

**Structural and functional characterization of the molecular
interaction between the chaperone IpgC and its substrates IpaB
and IpaC from *Shigella flexneri***

Inaugural – Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin

by

RAVI KUMAR LOKAREDDY

from Visakhapatnam, Andhra Pradesh, India

August, 2008

1st Reviewer: Prof. Dr. Udo Heinemann
Freie Universität Berlin

2nd Reviewer: Prof. Dr. Arturo Zychlinsky
Humboldt Universität Berlin

Date of Defence: 26.09.2008

SELBSTSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Berlin, den 26. August 2008

Ravi Kumar Lokareddy

ZUSAMMENFASSUNG

Shigellen sind Gram-negative pathogene Bakterien, die beim Menschen das Krankheitsbild der bakteriellen Ruhr hervorrufen. Bakterienruhr stellt speziell in Entwicklungsländern ein massives Gesundheitsproblem dar. In Westeuropa sind bakteriell verursachte Infektionskrankheiten in den letzten Jahren wieder auf dem Vormarsch, was auf die Verbreitung Antibiotika-resistenter Bakterienstämme speziell aus dem östlichen Europa zurückzuführen ist. Ein besseres Verständnis bakterieller Virulenzmechanismen ist mithin grundlegend für die Entwicklung neuer Medikamente.

Shigellen, wie auch viele andere Gram-negative Krankheitserreger, besitzen einen Multi-Protein-Komplex, um spezifisch Effektorproteine in die Membran sowie das Zytoplasma eukaryontischer Wirtszellen zu transportieren. Dieser transmembrane Proteinkomplex, genannt Typ III Sekretionssystem (TTSS), ist essentiell für die Pathogenität vieler human-, tier- als auch pflanzenspezifischer pathogener Bakterien. Aufbau und Funktion dieses Proteintransportsystems erfordert ein Zusammenspiel mehrerer Chaperone, die zusammen mit anderen Komponenten des TTSS im Bakteriengenom kodiert sind. Diese Chaperone erfüllen so vielfältige Funktionen wie die die Verhinderung der vorzeitigen Aggregation sowie Assoziation interagierender Proteine und spielen bei der Sekretion der Effektoren eine wichtige regulatorische Rolle. Abhängig von der Art der Effektorbindung werden Chaperone des TTSS in 3 Klassen eingeteilt (I-III). Bindemechanismus sowie der strukturelle Aufbau der Klasse I- und III-Chaperone sind in der Literatur ausführlich beschrieben. Über Klasse II-Chaperone hingegen ist wenig bekannt.

Im Rahmen dieser Arbeit wurde die 3-dimensionale Struktur und die Funktion des Klasse II-Chaperons IpgC aus *Shigella flexneri* untersucht. Darüber gelang es, die Chaperon-Bindedomäne (CBD) der Substrate IpaB und IpaC zu identifizieren und deren Wechselwirkungen mit dem Chaperon strukturell und funktionell zu charakterisieren. In der vorliegenden Arbeit wird die erste Kristallstruktur des funktionellen Klasse II-Chaperons IpgC vorgestellt. Die zur Kristallisation benutzte Mutante ist hinsichtlich Sekretion, Invasion und Zytotoxizität voll funktionsfähig im IpgC-Deletionsstamm. Die Kristallstruktur zeigt ein IpgC-Dimer. Dies ist in Übereinstimmung mit hier durchgeführten Lichtstreuexperimenten in Lösung, in denen ebenfalls ein Dimer gefunden wurde. IpgC weist eine vorwiegend α -helicale

Struktur auf, der ein sich wiederholendes Tetratricopeptid-Motif zur Grunde liegt. Die Kristallstruktur zeigt, dass die N-terminale Helix eine wichtige Rolle in der asymmetrischen IpgC-Dimerisierung spielt. Deletion der N-terminalen Helix führt zur Proteinaggregation *in vitro* und zu Funktionsverlust von IpgC *in vivo*. Diese Ergebnisse zeigen, dass die Dimerisierung essentiell für die Funktionalität des Chaperons und folglich die Funktion des gesamten TTSS ist.

