Supporting Information

Crystal Structure of the $MrkD_{1P}$ Receptor Binding Domain of *Klebsiella pneumoniae* and Identification of the Human Collagen V Binding Interface

Ana Toste Rêgo[1,2], Jeremiah G. Johnson[3,4], Sebastian Geibel[1], Francisco J. Enguita[5], Steven Clegg[3] and Gabriel Waksman[1,6*]

[1]Institute of Structural and Molecular Biology at University College London and Birkbeck College, Malet Street, London WC1E 7HX, UK, [2]Current address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK, [3]Department of Microbiology, College of Medicine, University of Iowa, Iowa City, Iowa, 52242, USA, [4]Current address: Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan, 48109, USA MRC, [5]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal, [6]Research Department of Structural and Molecular Biology, University College London, Gower Street, WC1E 6BT, UK.

* Author for correspondence

Professor Gabriel Waksman
Institute of Structural and Molecular Biology at University College London and Birkbeck College, Malet Street,
Malet Street London, WC1E 7HX, UK
Tel 020 7631 6833; Fax 020 7631 6803

Email: g.waksman@bbk.ac.uk or g.waksman@ucl.ac.uk

Supporting Figure Legends

Fig. S1. Diagram of the type 3 pilus structure (A) and type 3 pili plasmid-born gene cluster (B).

A) The periplasmic chaperone MrkB captures subunits at the exit of the Sec pathway and assists in their folding and targeting to the outer membrane usher MrkC. Once the usher is activated by incorporation of the first subunit adhesin MrkD, subunits polymerise in an ordered sequence until pilus completion. The extracellular, outer membrane and periplasmic spaces are bold labelled E, OM and P, respectively.

B) Type 3 plasmid-born gene cluster. Genes are represented in square boxes color-coded according to function as in A. Labels inside the boxes go from A-G and should be read with the cluster name (*mrk*) first and then the letter code.

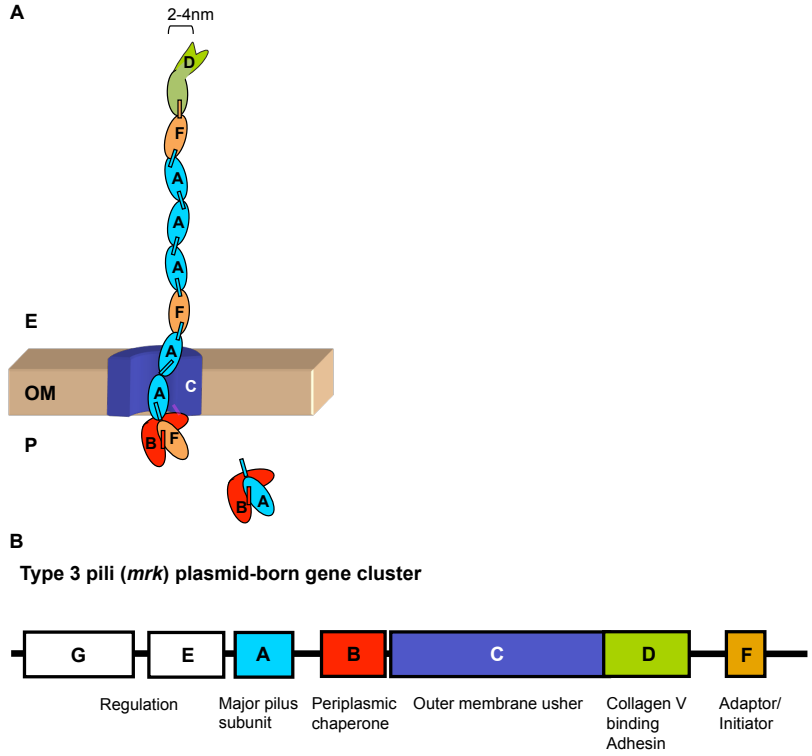

Fig. S2. Protein gel of the MrkB-MrkD$_{1P}$ complex trypsin digestion.

Samples taken at time 0, 5, 15, 30, 60 and 120 minutes after proteolysis by two different w/w ratios (1:100 and 1:1000) of trypsin:complex were loaded on an SDS gel and molecular weights estimated by comparison with the molecular standards shown on the first lane (PS- protein standards). Bands 1-3 were sequenced by Edman degradation. Band one was confirmed to contain the N-terminal receptor binding of domain of MrkD$_{1P}$.

