Mutagenesis of the Rns regulator of enterotoxigenic *Escherichia coli* reveals roles for a linker sequence and two helix–turn–helix motifs

Vivienne Mahon,^{1,2,3†} Cyril J. Smyth² and Stephen G. J. Smith^{1,2,3}

Correspondence
Stephen G. J. Smith
sgsmith@tcd.ie

¹Department of Clinical Microbiology, School of Medicine, Trinity College Dublin, St James's Hospital, Dublin 8, Ireland

²Department of Microbiology, School of Genetics and Microbiology, Moynihan Institute of Preventive Medicine, Trinity College Dublin, University of Dublin, Dublin 2, Ireland

³Institute of Molecular Medicine, Trinity Centre, Trinity College Dublin, St James's Hospital, Dublin 8, Ireland

The pathogenesis of diarrhoeal disease due to human enterotoxigenic *Escherichia coli* absolutely requires the expression of fimbriae. The expression of CS1 fimbriae is positively regulated by the AraC-like protein Rns. AraC-like proteins are DNA-binding proteins that typically contain two helix–turn–helix (HTH) motifs. A program of pentapeptide insertion mutagenesis of the Rns protein was performed, and this revealed that both HTH motifs are required by Rns to positively regulate CS1 fimbrial gene expression. Intriguingly, a pentapeptide insertion after amino acid C102 reduced the ability of Rns to transactivate CS1 fimbrial expression. The structure of Rns in this vicinity (NACRS) was predicted to be disordered and thus might act as a flexible linker. This hypothesis was confirmed by deletion of this amino acid sequence from the Rns protein; a truncated protein that lacked this sequence was no longer functional. Strikingly, this sequence could be functionally substituted *in vivo* and *in vitro* by a flexible seven amino acid sequence from another *E. coli* AraC-like protein RhaS. Our data indicate that HTH motifs and a flexible sequence are required by Rns for maximal activation of fimbrial gene expression.

Received 3 February 2010

Revised 12 May 2010

Accepted 22 May 2010

INTRODUCTION

Enterotoxigenic *Escherichia coli* (ETEC) is a major cause of diarrhoeal disease in humans and animals (Qadri *et al.*, 2005; Turner *et al.*, 2006). ETEC pathogenesis is dependent on production of both a toxin (heat-labile toxin and/or heat-stable toxin) and fimbriae (Gaastra & Svennerholm, 1996). The Rns protein was originally identified as being required for the expression of CS1 and CS2 fimbriae of ETEC (Caron *et al.*, 1989). Rns also positively autoregulates its gene expression (Froehlich *et al.*, 1994) and activates the genes *yjiS* (unknown function) (Munson *et al.*, 2002) and *cexE*, which encodes an uncharacterized extracytoplasmic protein (Pilonieta *et al.*, 2007). In addition, Rns can act as a

negative regulator, repressing the expression of NlpA (Bodero *et al.*, 2007).

The *coo* operon (*cooBACD*) that encodes the CS1 fimbriae is repressed by H-NS and activated by Rns (Murphree *et al.*, 1997). Rns activates the expression of the *coo* genes directly by binding to two sites upstream of the *coo* promoter. As site II overlaps the –35 hexamer, it has been suggested that Rns activates transcription from the *coo* promoter by forming direct contacts with RNA polymerase (Munson & Scott, 1999). A highly unconventional arrangement of binding sites are involved in Rns autoactivation. Rns binds to a site far upstream of the *rns* promoter and to two sites downstream of it (Munson & Scott, 2000). Downstream binding sites are usually associated with repressors of transcription, yet activation of the *rns* promoter *in vivo* is dependent on the upstream binding site and at least one of the downstream sites (Munson & Scott, 2000).

Rns is a member of the AraC family of transcriptional regulators (Caron *et al.*, 1989), a family of proteins defined by a 100 aa region of homology that contains two predicted helix–turn–helix (HTH) DNA-binding motifs (Gallegos *et al.*, 1997; Ibarra *et al.*, 2007). Among AraC

[†]Present address: NIBRT, Dublin City University, Glasnevin, Dublin 9, Ireland.

Abbreviations: EMSA, electrophoretic mobility shift assay; ETEC, enterotoxigenic *Escherichia coli*; HTH, helix–turn–helix; MBP, maltose-binding protein.

