A Putative Mucin-degrading Operon and a 6-Sulfo-N-acetyl-β-D-glucosaminidase in *Prevotella* strain RS2

Damian Paul Wright

Anaerobic Microbial Research Group

School of Biological Sciences

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Biological Sciences.

The University of Auckland, June, 2003

ABSTRACT

A protective layer of mucus lines the surface of the gastrointestinal tract. Agents or processes that interrupt the continuity of this layer can potentially compromise the integrity of the mucosa, causing disease. The anaerobic bacterium *Prevotella* strain RS2 utilises the major structural component of the mucus layer, the mucous glycoprotein or mucin, as an energy source. *Prevotella* strain RS2 possesses a novel mucin-desulfating sulfatase (MdsA) which removes the protective sulfate ester group from N-acetyl-β-D-glucosamine-6-sulfate sugar residues of mucin oligosaccharide side chains. The desulfation of sulfated sugar residues in mucin is thought to be the rate-limiting step in mucin degradation. The gene sequence for MdsA has previously been determined.

The 4650 bp of genomic DNA downstream of the gene encoding MdsA was amplified by inverse PCRs and sequenced. Two complete open reading frames, *mdsC* and *mdsD*, were identified and are predicted to encode putative proteins of 362 and 1073 amino acids respectively. The deduced MdsC protein exhibited homology to the catalytic domain of aminoglycoside phosphotransferases and may perform a similar function. The deduced MdsD protein is predicted to be transported to the periplasm and/or outer membrane, and based on homology, may have a multiprotein proteolytic function similar to the tricorn protease. Northern and Western blot analyses suggest *mdsA*, *mdsC* and *mdsD* are coexpressed and their expression is increased when *Prevotella* strain RS2 is grown on mucin. This work is the first mucin-degrading operon of genes to be described.

Also during the course of this research, a novel desulfating enzyme, 6-sulfo-N-acetyl-β-D-glucosaminidase (6-SNG) was discovered in *Prevotella* strain RS2. The enzyme can remove intact N-acetyl-β-D-glucosamine-6-sulfate residues from sulfated sugar substrates and is different to MdsA which removes the sulfate ester group. A procedure for the partial purification of active 6-SNG has been determined. The 6-SNG is predicted to corresponded to a 67.0 kDa SDS-PAGE protein band though several isoenzymes or fragmentation products were detected during the purification. The 6-SNG is located in the periplasm of *Prevotella* strain RS2 and is expressed at increased levels when grown on mucin. Inhibition studies for the partially purified enzyme have identified a number of potential enzyme inhibitors. This work represents the first 6-SNG to be described that may play a role in mucin-desulfation.

ACKNOWLEDGMENTS

I would like to thank Assoc. Prof. Tony Roberton for his supervision of the work described in this thesis and for his patience and assistance in the preparation of this manuscript.

I would also like to express my gratitude to Assoc. Prof. David Christie, Dr. Gillian Lewis and Dr. David Saul for advice and use of equipment from their laboratories. Thanks also to friends and colleagues within the School of Biological Sciences, The University of Auckland.

Further appreciation to The University of Auckland scholarship committee for receipt of a University of Auckland Maori and Pacific Islander Doctoral Scholarship.

Finally, aroha to Mum, Dad, Jared, Shontaal, Louise and 'peepi'.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATION

AA	amino acid(s)
ADNJ	2-acetamido-1,2-dideoxynojirimycin
Amp	ampicillin
apt	anthranilate phosphoribosyl transferase gene
ATCC	American Type Culture Collection
ATP	adenosine 5'-trisphosphate
bp	DNA base pairs
BSA	bovine serum albumin
°C	degrees celsius
Cam	chloramphenicol
cAMP	cyclic adenosine monophosphate
Cat. No.	catalogue number
cc	cubic centimetre
cDNA	complimentary DNA
CF	cystic fibrosis
CFB	Cytophaga/Flexibacter/Bacteroides
cm	centimetre
CPC	cetylpridinium chloride
cpm	counts per minute
CS	chondroitin sulfate
C-terminal	carboxyl terminal
CV	column volume
Da	Dalton(s)
dATP	2'-deoxy adenosine 5'-triphosphate
dCTP	2'-deoxy cytosine 5'-triphosphate
DEPC	diethylpyrocarbonate
DFP	diisopropylfluorophosphate
DGJ	deoxygalactonojirimycin
dGTP	2'-deoxyguanine 5'-triphosphate
DMDP	2,5-dideoxy-2,5-imino-D-mannitol
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DNJ	deoxynojirimycin
dNTP	2'-deoxyribonucleoside 5'-triphosphate
dsDNA	double stranded deoxyribonucleic acid
DTT	dithiolthreitol

