inheritance. *Plant J* **15**, 265-71.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-15.

Supplementary information

Supplementary figures

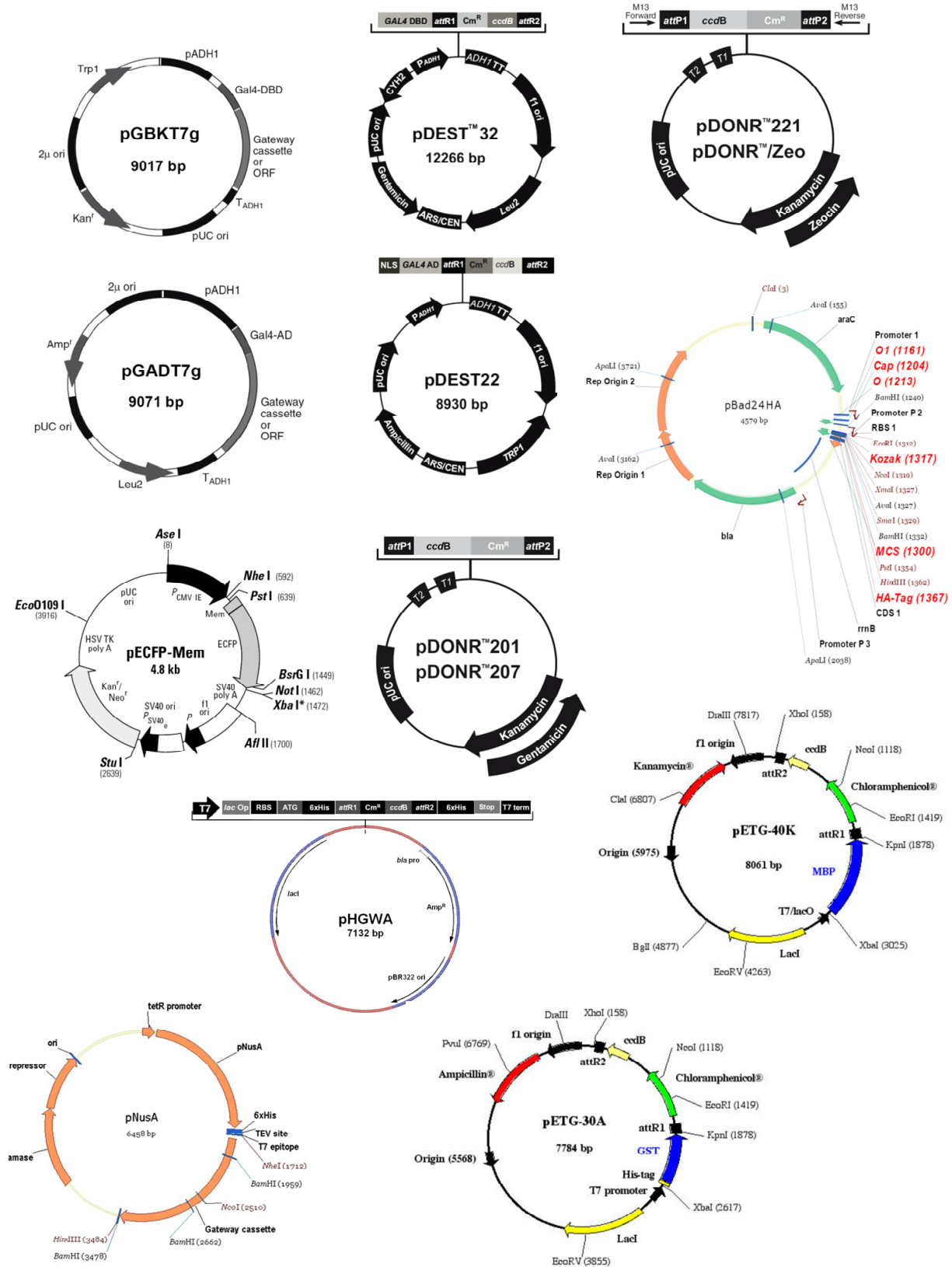


Fig. 49 Vector maps
References are given under Tab. 16.

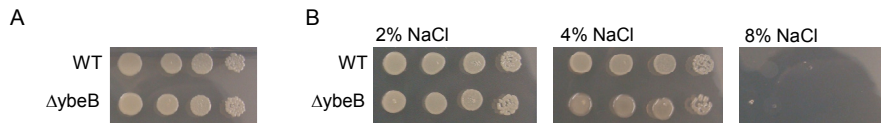


Fig. 50 General stress conditions tested for *ybeB* KO strain by agar plate assays
(A) Growth under high-temperature growth conditions (42°C). **(B)** High-salt condition including 2, 4, and 8% (w/v) sodium chloride in LB agar medium. Plates were grown at 37 °C. Pictures show serially diluted cells that were grown o/n at the indicated temperatures.