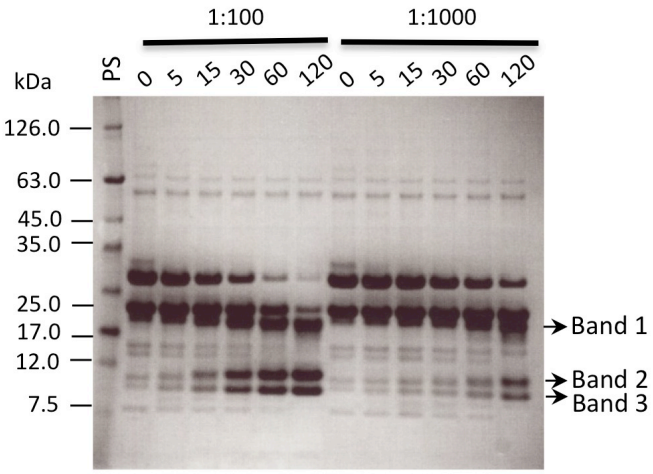

Fig. S3. Protein sequence alignment of adhesins MrkD, PapGII and FimH.

Alignment was generated by ClustalW (Larkin et al., 2007) and loaded to the ESPript server (Gouet et al., 1999). White characters in red boxes represent strict identity, red characters represent conserved substitutions and a blue frame represent semi-conserved substitutions. The secondary structure of the FimH crystal structure is shown (PDB: 1QUN). Residue numbering on the top is based on the FimH sequence. The end of the receptor binding domain and the beginning of the pilin domain are marked by red and green arrows below the protein sequence for FimH and PaGII respectively. Dashed green line underlines the possible localization of the C-terminus of MrkD receptor binding domain.

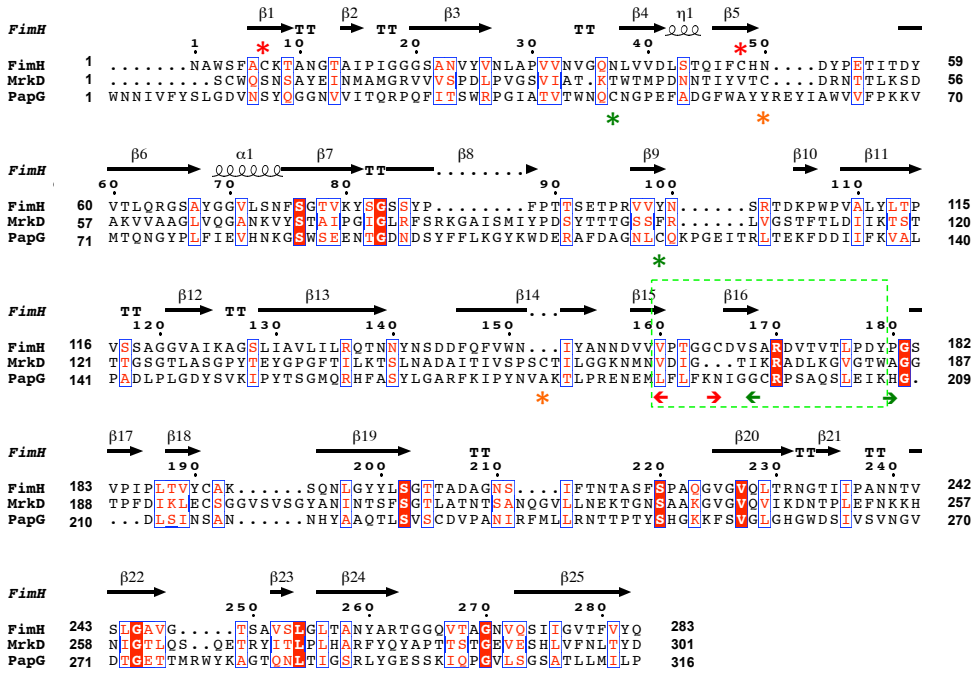

Fig. S4. Pictures of the MrkDrd native crystals. MrkD receptor domain crystals were grown in 0.1 M Sodium Acetate pH 4.6, 8 % PEG 4000 (sitting drop).

MrkD receptor binding domain was concentrated to 7 $mg.ml^{-1}$. The picture on panel A is identical to that on panel B but under polarized light.