Bisher wurden CBDs vorwiegend über genetische Screens sowie biochemische Methoden grob charakterisiert. In dieser Arbeit konnten durch den Einsatz mehrerer Techniken die Lokalisierungen der CBDs in IpaB und IpaC genauer kartiert werden. Im Fall von IpaB konnte darüber hinaus und zusätzliche CBD identifiziert werden.

Auf der Basis der gewonnenen Chaperon-Substrat-Interaktionen wurden im Rahmen dieser Arbeit Kokristalle von IpgC im Komplex mit den CBDs von IpaC sowie IpaB erhalten. Die Struktur der IpaB-CBD gebunden an IpgC wurde mittels Röntgenstrukturanalyse gelöst. IpgC stabilisiert die CBD in einer ungewöhnlichen Konformation, die durch konservierte Reste innerhalb des TPR-Motifs stabilisiert wird. Basierend auf der Kokristallstruktur wurde ein Sequenzmotif identifiziert, das von verschiedenen Klasse II-Chaperonen spezifisch erkannt wird.

Insgesamt zeigt diese Studie, wie Virulenzfaktoren von ihren Chaperonen gebunden werden. Sie erweitert unser Verständnis über diese Chaperon-Gruppe.

ABSTRACT

Shigellosis manifestation is a global health problem. Additionally, emerging drug-resistant strains accentuate the situation. Better understanding of the mechanisms underlying bacterial pathogenesis would provide an edge in developing suitable counter-strategies against pathogenic bacteria. In this respect detailed insight into the structure and function of components essential for bacterial invasion and survival in host cells could provide vital information.

Pathogenic Gram negative bacteria have evolved a complex molecular machinery to deliver virulence proteins into the membrane or inside the cytoplasm of the eukaryotic cell. This machinery termed type three secretion system (TTSS) is employed by both animal and plant pathogenic bacteria. Proper assembly and operation of TTSS requires cooperation of multiple pathogenicity associated chaperones. Chaperones have been implicated in many functions ranging from prevention of aggregation and premature association between the interacting components, serving as secretion pilots and role in regulation. Based on the type of the effector binding, chaperones of TTSS have been classified into three groups. The mode of binding and structures from class I and III are well characterized, but there is dearth of information regarding class II chaperones.

The aim of this study was to characterize structure and function of a class II chaperone, IpgC from *Shigella flexneri*. Further to map precisely the chaperone binding domains (CBDs) on its substrates- IpaB and IpaC and to decipher molecular interactions defining their interaction.

This study presents the first X-ray structure of a functional class II chaperone, IpgC from *Shigella flexneri*. The IpgC crystallized was fully functional for secretion, invasion and cytotoxicity when complemented in ipgC deleted *Shigella*. The crystal structure reveals the biological unit is a dimer in agreement with dimer in solution. It folds into an all- α helical structure with consecutive tetratricopeptide motifs forming a cleft-like scaffold. The crystal structure demonstrates the crucial role of N-terminal helix in the asymmetric dimerization of IpgC. Deletion of the N-terminal helix led to aggregation of the protein and in vivo could not restore the functionality of ipgC deleted *Shigella*, thereby indicating that dimerization is important for a functional chaperone.

Hitherto, largely, the CBDs on the substrates have been characterized using genetic screening and to a lesser extent biochemical methods. By using a combination of methods and techniques the boundaries of the CBDs in IpaB and IpaC were precisely mapped quelling the ambiguity of CBDs presented in the literature. Briefly, the contradicting reports existing regarding the CBD in IpaC was addressed, while for IpaB not only was the boundary of the known CBD redefined but also an additional CBD was mapped.

Further this study presents the first co-crystal structure of class II chaperone with its substrate. Crystals of IpgC in complexes with CBDs of IpaC and IpaB were obtained and the structure of IpaB CBD-bound IpgC was solved. The chaperone captures the CBD in an extended conformation stabilized by conserved residues lining the cleft. Analysis of the co-crystal structure reveals specific interactions which define how the substrate is bound. Importantly a sequence motif in substrates recognized by class II chaperones was identified. Consequentially the generality of the sequence motif which extends and holds true for other homologue substrates of class II chaperones was confirmed. Further this study shows how virulence factors are chaperoned. It significantly advances our understanding on this class of chaperone hitherto unavailable.