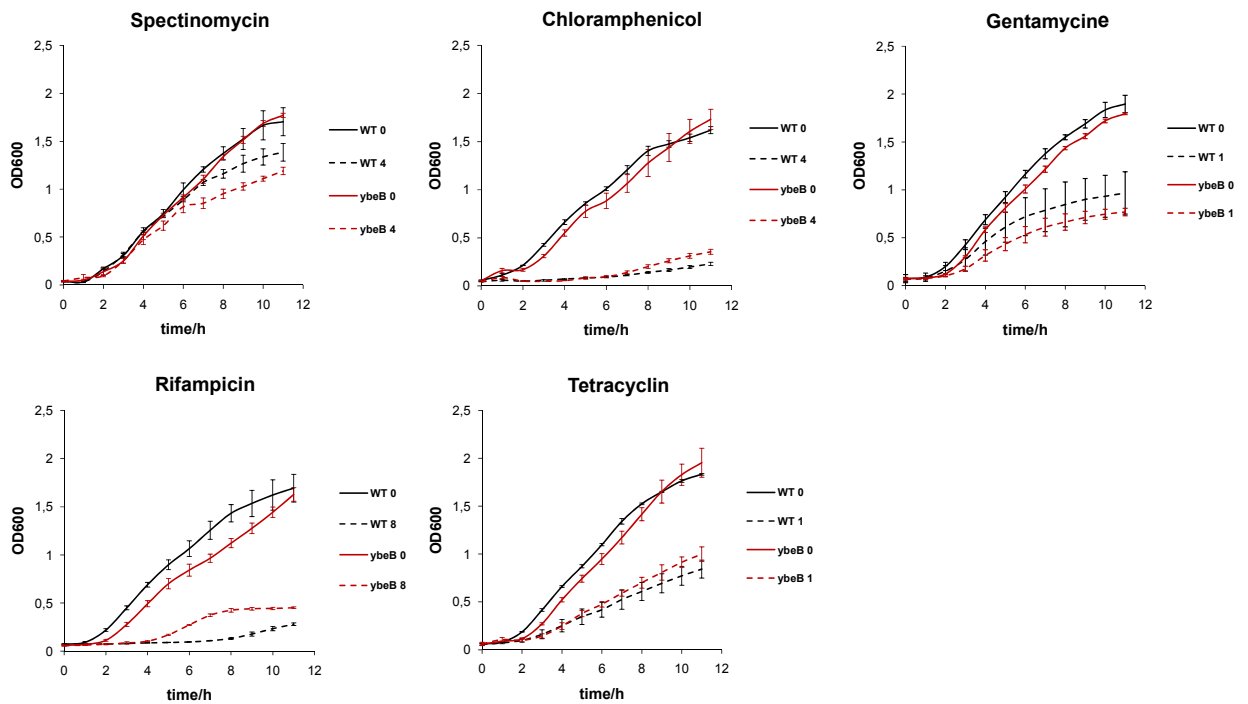
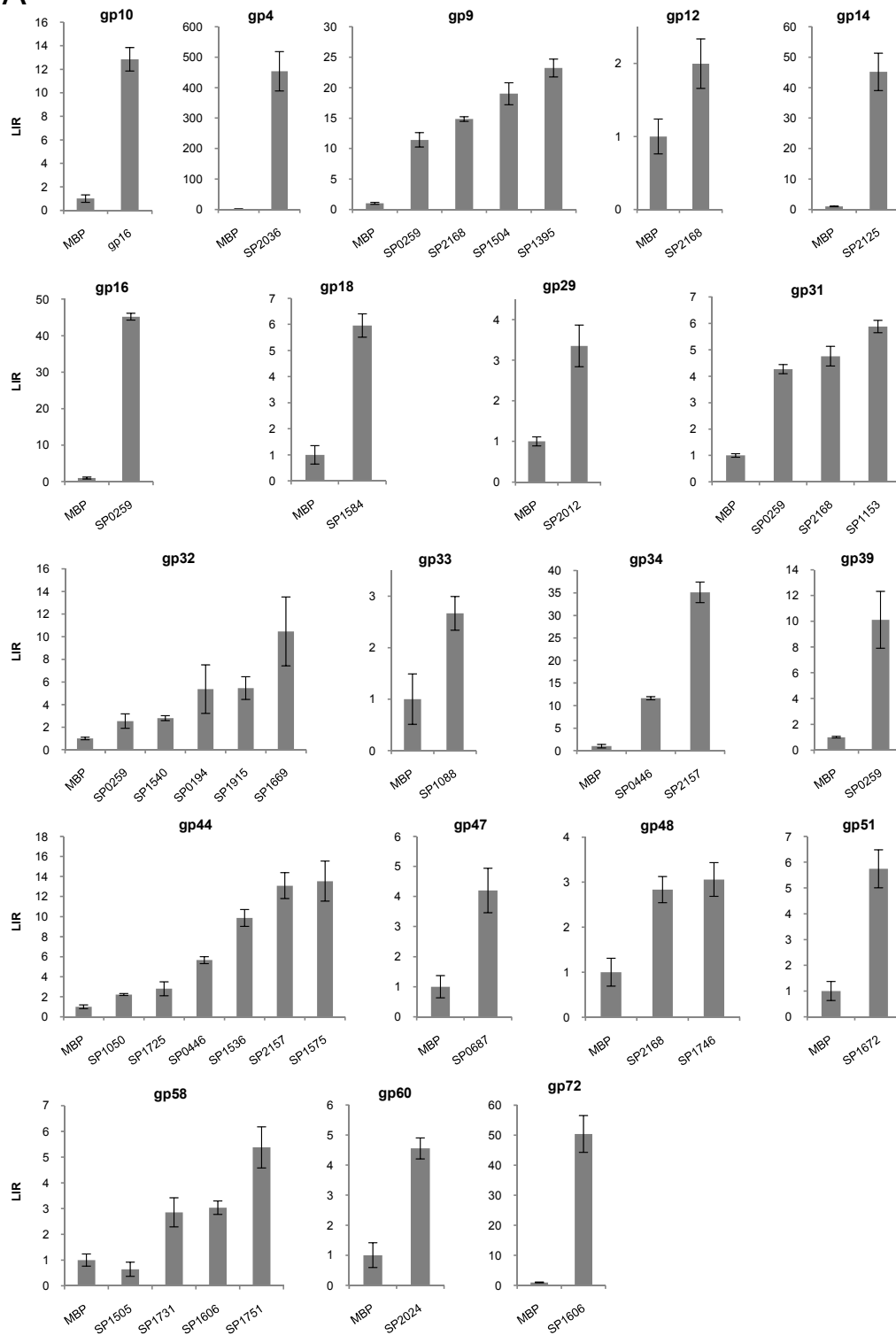


Fig. 51 Phenotyping of *E. coli* $\Delta ybeB$ by growth curve assays – test on chemical compounds that interfere with translation

Given are the control growth curves (no additives) and the curves when chemical compounds were added. Here only the corresponding curves for corresponding concentration (in μM , see number in legends) are represented that had a reasonable effect on bacterial growth. Note, that rifampicin blocks transcription.