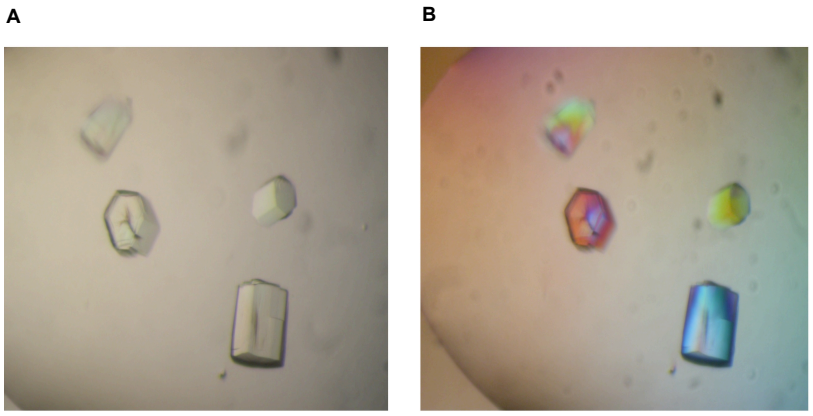

Fig. S5. Representative region of the 2 Fo-Fc electron density map, from PHENIX, around residues Met33 and Met35 of MrkDrd. Right panel: the magenta mesh represents the electron density map calculated using SAD phases and contoured at 1.5 sigma. Left panel: the blue mesh represents the 2 Fo-Fc electron density contoured at 1.5 sigma. In both panels, the orange mesh represents the Fo-Fc difference density contoured at 3.6 sigma, clearly showing the difference density for the selenium atom of two seleno-methionines.

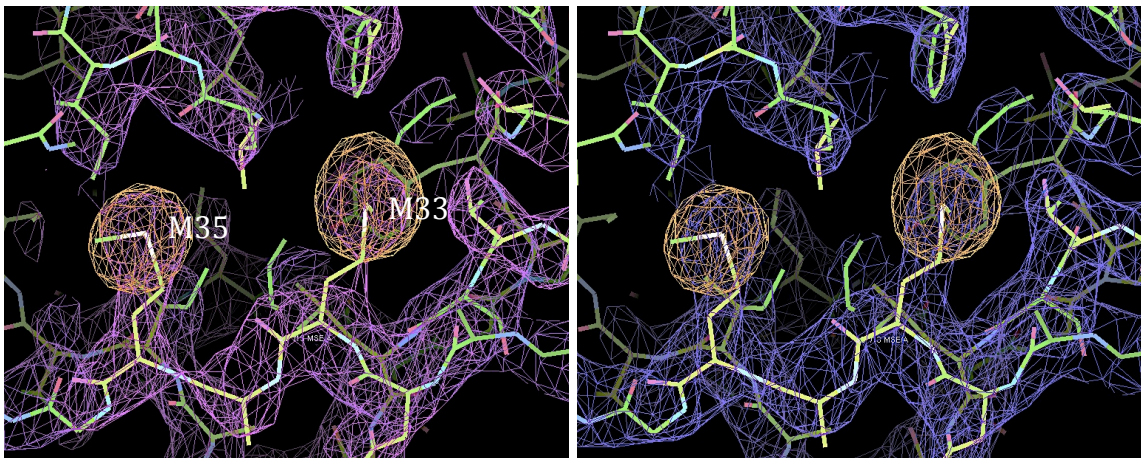

Fig. S6. Crystal structures of the receptor binding domains of adhesins $MrkD_{1P}$, GafD/F17-G, FimH, PapGII.

From left to right: $MrkD_{1P}$ receptor domain (this work), GafD/F17-G bound to N-acetyl-D-glucosamine (GlcNAc) (pdb 1OIO), FimH in complex with Man α 1,3 Man β 1,4 GlcNAc β 1,4 GlcNAc (oligomannose-3) (pdb 2VCO) and PapGII bound to GalNAc β 1-3Gal α 1-4Gal β 1-4Glc (galabiose) (pdb 1J8R). Structures are shown as ribbon representation. Disulfide bridges are presented in orange in all structures and sugars in blue.