TABLE OF CONTENTS

1	INTRODUCTION	1
1.1	<i>Shigella</i>	1
1.1.1	Properties	1
1.1.2	Epidemiology	2
1.1.3	Pathogenicity	2
1.2	Virulence Factors	4
1.3	Type three secretion system (TTSS)	5
1.4	Chaperones associated with TTSS	6
1.4.1	Classification of chaperones	7
1.4.2	Class I chaperones	7
1.4.3	Class III chaperones	8
1.4.5	Substrates of the class II chaperone IpgC in <i>Shigella</i>	11
1.5	Protein X-ray crystallography	13
1.5.1	Principles of X-ray diffraction	13
1.5.2	Calculation of the electron density	17
1.5.3	Solution of the crystallographic phase problem	18
1.5.4	Principles of MAD phasing	18
1.5.5	Choice of wavelengths	19
1.5.6	Methods for locating anomalous scattering atoms	19
1.6	Isothermal titration calorimetry (ITC)	20
1.7	Goals of this study	22
2	MATERIALS AND METHODS	23
2.1	Materials	23
2.1.1	Strains and Cell lines	23
2.1.2	Plasmids	24
2.1.3	DNA oligos for cloning and site-directed mutagenesis	26
2.1.4	Peptides	28
2.1.5	Culture media and antibiotics	29
2.1.6	Buffers and solutions	30
2.1.7	Enzymes and chemicals	30
2.1.8	Instrumentation	31
2.1.9	Software for X-ray structure determination	31

2.2 Methods	32
2.2.1 Molecular biology methods	32
2.2.2 Cell based methods	34
2.2.3 Protein methods	35
2.2.4 Spectroscopic methods	41
2.2.5 Isothermal Titration Calorimetry (ITC)	42
3 RESULTS	43
3.1 Part I: Structural characterization of the functional class II chaperone IpgC	44
3.1.1 Crystallization of IpgC	44
3.1.2 IpgC crystallized is fully functional	49
3.1.3 The structure of IpgC reveals TPR domains	51
3.2 Part II: Characterization of the chaperone binding domain (CBD) of IpaC	57
3.2.1 IpaC is unstructured	57
3.2.2 A novel approach to identify CBDs of IpaC	58
3.2.3 IpgC binds to N-terminal moiety of IpaC	59
3.2.4 Defining the boundaries of the CBD in IpaC	62
3.2.5 Crystallization of the IpaC-IpgC complex	64
3.3 Part III: Characterization of the CBD of IpaB	65
3.3.1 IpaB is toxic to <i>E. coli</i>	65
3.3.2 IpaB yields a stable core upon proteolytic cleavage	65
3.3.3 CBD on IpaB spans from amino acids 51 to 72	67
3.3.4 CBD mutant IpaB ^{Δ51-72} shows similar activity as full-length IpaB	67
3.3.5 An additional CBD in IpaB	70
3.3.6 Crystallization of IpaB-IpgC complex	72
3.3.7 Structure of the IpaB-IpgC complex	73
3.3.8 Specificity of pocket binding residues in IpaB	75
3.3.9 IpgC-IpaB forms a ternary complex	77
3.3.10 Crystallization of IpaB encompassing both CBDs	78
4 DISCUSSION	82
4.1 Crystallizing a functional chaperone	82
4.2 IpgC reveals TPR-like repeats	83
4.3 Polypeptide binding groove of the TPR domain	84
4.4 Functional IpgC is a dimer	87
4.5 Role of H1 in dimeric interaction	88
4.6 Characterization of the IpaC-IpgC interaction	90
4.6.1 CBD in IpaC spans from residues 36-68	90