SUPPLEMENTARY INFORMATION

A



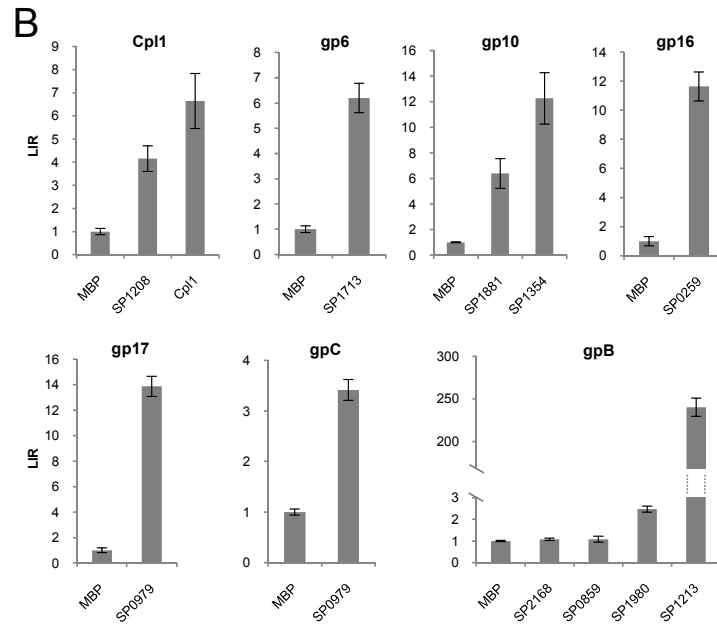


Fig. 52 Verification of phage-host PPIs by LuMPIS

(A) All tests for Dp-1-host PPIs and (B) for Cp-1 identified by the Y2H screens. Single histograms represent the LIR mean values for single GFP-luc tagged phage prey proteins tested against as single or a set of MBP-tagged host prey proteins. Tests were done in quadruplicates and the corresponding standard deviation of the mean is indicated by error bars. MBP is the negative control (MBP w/o a tagged ORF).

Supplementary tables

Tab. 35 Intra-viral PPIs of Dp-1

The table represents a non-redundant list of Dp-1 intra-viral PPIs that were reproducible in Y2H retest experiments. DBD-/AD- (bait/prey) interaction direction and the vector system the PPIs were detected with are combined. (VS) indicates the positive tested vector system ((D) pDEST32/22 or (G) pGBKT7g/pGADT7g). The 3-AT score represents the concentration range of 3-AT in the readout medium where yeast growth was visible compared to self-activation background growth of this bait. A “+” indicates that yeast grew still on the highest tested concentration (50 mM 3-AT).

Protein A	Protein B	VS	3-AT score/mM
Gp10 DNA polymerase III beta subunit DnaN, putative	Gp10 DNA polymerase III beta subunit DnaN, putative	D	0.5
Gp10 DNA polymerase III beta subunit DnaN, putative	Gp15 Recombination protein RecA	D	50 ⁺
Gp10 DNA polymerase III beta subunit DnaN, putative	Gp16 NAD-dependent DNA ligase, putative	D G D	50 ⁺ 50 ⁺ 50 ⁺
Gp10 DNA polymerase III beta subunit DnaN, putative	Gp23 Terminase-like protein	D	2
Gp10 DNA polymerase III beta subunit DnaN, putative	Gp71 DNA polymerase I	D G D	2.5 50 ⁺ 50 ⁺
Gp11 Cas4 RecB like exonuclease, putative	Gp11 Cas4 RecB like exonuclease, putative	D	2.5
Gp44 Rho-like domain lipoprotein, putative	Gp11 Cas4 RecB like exonuclease, putative	D	2.25
Gp11 Cas4 RecB like exonuclease, putative	Gp69 Sporulation sigma factor SigK, putative	G	0
Gp13 Hypothetical protein	Gp13 Hypothetical protein	D	5
Gp18 DNA polymerase III, delta' subunit HolB, putative	Gp13 Hypothetical protein	D	0.1
Gp31 Hypothetical protein	Gp13 Hypothetical protein	D	2.5
Gp32 Hypothetical protein	Gp13 Hypothetical protein	D	7.5
Gp33 Hypothetical protein	Gp13 Hypothetical protein	D	0.1
Gp45 Hypothetical protein	Gp13 Hypothetical protein	D	0
Gp53 Tail protein, putative	Gp13 Hypothetical protein	D	0
Gp68 DNA primase DnaG	Gp13 Hypothetical protein	D	47.5 ⁺
Gp13 Hypothetical protein	Gp69 Sporulation sigma factor SigK, putative	G	0.1
Gp13 Hypothetical protein	Gp70 Hypothetical protein	D D	2.5 25
Gp8 No similarity	Gp13 Hypothetical protein	D	0.25
Gp9 No similarity	Gp13 Hypothetical protein	D	2.5
Gp15 Recombination protein RecA	Gp25 Hypothetical protein	D G	50 ⁺ 50 ⁺
Gp15 Recombination protein RecA	Gp31 Hypothetical protein	G	0.1
Gp21 Metal-dependent phosphohydrolase HD, putative	Gp16 NAD-dependent DNA ligase, putative	G	2.5
Gp16 NAD-dependent DNA ligase, putative	Gp23 Terminase-like protein	D	50 ⁺
Gp32 Hypothetical protein	Gp16 NAD-dependent DNA ligase, putative	G	1
Gp70 Hypothetical protein	Gp16 NAD-dependent DNA ligase, putative	G	2.5
Gp17 DNA polymerase III gamma/tau subunit DnaX	Gp17 DNA polymerase III gamma/tau subunit DnaX	D G	10 49.75 ⁺
Gp17 DNA polymerase III gamma/tau subunit DnaX	Gp18 DNA polymerase III, delta' subunit HolB, putative	D G D	25 49.75 ⁺ 50 ⁺
Gp19 DNA polymerase III delta subunit HolA, putative	Gp17 DNA polymerase III gamma/tau subunit DnaX	D	5