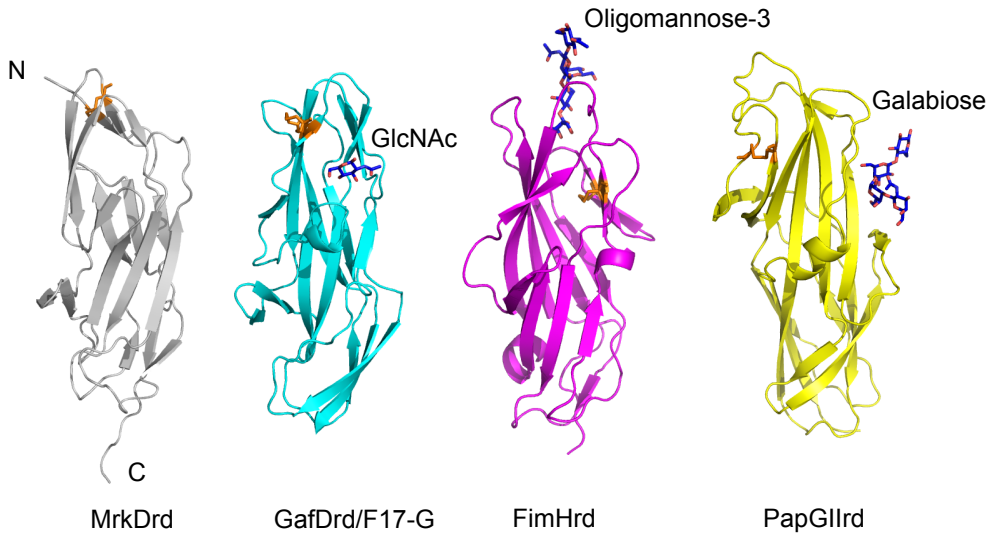

Fig. S7. The "Collagen Hug model" of adhesin CNA. Crystal structure of CNA binding to collagen (PDB: 2F6A). CNA holds the collagen triple helix by two subdomains N1 (cyan) and N2 (magenta) that are connected by a long hydrophobic linker (blue). A C-terminal extension of N2 then "traps" collagen by insertion in the latching trench of N1 in a 'strand addition' mode.

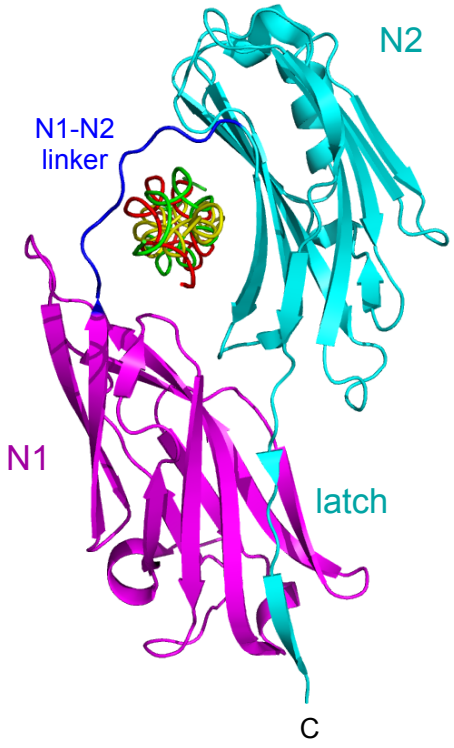

Fig. S8. Structural alignment of the receptor domains of FimH and *MrkD$_{1P}$* .

Structural alignment of FimH (purple) and MrkD$_{1P}$ (grey) shows that residue 62 in FimH reported as being fundamental for Collagen binding is located in a different side of the molecule having no analogy to the described hydrophobic patch described in this study for MrkD collagen binding.