dTTP	2'-deoxythymine 5'-triphosphate
EC	enzyme classification
EDTA	ethylenediaminetetraacetic acid
EGTA	ethyleneglycol-bis(β-aminoethylether)-N,N'-tetraacetic acid
ELISA	enzyme linked immunosorbant assay
Fig.	figure
Fuc	L-fucose
G	gauge of needle
xg	unit of gravity
GAG	glycosaminoglycan
Gal	D-galactose
Gal-3-SO_4	D-galactose-3-sulfate
Gal-6-SO_4	D-galactose-6-sulfate
GalNAc	N-acetyl-D-galactosamine
GI	gastrointestinal
Glc	D-glucose
Glc-6-SO_4	D-glucose-6-sulfate
GlcNAc	N-acetyl-D-glucosamine
GlcNAc-6-SO_4	N-acetyl-D-glucosamine-6-sulfate
GOR	glucose oxidase reagent
GST	glutathione S-transferase
h	hour(s)
^3H	tritium
HA	hyaluronic acid
HKY	Hasegawa Kishino Yano
H_2O_2	hydrogen peroxide
HPLC	high performance liquid chromotography
InvPCR	inverse polymerase chain reaction
IPTG	isopropyl-β-D-thiogalactopyranoside
IRL	Industrial Research Limited
IRL/GEC160	6-acetamido-6-deoxycastanospermine
IRL/JMC344	8-benzylamino-8-deoxyaustraline
kb	kilobase pairs
kDa	kiloDalton(s)
K_m	Michaelis constant
kPa	kilo Pascals
L	litre(s)

LB	Luria broth
Le	Lewis antigen
mL	millilitre(s)
μL	microlitre(s)
M	moles per litre
MDa	mega Dalton(s)
mds	mucin-desulfating sulfatase (gene)
Mds	mucin-desulfating sulfatase (protein)
mg	milligram(s)
μg	microgram(s)
MHO	reciprocal of Ohm (Ohm^{-1})
min	minute(s)
μm	micrometre(s)
μM	micromoles per litre
mM	millimoles per litre
mRNA	messenger RNA
MOD	mucin oligosaccharide degrading
mol	mole(s)
MOPS	3-[N-Morpholino]propane-sulphonic acid
MPa	mega Pascals
MSc	Master of Science
MUBF	4-methylumbelliferone
MUC	human mucin gene
Mw	molecular weight
NADH	reduced nicotinamide adanine dinucleotide
NANA	α-D-N-acetylneuraminic acid
NCBI	National Centre for Biotechnology Information
NCTC	National Culture Type Collection
N/D	no data available
Neu5Ac	5-N-acetylneuraminic acid
ng	nanogram(s)
NIH	National Institute of Health
nm	nanometre(s)
nmol	nanomole(s)
NMR	nuclear magnetic resonance
N-terminal	amino terminal
OD	optical density
Ohm	unit of electrical resistence

ORF	open-reading frame
OTU	operational taxonomic unit
^{32}P	radioisotope of phosphorous
PAPS	3'-phosphoadenodine-5'-phosphosulphate
PAUP	Phylogenetic Analysis Using Parsimony
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	protein databank
PEG	polyethyleneglycol
pI	isoelectric point
PGM	pig gastric mucin
PGMt	trypsinised pig gastric mucin
pH	$-\log_{10}$ hydrogen ion (H^+) concentration
PhD	Doctor of Philosophy
pNP	para-nitrophenol
pNP-	para-nitrophenyl (substituted)
ppm	parts per million
PSGL	P-selectin glycoprotein ligand-1
p.s.i.	pounds per square inch
PSI	position specific iterated
RDP	ribosomal database project II
Reg. no.	registration number
R_f	relative migration
RFDNA	replicative form deoxyribonucleaic acid
RP-HPLC	reverse phase-high performance liquid chromatography
RNA	ribonucleic acid
RNase	ribonuclease
rDNA	ribosomal deoxyribonucleic acid
rRNA	ribosomal ribonucleic acid
^{35}S	radioisotope of sulfur
SDS	soduim dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
sec	second(s)
6-SNG	6-sulfo-N-acetyl-β-D-glucosaminidase
sp.	species
spp.	more than one species

SSC	sodium chloride, sodium citrate buffer
ssDNA	single stranded deoxyribonucleaic acid
SSPE	sodium chloride, sodium dihydrogen phosphate, EDTA buffer
STEL	Tris.HCl, EDTA, lysozyme, sodium deoxycholate
STET	sodium chloride, Tris.HCl, EDTA, Triton X-100™
SUB	solubilisation buffer
TAE	Tris.acetate, EDTA buffer
TBE	Tris.borate, EDTA buffer
TBR	Tree Bisection Reconnection
TBS	Tris.HCl buffered saline
TBE-T	Tris.HCl buffered saline Tween® 20
TE	Tris.HCl, EDTA buffer
Tet	tetracycline.HCl
TFA	trifluoroacetic acid
T_m	melting temperature
TNE	Tris.HCl, sodium chloride, EDTA
TNE-T	Tris.HCl, sodium chloride, EDTA, Triton X-100™
Tp	primer annealing temperature
Tris.HCl	2-amino-2-hydroxymetyl-propane1,3-diol hydrogen chloride
Triton X-100	iso-octylphenoxypolyethoxyethanol
tRNA	transfer RNA
UC	ulcerative colitis
UV	ultra violet light
UWGCG	University of Wisconsin Genetics Computer Group
V	volts
VFA	volatile fatty acid(s)
V_{max}	maximum velocity
v/v	volume per volume
w/v	weight per volume
w/w	weight per weight
X-gal	5-bromo-4-chloro-3-indolyl β-D-galacto-pyranoside
zbp	zinc binding protein gene

The standard three letter and one letter codes were used to abbreviate amino acids appropriately. Standard restriction endonuclease and plasmid/vector abbreviations were used.