SUPPLEMENTARY INFORMATION

Protein A	Protein B	VS	3-AT score/mM
Gp17 DNA polymerase III gamma/tau subunit DnaX	Gp31 Hypothetical protein	G	9.75
Gp17 DNA polymerase III gamma/tau subunit DnaX	Gp69 Sporulation sigma factor SigK, putative	G	2.25
Gp17 DNA polymerase III gamma/tau subunit DnaX	Gp70 Hypothetical protein	G	2.25
Gp18 DNA polymerase III, delta' subunit HolB, putative	Gp27 Hypothetical protein	D	2.5
Gp18 DNA polymerase III, delta' subunit HolB, putative	Gp69 Sporulation sigma factor SigK, putative	G	9.75
Gp1 Queuosine biosynthesis protein QueF	Gp1 Queuosine biosynthesis protein QueF	D	50 ⁺
Gp44 Rho-like domain lipoprotein, putative	Gp1 Queuosine biosynthesis protein QueF	D	0.75
Gp21 Metal-dependent phosphohydrolase HD, putative	Gp21 Metal-dependent phosphohydrolase HD, putative	D	1
Gp21 Metal-dependent phosphohydrolase HD, putative	Gp61 Superfamily II DNA/RNA helicases, SNF2 family	G	2.5
Gp22 Hypothetical protein	Gp23 Terminase-like protein	G	9.75
Gp22 Hypothetical protein	Gp41 Minor capsid protein	G	2.25
Gp22 Hypothetical protein	Gp71 DNA polymerase I	G	2.25
Gp28 Hypothetical protein	Gp23 Terminase-like protein	G	0.1
Gp33 Hypothetical protein	Gp23 Terminase-like protein	G	0.1
Gp39 Zinc finger domain protein, putative	Gp23 Terminase-like protein	G	0.25
Gp43 Major capsid protein	Gp23 Terminase-like protein	G	2.5
Gp48 Hypothetical protein	Gp23 Terminase-like protein	G	4.75
Gp49 Listeria-Bacteroides repeat domain protein	Gp23 Terminase-like protein	G	4.75
Gp53 Tail protein, putative	Gp23 Terminase-like protein	G	0.25
Gp71 DNA polymerase I	Gp23 Terminase-like protein	D	50 ⁺
Gp24 Sigma factor (region 4), putative	Gp24 Sigma factor (region 4), putative	D G	2.5 1
Gp24 Sigma factor (region 4), putative	Gp31 Hypothetical protein	G	0.1
Gp24 Sigma factor (region 4), putative	Gp69 Sporulation sigma factor SigK, putative	G	1
Gp24 Sigma factor (region 4), putative	Gp70 Hypothetical protein	G	1
Gp25 Hypothetical protein	Gp25 Hypothetical protein	D	25
Gp32 Hypothetical protein	Gp25 Hypothetical protein	G	1
Gp31 Hypothetical protein	Gp27 Hypothetical protein	D	2.5
Gp32 Hypothetical protein	Gp27 Hypothetical protein	D	24
Gp34 Hypothetical protein	Gp27 Hypothetical protein	D	0.75
Gp39 Zinc finger domain protein, putative	Gp27 Hypothetical protein	D	1
Gp44 Rho-like domain lipoprotein, putative	Gp27 Hypothetical protein	D	2.25
Gp48 Hypothetical protein	Gp27 Hypothetical protein	G	2.25
Gp27 Hypothetical protein	Gp69 Sporulation sigma factor SigK, putative	G	0.25
Gp9 No similarity	Gp27 Hypothetical protein	D	0.25
Gp28 Hypothetical protein	Gp69 Sporulation sigma factor SigK, putative	G	0.1
Gp28 Hypothetical protein	Gp7 Hypothetical protein	G	0.1
Gp28 Hypothetical protein	Gp8 No similarity	G	0.1