Five supplementary figures, showing expression and purification of His-tagged Rns and mutant derivatives and models of the Rns N- and C-termini [in 2D and 3D (PDB) format], are available with the online version of this paper.

family members, there is a low level of sequence conservation at the first predicted HTH motif (HTH1). This may represent involvement of HTH1 in the recognition of the different target sites of each member (Gallegos *et al.*, 1997). In contrast, the sequence of the second predicted HTH motif (HTH2) was found to be much less variable, prompting the suggestion that this motif may be involved in a common role other than DNA binding, such as forming contacts with RNA polymerase (Gallegos *et al.*, 1997). Rns has two predicted HTH motifs in its C-terminal domain (Munson *et al.*, 2001), so it is assumed that this area of Rns is involved in DNA binding, although this has not been verified previously.

Less is known about the role of the N-terminal domain of Rns. Rns does not appear to respond to an exogenous ligand (Basturea *et al.*, 2008). Thus, while the possibility cannot be ruled out, it is unlikely that the N-terminal portion of Rns is an effector-binding domain. Additionally, it has been reported that the N terminus of Rns does not act as a dimerization domain (Basturea *et al.*, 2008). However, the N-terminal domain has been found to be essential for transcriptional regulation and DNA binding by Rns (Basturea *et al.*, 2008).

Because of its importance in regulation of fimbrial gene expression among human ETEC (Anantha *et al.*, 2004; Boderio *et al.*, 2008), mutagenesis of the *rns* gene was undertaken to define functionally relevant parts of the Rns protein.

METHODS

Bacterial strains and growth conditions. Plasmids and *E. coli* strains used in this study are listed in Table 1. Bacteria were cultured at 37 °C in Luria broth or on Luria agar. Where appropriate, antibiotics were used at the following concentrations: 50 µg carbenicillin ml⁻¹, 50 µg kanamycin ml⁻¹, 50 µg spectinomycin ml⁻¹, 10 µg tetracycline ml⁻¹ and 30 µg chloramphenicol ml⁻¹.

DNA manipulations. DNA manipulations were performed according to standard protocols (Sambrook *et al.*, 1989). Plasmid DNA was purified using the HiYield Plasmid kit (Real Biotech Corporation) according to the manufacturer's instructions. PCR was performed with Phusion High Fidelity DNA Polymerase (Finnzymes). Restriction endonucleases and Quick T4 DNA Ligase were purchased from New England Biolabs and used according to the manufacturer's instructions. DNA samples were purified by agarose gel electrophoresis followed by gel extraction with the GeneClean II kit (QBiogene) as described by the manufacturer. Standard methods were used for the transformation of plasmid DNA (Dower *et al.*, 1988). DNA sequencing and synthesis of all oligonucleotide primers were performed by MWG-Biotech (Ebersberg, Germany). Site-directed mutagenesis was carried out with the QuikChange kit (Stratagene).

Pentapeptide scanning mutagenesis. The GPS-LS linker-scanning system (NEB) was used according to the manufacturer's instructions to introduce transprimer insertions at random positions throughout the *rns* gene in the plasmid pSS2192. The resulting mutagenized plasmids were transformed into *E. coli* strain XL-1. The several thousand resultant colonies were pooled and grown overnight with antibiotic selection for both the target plasmid and the

transprimer resistance markers. Plasmid DNA prepared from this overnight culture was digested with restriction endonuclease *PmeI* to remove all but 15 bp of the transprimer. The transprimer-free backbone was religated and transformed into *E. coli* strain XL-1 to amplify the library of plasmids containing 15 bp insertions. Plasmid DNA prepared from a pool of the resultant colonies was transformed into *E. coli* strain XL-1 containing the reporter plasmid pCooGFP. Bacteria harbouring mutant *rns* genes were discerned by observing their colony fluorescence under UV light. Colonies with altered fluorescence (relative to that seen with bacteria expressing the wild-type *rns* gene) were selected and the presence of a 15 bp insertion in their *rns* genes was verified by a PCR/restriction endonuclease screen. The precise sequence and location of the insertion was determined by DNA sequencing.

Western blotting. Western blotting was performed as described previously (Fagan *et al.*, 2008). Antisera were generated commercially by Convance. Anti-Rns sera were raised against a purified 6 × His-tagged Rns (derived from pHisRns) by three successive immunizations with 500 µg protein. The serum was absorbed against lysates of *E. coli* XL-1 Blue harbouring pQE30 and used in Western blotting at a dilution of 1 : 1000.