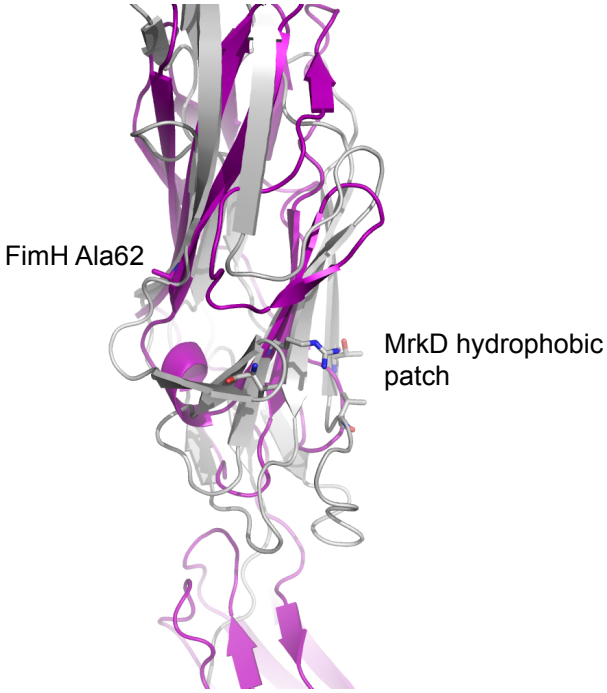

Table S1: Oligonucleotides used in this study

Oligonucleotide	Sequence (5'-3')
MrkD21_IBA12-f	ATGGTAGGTCTCACTCCTCATGTTGGCAATCTAATAGTGCC
MrkD198_IBA12-r	ATGGTAGGTCTCATATCATTTTAAATCAGCTCGTTTAATCGTGC
ΔmrkDForXhoI	CATGCACTCGAGCTGGTCTTTAACCTGACGTACGAT
ΔmrkDRevXhoI	CATGCACTCGAGCGTTGTCCCCAGCGCCATTAAGCCAAT
5'mrkDBamHI	CATGCAGGATCCACAGTGCTCGGTGGAATTAAGT
3'mrkDHindIII	CATGCAAAGCTTTCAGGGCCAACTGAATGAGTCG
MrkD W23A For	CGTCAGCATGGGCATCATGTGCGCAATCTAATAGTGCCTATGAA
MrkD W23A Rev	TTCATAGGCACTATTAGATTGCGCACATGATGCCCATGCTGACG
MrkD V39G For	AATATGGCTATGGGGCGCGTGGGTGTTAGCCCGGACTTACCAGTA
MrkD V39G Rev	TACTGGTAAGTCCGGGCTAACACCCACGCGCCCCATAGCCATATT
MrkD V49G For	GCCCGGACTTACCAGTAGGGAGTGGCATTGCAACTAAAACATGGACAATG
MrkD V49G Rev	CATTGTCCATGTTTTAGTTGCAATGCCACTCCCTACTGGTAAGTCCGGGC
MrkD V49A For	GAC TTA CCA GTA GGG AGT GCC ATT GCA ACT AAA ACA TGG
MrkD V49A Rev	CTG AAT GGT CAT CCC TCA CGG TAA CGT TGA TTT TGT ACC
MrkD T52A For	GTA GGG AGT GTC ATT GCA GCT AAA ACA TGG ACA ATG C
MrkD T52A Rev	CAT CCC TCA CAG TAA CGT CGA TTT TGT ACC TGT TAC G
MrkD T52S For	GTA GGG AGT GTC ATT GCA TCT AAA ACA TGG ACA ATG C
MrkD T52S Rev	CAT CCC TCA CAG TAA CGT AGA TTT TGT ACC TGT TAC G
MrkD T54A For	GGA GTG TCA TTG CAA CTA AAG CAT GGA CAA TGC CG
MrkD T54A Rev	CCT CAC AGT AAC GTT GAT TTC GTA CCT GTT ACG GC
MrkD V85G For	GCGAAAGTTGTTGCTGCTGGTTTGGGTCAGGGGGCCAATAAAGTCTA
MrkD V85G Rev	TAGACTTTATTGGCCCCCTGACCCAAACCAGCAGCAACAACTTTCGC
MrkD V91G For	GGTTCAGGGGGCCAATAAAGGCTATTCAACTGCAATTCCTGG
MrkD V91G Rev	CCAGGAATTGCAGTTGAATAGCCTTTATTGGCCCCCTGAACC
MrkD R102G For	GCAATTCCTGGAATTGGTTTAGGTTTCTCACGTAAAGGGGCGATC
MrkD R102G Rev	GATCGCCCCTTTACGTGAGAAACCTAAACCAATTCCAGGAATTGC
MrkD R105E For	CTGGAATTGGTTTACGTTTCTCAGAGAAAGGGGCGATCAGTATGATC
MrkD R105E Rev	GATCATACTGATCGCCCCTTTCTCTGAGAAACGTAAACCAATTCCAG
MrkD K106A For	GGTTTACGTTTCTCACGTGCGGGGGCGATCAGTATGATCTAC
MrkD K106A Rev	GTAGATCATACTGATCGCCCCCGCACGTGAGAAACGTAAACC
MrkD T130A For	CCT CCT TTA GAC TCG TAG GTT CAG CAT TCA CAT TAG ATA TAA TTA AGA
MrkD T130A Rev	GGA GGA AAT CTG AGC ATC CAA GTC GTA AGT GTA ATC TAT ATT AAT TCT
MrkD T132A For	TTA GAC TCG TAG GTT CAA CAT TCG CAT TAG ATA TAA TTA AGA CCA GT
MrkD T132A Rev	AAT CTG AGC ATC CAA GTT GTA AGC GTA ATC TAT ATT AAT TCT GGT CA
MrkD I136A For	CTC GTA GGT TCA ACA TTC ACA TTA GAT ATA GCT AAG ACC AGT ACC AC
MrkD I136A Rev	GAG CAT CCA AGT TGT AAG TGT AAT CTA TAT CGA TTC TGG TCA TGG TG
MrkD I136G For	TCAACATTCACATTAGATATAGGCAAGACCAGTACCACAACAGGGAGTGG
MrkD I136G Rev	CCACTCCCTGTTGTGGTACTGGTCTTGCCTATATCTAATGTGAATGTTGA
MrkD Y155A For	AGCCAGTGGGCCGTATACAGAGGCCGGACCAGGATTTACAATCCTT
MrkD Y155A Rev	AAGGATTGTAAATCCTGGTCCGGCCTCTGTATACGGCCCACTGGCT
MrkD Y155F For	AGCCAGTGGGCCGTATACAGAGTTTGGACCAGGATTTACAATCCTT
MrkD Y155F Rev	AAGGATTGTAAATCCTGGTCCAAACTCTGTATACGGCCCACTGGCT
MrkD K163A For	GGACCAGGATTTACAATCCTTGCGACCAGCCTTAATGCTGATGCC
MrkD K163A Rev	GGCATCAGCATTAAGGCTGGTCGCAAGGATTGTAAATCCTGGTCC
MrkD V174G For	GCTGATGCCATTACAATTGGTTCACCTTCTTGTACCATT
MrkD V174G Rev	AATGGTACAAGAAGGTGAACCAATTGTAATGGCATCAGC