SUPPLEMENTARY INFORMATION

Protein A	Protein B	VS	3-AT score/mM
<i>Gp9</i> No similarity	<i>Gp28</i> Hypothetical protein	D	0.25
<i>Gp29</i> Hypothetical protein	<i>Gp29</i> Hypothetical protein	D	1
<i>Gp2</i> Queuosine biosynthesis protein QueC	<i>Gp2</i> Queuosine biosynthesis protein QueC	D	47.5 ⁺
<i>Gp2</i> Queuosine biosynthesis protein QueC	<i>Gp27</i> Hypothetical protein	D	47.5 ⁺
<i>Gp31</i> Hypothetical protein	<i>Gp31</i> Hypothetical protein	D G	0.25 0.25
<i>Gp33</i> Hypothetical protein	<i>Gp31</i> Hypothetical protein	G	0.25
<i>Gp39</i> Zinc finger domain protein, putative	<i>Gp31</i> Hypothetical protein	G	1
<i>Gp31</i> Hypothetical protein	<i>Gp40</i> Minor capsid protein, putative	G D	1 0.25
<i>Gp48</i> Hypothetical protein	<i>Gp31</i> Hypothetical protein	G	4.75
<i>Gp53</i> Tail protein, putative	<i>Gp31</i> Hypothetical protein	G	0.5
<i>Gp54</i> Antireceptor	<i>Gp31</i> Hypothetical protein	G	4
<i>Gp59</i> Endolysin Pal	<i>Gp31</i> Hypothetical protein	G	2.25
<i>Gp64</i> Hypothetical protein	<i>Gp31</i> Hypothetical protein	G	1
<i>Gp68</i> DNA primase DnaG	<i>Gp31</i> Hypothetical protein	G	2.5
<i>Gp31</i> Hypothetical protein	<i>Gp69</i> Sporulation sigma factor SigK, putative	G	1
<i>Gp31</i> Hypothetical protein	<i>Gp70</i> Hypothetical protein	D G G	0.25 0.25 25
<i>Gp71</i> DNA polymerase I	<i>Gp31</i> Hypothetical protein	G	25
<i>Gp32</i> Hypothetical protein	<i>Gp47</i> Hypothetical protein	G	1
<i>Gp32</i> Hypothetical protein	<i>Gp68</i> DNA primase DnaG	G	2.5
<i>Gp32</i> Hypothetical protein	<i>Gp7</i> Hypothetical protein	G	2.5
<i>Gp32</i> Hypothetical protein	<i>Gp70</i> Hypothetical protein	G	1
<i>Gp32</i> Hypothetical protein	<i>Gp71</i> DNA polymerase I	G	1
<i>Gp32</i> Hypothetical protein	<i>Gp8</i> No similarity	G	1
<i>Gp32</i> Hypothetical protein	<i>Gp9</i> No similarity	G	1
<i>Gp33</i> Hypothetical protein	<i>Gp33</i> Hypothetical protein	G	0.25
<i>Gp33</i> Hypothetical protein	<i>Gp4</i> Queuosine biosynthesis protein QueE	D	1
<i>Gp33</i> Hypothetical protein	<i>Gp39</i> Zinc finger domain protein, putative	G	2.5
<i>Gp33</i> Hypothetical protein	<i>Gp44</i> Rho-like domain lipoprotein, putative	G	50 ⁺
<i>Gp48</i> Hypothetical protein	<i>Gp33</i> Hypothetical protein	G	0.75
<i>Gp53</i> Tail protein, putative	<i>Gp33</i> Hypothetical protein	G	0.25
<i>Gp33</i> Hypothetical protein	<i>Gp69</i> Sporulation sigma factor SigK, putative	G	1
<i>Gp36</i> Hypothetical protein	<i>Gp36</i> Hypothetical protein	D	49 ⁺
<i>Gp37</i> Terminase, large subunit	<i>Gp36</i> Hypothetical protein	D	50 ⁺
<i>Gp39</i> Zinc finger domain protein, putative	<i>Gp36</i> Hypothetical protein	D	25
<i>Gp63</i> Resolvase domain protein, putative	<i>Gp36</i> Hypothetical protein	D G	50 ⁺ 25
<i>Gp33</i> Hypothetical protein	<i>Gp3</i> Queuosine biosynthesis protein QueD	D	1
<i>Gp4</i> Queuosine biosynthesis protein QueE	<i>Gp3</i> Queuosine biosynthesis protein QueD	D	50 ⁺
<i>Gp39</i> Zinc finger domain protein, putative	<i>Gp40</i> Minor capsid protein, putative	G	10
<i>Gp39</i> Zinc finger domain protein, putative	<i>Gp43</i> Major capsid protein	G G	2.5 10
<i>Gp39</i> Zinc finger domain protein, putative	<i>Gp68</i> DNA primase DnaG	G	10
<i>Gp39</i> Zinc finger domain protein, putative	<i>Gp69</i> Sporulation sigma factor SigK, putative	G	1
<i>Gp39</i> Zinc finger domain protein, putative	<i>Gp7</i> Hypothetical protein	G	0.25
<i>Gp39</i> Zinc finger domain protein, putative	<i>Gp70</i> Hypothetical protein	G	1

SUPPLEMENTARY INFORMATION

Protein A	Protein B	VS	3-AT score/mM
Gp39 Zinc finger domain protein, putative	Gp8 No similarity	G	0.25
Gp43 Major capsid protein	Gp41 Minor capsid protein	D G	50 ⁺ 25
Gp42 Hypothetical protein	Gp42 Hypothetical protein	D	1
Gp43 Major capsid protein	Gp44 Rho-like domain lipoprotein, putative	D G	25 50 ⁺
Gp43 Major capsid protein	Gp69 Sporulation sigma factor SigK, putative	G	2.5
Gp43 Major capsid protein	Gp70 Hypothetical protein	G	1
Gp45 Hypothetical protein	Gp45 Hypothetical protein	D	0.5
Gp45 Hypothetical protein	Gp46 Hypothetical protein	G	2.25
Gp45 Hypothetical protein	Gp58 Holin	G	4.75
Gp45 Hypothetical protein	Gp69 Sporulation sigma factor SigK, putative	G	2.25
Gp45 Hypothetical protein	Gp70 Hypothetical protein	G	2.25
Gp48 Hypothetical protein	Gp46 Hypothetical protein	G	0.75
Gp48 Hypothetical protein	Gp69 Sporulation sigma factor SigK, putative	G	24.75
Gp48 Hypothetical protein	Gp70 Hypothetical protein	G	9.75
Gp4 Queuosine biosynthesis protein QueE	Gp4 Queuosine biosynthesis protein QueE	D	50 ⁺
Gp49 Listeria-Bacteroides repeat domain protein	Gp69 Sporulation sigma factor SigK, putative	G	24.75
Gp49 Listeria-Bacteroides repeat domain protein	Gp70 Hypothetical protein	G	9.75
Gp50 Hypothetical protein	Gp50 Hypothetical protein	D	50 ⁺
Gp53 Tail protein, putative	Gp51 Hypothetical protein	G	10
Gp55 Tail protein, putative	Gp51 Hypothetical protein	G	1.5
Gp53 Tail protein, putative	Gp69 Sporulation sigma factor SigK, putative	G	5
Gp53 Tail protein, putative	Gp70 Hypothetical protein	G	2.5
Gp54 Antireceptor	Gp55 Tail protein, putative	D G	1 49 ⁺
Gp54 Antireceptor	Gp69 Sporulation sigma factor SigK, putative	G	9
Gp54 Antireceptor	Gp70 Hypothetical protein	G	4
Gp55 Tail protein, putative	Gp55 Tail protein, putative	D G	0.5 24
Gp6 Queuosine biosynthesis intermediate transporter QueT	Gp55 Tail protein, putative	G	0.1
Gp56 Hypothetical protein	Gp56 Hypothetical protein	D G	50 ⁺ 25
Gp59 Endolysin Pal	Gp59 Endolysin Pal	D	50 ⁺
Gp59 Endolysin Pal	Gp69 Sporulation sigma factor SigK, putative	G	2.25
Gp59 Endolysin Pal	Gp70 Hypothetical protein	G	2.25
Gp60 Hypothetical protein	Gp64 Sporulation sigma factor SigK, putative	D D	25 0.25
Gp63 Resolvase domain protein, putative	Gp69 Sporulation sigma factor SigK, putative	G	0.25
Gp63 Resolvase domain protein, putative	Gp70 Hypothetical protein	G	0.1
Gp64 Hypothetical protein	Gp64 Hypothetical protein	D	5
Gp64 Hypothetical protein	Gp69 Sporulation sigma factor SigK, putative	G	0.5
Gp67 Replicative DNA helicase DnaB	Gp67 Replicative DNA helicase DnaB	D	1
Gp68 DNA primase DnaG	Gp68 DNA primase DnaG	G	25
Gp68 DNA primase DnaG	Gp69 Sporulation sigma factor SigK, putative	G	2.5
Gp70 Hypothetical protein	Gp68 DNA primase DnaG	G	1
Gp68 DNA primase DnaG	Gp71 DNA polymerase I	G	2.5