4.7 Characterization of the IpaB-IpgC interaction	92
4.7.1 Characterizing the CBD in IpaB	92
4.7.2 An additional CBD in IpaB (CBD1)	93
4.7.3 <i>In vivo</i> relevance of CBD deletions	93
4.8 Polypeptide binding to groove of the TPR domain	94
4.9 A conserved chaperone binding motif	96
5 REFERENCES	99
6 APPENDIX	110

ABBREVIATIONS

α	alpha
Δ	truncation (amino acids) / deletion (gene)
$^{\circ}\text{C}$	degree Centigrade (Celsius)
μ	micro (10^{-6})
A	adenosine
APS	ammonium persulfate
ATP	adenosine-triphosphate
BESSY	Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung
bp	base pair
BSA	bovine serum albumin
C	cytosine / carboxy terminal end of polypeptide / protein
CBD	chaperone binding domain
CBD1	CBD in IpaB18-35 (residues 18-35)
CBD2	CBD in IpaB 51-72 (residues 51-72)
C-ter	carboxy terminal end of polypeptide / protein
Da	molecular weight unit, Dalton, \sim g/mol
DESY	Deutsches Elektronen-Synchrotron
DMEM	Dulbecco's modified Eagle medium
DNA	deoxyribonucleic acid
ECL	enhanced chemiluminescence
EDTA	ethylene diamine tetraacetate
ESRF	European Synchrotron Radiation Facility, Grenoble, France
FCS	fetal calf serum
FPLC	fast performance liquid chromatography
G	guanosine
HEK	human embryonic kidney cell
HEPES	2-[4-(2-hydroxyethyl)-1-piperazino]-ethansulfonic acid
His	Histidine
Hsp	Heat shock protein
IL	interleukin
IpaB	invasion plasmid antigen B
IpaC	invasion plasmid antigen C
IpgC	invasion plasmid gene C
IPTG	isopropyl thiogalactoside
ITC	isothermal titration calorimetry
k	kilo, 10^3
kb	kilo base
kDa	kilo Dalton
l	volume unit, litre
LB	Luria-Bertani medium
m	milli, 10^{-3}
M	concentration unit, molar, moles per liter
MAD	multiwavelength anomalous dispersion
MALDI	matrix assisted laser desorption/ ionization
MALLS	Multi-angle laser light scattering
min	time unit, minute, equals 60 seconds
MIR	multiple isomorphous replacement
mol	molecular unit, mole, equals 6.023×10^{23} particles of a substance
MR	molecular replacement

MS	mass spectrometry
N	amino terminal end of polypeptide / protein
NCS	non-crystallographic symmetry
NTA	nitrilotriacetic acid
N-ter	amino terminal end of polypeptide / protein
OD	optical density
pBpa	para-benzoyl phenylalanine
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PBS	phosphate-buffered saline
PDB	Protein Data Bank
PEG	polyethylene glycol
PMSF	phenylmethylsulfonyl fluoride
pI	isoelectric point
psi	pressure unit, Pascal
PVDF	polyvinylidene fluoride
QCM	QuikChange mutagenesis
Rmsd	root mean square deviation
RPMI	medium developed at 'Roswell Park Memorial Institute'
s	time unit, second
SDS	sodium dodecylsulfate
SEC	size exclusion chromatography
SU	subunit
SycD	specific yop chaperone D
T	thymidine
TBE	TRIS-borate-EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
TPR	tetratricopeptide-like repeat
TRIS	Tris(hydroxymethyl)-aminomethane
TTSS	type three secretion system
UV	ultraviolet radiation
V	voltage
wt	wild-type
Yop	Yersinia outermembrane protein

Three letter and single letter codes for amino acids

Ala/A	Alanine	Leu/L	Leucine
Arg/R	Arginine	Lys/K	Lysine
Asn/N	Asparagine	Met/M	Methionine
Asp/D	Aspartate	Phe/F	Phenylalanine
Cys/C	Cysteine	Pro/P	Proline
Gly/G	Glycine	Ser/S	Serine
Gln/Q	Glutamine	Thr/T	Threonine
Glu/E	Glutamate	Trp/W	Tryptophan
His/H	Histidine	Tyr/Y	Tyrosine
Ile/I	Isoleucine	Val/V	Valine