Table S2: Plasmids used in this study

Plasmid	Description	Source
pFK12	Cam^r; pACYC184-based plasmid containing IA565-derived determinants *mrkABCD*	Gerlach et al., 1989
pFK68	Cam^r; pACYC184-based plasmid containing IA565-derived determinants *mrkABCDF*	Allen et al., 1991
pFK68 Δ *mrkD*	Cam^r; *mrkD* deleted derivative of pFK68	This study
pGEM-T Easy	Amp^r; subcloning vector	Promega (Madison, WI)
pTrc99A	Amp^r; expression vector	Amersham-Pharmacia (Piscataway, NJ)
pTrc99*mrkD*	Amp^r; expresses wild-type IA565 MrkD	This study
pTrc99*mrkD*$_{W23A}$	Amp^r; expresses IA565 MrkD(W23A)	This study
pTrc99*mrkD*$_{V39G}$	Amp^r; expresses IA565 MrkD(V39G)	This study
pTrc99*mrkD*$_{V49G}$	Amp^r; expresses IA565 MrkD(V49G)	This study
pTrc99*mrkD*$_{V49A}$	Amp^r; expresses IA565 MrkD(V49A)	This study
pTrc99*mrkD*$_{T52A}$	Amp^r; expresses IA565 MrkD(T52A)	This study
pTrc99*mrkD*$_{T52S}$	Amp^r; expresses IA565 MrkD(T52S)	This study
pTrc99*mrkD*$_{T54A}$	Amp^r; expresses IA565 MrkD(T54A)	This study
pTrc99*mrkD*$_{V85G}$	Amp^r; expresses IA565 MrkD(V85G)	This study
pTrc99*mrkD*$_{V91G}$	Amp^r; expresses IA565 MrkD(V91G)	This study
pTrc99*mrkD*$_{R102G}$	Amp^r; expresses IA565 MrkD(R102G)	This study
pTrc99*mrkD*$_{R105E}$	Amp^r; expresses IA565 MrkD(R105E)	This study
pTrc99*mrkD*$_{K106A}$	Amp^r; expresses IA565 MrkD(K106A)	This study
pTrc99*mrkD*$_{T130A}$	Amp^r; expresses IA565 MrkD(T130A)	This study
pTrc99*mrkD*$_{T132A}$	Amp^r; expresses IA565 MrkD(T132A)	This study
pTrc99*mrkD*$_{I136A}$	Amp^r; expresses IA565 MrkD(I136A)	This study
pTrc99*mrkD*$_{I136G}$	Amp^r; expresses IA565 MrkD(I136G)	This study
pTrc99*mrkD*$_{Y155A}$	Amp^r; expresses IA565 MrkD(Y155A)	This study
pTrc99*mrkD*$_{Y155F}$	Amp^r; expresses IA565 MrkD(Y155F)	This study
pTrc99*mrkD*$_{K163A}$	Amp^r; expresses IA565 MrkD(K163A)	This study
pTrc99*mrkD*$_{V174G}$	Amp^r; expresses IA565 MrkD(V174G)	This study