SUPPLEMENTARY INFORMATION

Protein A	Protein B	VS	3-AT score/mM
<i>Gp70</i> Hypothetical protein	<i>Gp69</i> Sporulation sigma factor SigK, putative	G	50
<i>Gp71</i> DNA polymerase I	<i>Gp69</i> Sporulation sigma factor SigK, putative	G	1
<i>Gp71</i> DNA polymerase I	<i>Gp70</i> Hypothetical protein	G	0.5
<i>Gp9</i> No similarity	<i>Gp70</i> Hypothetical protein	D	1
<i>Gp7</i> Hypothetical protein	<i>Gp7</i> Hypothetical protein	G	0.1
<i>Gp7</i> Hypothetical protein	<i>Gp8</i> No similarity	G	0.1

Tab. 36 All TP0738 (YbeB) interactions (non-redundant) identified in *T. pallidum* Y2H screens
 Refereces (1) this work, (2) (Titz et al., 2008) dataset pc50.

Locus tag	Description	Ref.	Locus tag	Description	Ref.
TP0024	conserved hypothetical protein	2	TP0519	response regulatory protein (atoC)	2
TP0048	conserved hypothetical protein	1, 2	TP0530	V-type ATPase, subunit E, putative	2
TP0060	ribosomal protein L9	1	TP0552	hypothetical protein	2
TP0067	conserved hypothetical protein	2	TP0554	phosphoglycolate phosphatase (gph)	1, 2
TP0078	spore coat polysaccharide biosynthesis protein (spsC)	2	TP0559	conserved hypothetical protein	1
TP0080	quinoline 2-oxidoreductase	2	TP0569	aminopeptidase P	2
TP0095	hypothetical protein	2	TP0587	hypothetical protein	2
TP0097	translation initiation factor 1 (infA)	2	TP0607	hypothetical protein	2
TP0113	Lambda CII stability-governing protein (hflK)	2	TP0618	hypothetical protein	2
TP0197	ribosomal protein L29 (rpmC)	1, 2	TP0626	exonuclease, putative	2
TP0199	ribosomal protein L14 (rplN)	1	TP0661	hypothetical protein	1
TP0233	anti-sigma F factor antagonist, putative	1, 2	TP0664	flagellar filament outer layer protein (flaA)	2
TP0247	N-acetylmuramoyl-L-alanine amidase (amiA)	2	TP0684	methylgalactoside ABC transporter, periplasmic galactose-binding protein	1
TP0255	ribosomal protein L31 (rpmE)	2	TP0704	single-stranded-DNA-specific exonuclease (recJ)	2
TP0257	glycerophosphodiester phosphodiesterase (glpQ)	2	TP0711	conserved hypothetical protein	2
TP0258	conserved hypothetical protein	1	TP0738	conserved hypothetical protein	1
TP0286	conserved hypothetical protein	2	TP0757	polypeptide deformylase (def)	1, 2
TP0297	conserved hypothetical protein	2	TP0773	periplasmic serine protease DO (htrA)	2
TP0334	conserved hypothetical protein	2	TP0807	ribosomal protein L32	1
TP0341	UDP-N-acetylmuramate--alanine ligase (murC)	2	TP0833	hypothetical protein	2
TP0354	thymidylate kinase (tmk)	2	TP0907	conserved hypothetical protein	1
TP0359	hypothetical protein	2	TP0943	flagellar protein (fliS)	1, 2
TP0375	hypothetical protein	2	TP0945	ribulose-phosphate 3-epimerase (cfxE)	1, 2
TP0393	smf protein (smf)	2	TP0946	glucose-inhibited division protein B (gidB)	2
TP0397	flagellar basal-body rod protein (flgC)	2	TP0965	membrane fusion protein, putative	2
TP0408	hypothetical protein	2	TP0974	hypothetical protein	1, 2
TP0412	conserved hypothetical protein	2	TP0981	sensory transduction histidine kinase, putative	2
TP0443	conserved hypothetical protein	2	TP0997	protease IV (sppA)	2
TP0445	4-methyl-5(b-hydroxyethyl)-thiazole monophosphate biosynthesis enzyme (thiJ)	2	TP1005	DNA polymerase III, subunits gamma and tau (dnaH)	2
TP0461	hypothetical protein	1	TP1019	glu-tRNA amidotransferase, subunit C	1
TP0465	hypothetical protein	2	TP1023	recX protein (recX)	1, 2

SUPPLEMENTARY INFORMATION

Tab. 37 All known protein-protein interactions of YbeB orthologs

Species	YbeB ortholog	Binding partner	Description interaction partner	Method	Reference
<i>E. coli</i>	b0637	50S subunit	large ribosomal subunit	pull down and iTRAQ	(Jiang et al., 2007)
		b3310	50S ribosomal subunit protein L14	Y2H, pull down	this work
		b1200	dihydroxyacetone kinase, N-terminal domain	LC-MS	(Butland et al., 2005)
		b2119	predicted transporter subunit: ATP-binding component of ABC superfamily		
		b2122	predicted protein		
		b2606	50S ribosomal subunit protein L19		
		b3056	fused tRNA nucleotidyl transferase -!- 2'3'-cyclic phosphodiesterase and 2'nucleotidase and phosphatase CCA		
		b3310	50S ribosomal subunit protein L14		
		b3319	50S ribosomal subunit protein L4		
		b3882	predicted oxidoreductase with NAD(P)-binding Rossmann-fold domain		
		b3986	50S ribosomal subunit protein L7/L12		
		b4066	conserved protein		
		<i>Treponema pallidum</i>	TP0738	TP0060	ribosomal protein L9
TP0097	translation initiation factor 1 InfA				
TP0197	ribosomal protein L29				
TP0199	ribosomal protein L14				
TP0255	ribosomal protein L31				
TP0559	Probable tRNA sulfurtransferase, tRNA 4-thiouridine synthase ThiI				
TP0757	peptide deformylase Def				
TP0807	50S ribosomal protein L32				
TP0907	Ribosome maturation factor RimM				
TP1019	Glutamyl-tRNA(Gln) amidotransferase subunit C GatC				
TP0530	V-type ATPase, subunit E, putative				
various others	various other interactions known (Tab. 36)				
<i>Campylobacter jejuni</i>	Cj1405			Cj1708c	30S ribosomal protein S10
		Cj1697c	50S ribosomal protein L14		
		Cj0670	RNA polymerase sigma-54 factor RpoN		
		Cj0307	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase BioA		
		Cj0542	Glutamyl-tRNA reductase HemA		
		Cj1454c	Ribosomal protein S12 methylthiotransferase RimO		
		Cj0389	Seryl-tRNA synthetase SerS		
<i>Helicobacter pylori</i>	HP1414	HP0187	Putative uncharacterized protein	Y2H	(Rain et al., 2001)
		HP0025	Outer membrane protein (Omp2)		
		HP0060	Putative uncharacterized protein		
		HP0205	Putative uncharacterized protein		
		HP0281	Queuine tRNA-ribosyltransferase tgt		
		HP0522	Cag pathogenicity island protein (Cag3)		
		HP0655	Protective surface antigen D15		
		HP1124	Putative uncharacterized protein		

SUPPLEMENTARY INFORMATION

Species	YbeB ortholog	Binding partner	Description interaction partner	Method	Reference
<i>S. pneumoniae</i> TIGR4	SP1744	SP0219	50S ribosomal protein L14	Y2H	this work
<i>Synechocystis</i> PCC 6803	slr1886	sll1806	50S ribosomal protein L14	Y2H	this work
<i>Homo sapiens</i>	C7orf30	MRPL14	39S ribosomal protein L14, mitochondrial	BiFC, pull down	this work
		UBE2V2	Enterocyte differentiation-associated factor EDAF-1, Enterocyte differentiation-promoting factor 1	Y2H	(Markson et al., 2009)
<i>Zea mays</i>	IJ (Iojap)	rpl14	50S ribosomal protein L14, chloroplastic	Pull down	this work
		50S subunit	50S chloroplast ribosomal subunit	IP	(Han and Martienssen, 1995)
<i>Saccharomyces cerevisiae</i>	YMR098C	54S subunit	Mitochondrial mitochondrial large subunit		(Gavin et al., 2006)
		YLR189C	UDP-glucose:sterol glucosyltransferase, conserved enzyme involved in synthesis of sterol glucoside membrane lipids	MALDI-TOF MS	(Krogan et al., 2006)
		YCR046C	54S ribosomal protein IMG1, mitochondrial IMG1; Pfam domain: Ribosomal L19		
		YGR220C	54S ribosomal protein L9, mitochondrial		
		Various others			MALDI-TOF MS; LC-MS/MS
<i>Drosophila melanogaster</i>	FBgn0029514	FBgn0036029	no description; contains DUF1042 signature; GO:0007498(NAS)=mesoderm development	Y2H	(Giot et al., 2003)

List of abbreviations

3-AT	3-Amino-1,2,4-triazole
AD	Activation domain
ADP	Adenosine diphosphate
APS	Ammonium persulfate
ATC	Anhydrotetracycline
ATP	Adenosine triphosphate
BCAA	Branched chain amino acids
bp	Base pair
BSA	Bovine serum albumin
CFP	Cyan fluorescent protein
CHP	Conserved hypothetical protein
Da	Dalton
DBD	DNA-binding domain
DCS	Donor calf serum
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
FCS	Fetal calf serum
FPLC	Fast protein liquid chromatography
GFP	Green fluorescent protein
GST	Glutathione S-transferase
GW	Gateway®
H ₂ O dd.	Water, double distilled
H ₂ O dist.	Water, distilled
HMM	Hidden Markov Model
HRP	Horse radish peroxidase
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITR	Inverted terminal repeats
kbp	Kilo base pairs
kDa	Kilo Dalton
KO	Knock-out
LB	Lysogeny broth
LIR	Luminescence intensity ratio
LPS	Lipopolysaccharide
LUMIER	Luminescence-based mammalian interactome mapping
LuMPIS	Luminescence-based MBP pull-down Interaction screening system
mAU	Milliabsorbance units
MBP	Maltose binding protein
MCS	Multiple cloning site
MHP	Major head protein
MWG	Molecular weight
OD	Optical density
ONPG	2-Nitrophenyl β -D-galactopyranoside
ORF	Open reading frame
ORFeome	Library collection of all open reading frames of a genome
ori	Origin of replication
P.A.	Pro analysis
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylenglykole
PMSF	Phenylmethylsulfonyl fluoride
PPI	Protein-protein interaction
PTS	(Phosphoenolpyruvate sugar) phosphotransferase system
Q	Queuosine
RBS	Ribosome binding site
R _f	Retention factor
RNR	Ribonucleotide reductase
rpm	Rotations per minute
RT	Reverse transcription/reverse transcriptase
SAP	Shrimp alkaline phosphatase

SDS	Sodium lauryl (dodecyl) sulfate
TCS	Two-component system
TEMED	N,N,N',N'-Tetramethylethylenediamine
TLC	Thin layer chromatography
T _m	Melting temperature
TSS	Transformation and storage solution
U	Enzyme unit
UMP	Uridine monophosphate
WT	Wild-type
X-Gal	5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside
Y2H	Yeast Two-Hybrid

List of figures

Fig. 1	Comprehensive, functional phage studies are underrepresented compared to phage genome sequencing projects	4
Fig. 2	Phage systematics	5
Fig. 3	Scheme of a lytic and temperate phage's reproduction cycle	7
Fig. 4	<i>Streptococcus pneumoniae</i> cell chains	9
Fig. 5	Cp-1 virion	12
Fig. 6	Genome map of Cp-1 and transcriptional organization	13
Fig. 7	EM micrographs of Dp-1 virions	14
Fig. 8	Lojap-affected maize	17
Fig. 9	A translating ribosome during peptidyl chain elongation	20
Fig. 10	The Y2H principle	23
Fig. 11	A matrix-based screen	25
Fig. 12	Principle of mini-pool screens	27
Fig. 13	Gateway® technology	41
Fig. 14	BP reaction	42
Fig. 15	LR reaction	43
Fig. 16	Illustration of mutagenic fusion PCRs of <i>E. coli</i> L14 mutant constructs	46
Fig. 17	Dp-1 genome map	80
Fig. 18	Homology-based Dp-1 protein annotation	81
Fig. 19	Sequence features of some Dp-1 proteins	83
Fig. 20	Q biosynthesis	90
Fig. 21	Comparison of the DNA polymerase III system of Dp-1 and its host	92
Fig. 22	Dp-1 PPI and net-work properties	96
Fig. 23	Intra-viral Dp-1 PPI network connects most of the Dp-1 proteins	97
Fig. 24	Cp-1 PPI properties	98
Fig. 25	Relationship of the Dp-1 genome context and the intra-viral PPI network	101
Fig. 26	Intra-viral PPI network of Cp-1 and relationship to its genomic context	102
Fig. 27	A screen and retest example	104
Fig. 28	Dp-1 and Cp-1 interactions with the host PPI network	106
Fig. 29	Comparison of host-phage PPI network properties	109
Fig. 30	LuMPIS verification of phage-host PPIs identified in the Y2H screens	112
Fig. 31	Domain interaction mapping of Cpl1-Udk PPI	114
Fig. 32	Effects of Cpl1 onto the quaternary structure of Udk	116
Fig. 33	Competitive Udk enzyme assay	117
Fig. 34	YbeB/DUF143 orthologs are widely distributed but absent in a few clades	120
Fig. 35	YbeB orthologs exhibit a low sequence conservation degree	123
Fig. 36	Mapping of conserved protein docking sites of YbeB with ribosome-related proteins	125
Fig. 37	Y2H cross-species interactions of YbeB and L14 orthologs	127
Fig. 38	<i>E. coli</i> , maize, and human YbeB-L14 pairs interact <i>in vitro</i>	128
Fig. 39	Human C7orf30 and L14 _{mt} co-localize <i>in vivo</i> into the mitochondrial compartment	129
Fig. 40	C7orf30 and L14 _{mt} interact <i>in vivo</i>	130
Fig. 41	Conserved amino acids mediate the interaction of L14 with YbeB	132

Fig. 42	The <i>E. coli ybeB</i> KO leads to reduced fitness.....	134
Fig. 43	The loss of YbeB results in a more rapid onset of protein translation.....	136
Fig. 44	YbeB interferes with 50S-30S assembly on the 3D-level.....	138
Fig. 45	A Dp-1 virion model model	146
Fig. 46	A Cp-1 virion PPI linkage map.....	148
Fig. 47	Putative effects on host gene expression mediated by phage-host PPIs	158
Fig. 48	Putative targets of Cp-1 and Dp-1 in the central host metabolism and its uptake systems.	159
Fig. 49	Vector maps.....	182
Fig. 50	General stress conditions tested for <i>ybeB</i> KO strain by agar plate assays.....	183
Fig. 51	Phenotyping of <i>E. coli ΔybeB</i> by growth curve assays – test on chemical compounds that interfere with translation.....	183
Fig. 52	Verification of phage-host PPIs by LuMPIS	185

List of tables

Tab. 1	Model phages.....	3
Tab. 2	Phage-derived tools and methods	3
Tab. 3	Bradley Classification of bacteriophages.....	6
Tab. 4	Formerly isolated <i>Pneumococcal</i> bacteriophages and some properties.....	11
Tab. 5	An updated annotation list of Cp-1 gene products.....	14
Tab. 6	Applications and use of two-hybrid assays besides protein-protein interaction screens ...	23
Tab. 7	Published proteome-scale and comprehensive Y2H projects	24
Tab. 8	Y2H pooling strategies	26
Tab. 9	Chemicals	31
Tab. 10	Enzymes and antibodies.....	33
Tab. 11	Instruments	34
Tab. 12	Kits and reagents.....	34
Tab. 13	Consumable materials.....	35
Tab. 14	Common primers	35
Tab. 15	Gene-specific primers.....	36
Tab. 16	Plasmids.....	37
Tab. 17	Bacterial strains	38
Tab. 18	Yeast strains and human cell lines.....	38
Tab. 19	Basic PCR templates for gene-specific amplification	38
Tab. 20	Plasmid selection conditions in <i>E. coli</i>	61
Tab. 21	Chemical components for phenotyping assays.....	62
Tab. 22	Fluorophores used in confocal microscopy	66
Tab. 23	Plasmid selection conditions in yeast	71
Tab. 24	Dp-1 ORFs and gene product annotation.....	84
Tab. 25	Predicted promoters	87
Tab. 26	Predicted Rho-independent termination sites	88
Tab. 27	Que synthesis enzymes of Dp-1.....	90
Tab. 28	Intra-viral PPIs of Cp-1.....	99
Tab. 29	Phage-host PPIs detected by Y2H screens and verified by Y2H retests	105
Tab. 30	Study-related YbeB and L14 orthologs	121
Tab. 31	DUF143/Iojap signatures appear mainly as single-domain proteins	122
Tab. 32	Interactions of YbeB that are not conserved in <i>E. coli</i>	126
Tab. 33	Putative false-negatives that failed detection in the screens.....	141
Tab. 34	Interactions of YbeB orthologs with L14 and the large ribosomal subunit.....	162
Tab. 35	Intra-viral PPIs of Dp-1	186
Tab. 36	All TP0738 (YbeB) interactions (non-redundant) identified in <i>T. pallidum</i> Y2H screens.	190
Tab. 37	All known protein-protein interactions of YbeB orthologs	191