RNA isolation and reverse transcriptase PCR. Total RNA was isolated from mid-exponential-phase bacterial cultures using the RiboPure-Bacteria kit (Ambion). Contaminating genomic DNA was removed by DNase I treatment. RT-PCR was performed using 20 ng template total RNA and the Qiagen OneStep RT-PCR kit according to the manufacturer's instructions. The PCR products were visualized on a 2 % (w/v) agarose gel.

β-Galactosidase assay. The β-galactosidase levels were measured by using the method of Miller (1992). Assays were performed in duplicate on samples that had been grown in duplicate. The data were expressed as the means of four measurements. The SDs are indicated as error bars in each figure. Assays were performed independently at least twice.

Spectrofluorimetry. GFP levels were measured from supernatant fractions of bacteria lysed with CelLytic B, a non-denaturing detergent (Sigma). Duplicate 100 µl aliquots of each cell lysate were dispensed to a black, flat-bottomed 96-well plate. Fluorescence levels were measured in a Wallac 1420 Victor² multilabel counter set at an excitation wavelength of 485 nm and an emission wavelength of 535 nm. The values obtained were standardized with respect to optical density and expressed per ml culture. The final values were expressed as averages of all of the readings obtained for each strain. Assays were performed independently at least twice. A representative dataset is shown.

Maltose-binding protein (MBP) purification. *E. coli* strain KS1000 harbouring plasmids pRare and pMRns5 (or its mutant derivatives) was used for overexpression and purification of MBP::Rns fusions. Bacteria that had been induced for expression were harvested and resuspended in 1 ml sonication buffer [50 mM Tris/HCl pH 7.5, 10 % (v/v) sucrose, 100 mM NaCl, 1 mM EDTA] and then frozen at -80 °C for 2–3 h. The cell suspensions were thawed on ice and then lysed by sonication. The insoluble material was removed by centrifugation. The soluble supernatant fraction was collected and made up to a volume of 50 ml in column buffer (20 mM Tris/HCl, 200 mM NaCl, 1 mM EDTA, pH 7.4) prior to being loaded onto an amylose column. MBP fusion protein was purified using amylose resin (New England Biolabs) as described by the manufacturer.

Electrophoretic mobility shift assay (EMSA). The binding of purified MBP::Rns (or mutant derivatives) to the *coo* and *rns* promoter regions was analysed by EMSA using the LightShift

Table 1. Strains and plasmids used in this study

Strain/plasmid	Relevant details	Reference or source
Strains		
KS1000	F' <i>lacI^q lac⁺ pro⁺ / ara Δ(lac-pro) Δ(tsp)=Δ(prc)::Kan^R eda51::Tn10 (Tet^R) gyrA (Nal^R) rpoB thi-1 argI (am)</i>	Silber & Sauer (1994)
XL-1 Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac</i> [F' <i>proAB lac^dΔM15 Tn10 (Tet^R)</i>]	Stratagene
LMC10	A <i>lac</i> deletion restriction-negative derivative of the O6:K15:H16 ETEC-derived strain C921b-2, which was cured of the plasmid encoding Rns	Caron <i>et al.</i> (1989)
Plasmids		
pC102	pSS2192 derivative with 15 bp insertion encoding RCLNS after C102	This study
pCL1920	pSC101 replicon, low-copy-number cloning vector.	Lerner & Inouye (1990)
pCooGFP	Ap ^r Cm ^r ColE1 replicon, reporter plasmid containing <i>coo</i> promoter (from bp -393 to bp +62) fused to a promoterless <i>gfp</i> gene in plasmid pZEP08	This study
pCooGFP-2	Kan ^R plasmid p15A replicon, reporter plasmid containing <i>coo</i> promoter fused to a promoterless <i>gfp</i> gene	This study
pF205	pSS2192 derivative with 15 bp insertion encoding MFKHF after F205	This study
pGPS5	Kan ^R transprimer donor	New England Biolabs
pHisΔNACRS	pHisRns minus the nucleotides encoding the NACRS linker region	This study
pHisC102	pHisRns containing the 15 bp insertion from plasmid pC102	This study
pHisF205	pHisRns containing the 15 bp insertion from plasmid pF205	This study
pHisL198	pHisRns containing the 15 bp insertion from plasmid pL198	This study
pHisQ227	pHisRns containing the 15 bp insertion from pQ227	This study
pHisRns	<i>rns</i> ORF inserted into pQE-30	This study
pHisV253Δ	pHisRns containing the 15 bp insertion from pV253Δ	This study
pHisY242	pHisRns containing the 15 bp insertion from pY242	This study
pK256	pSS2192 with 15 bp insertion encoding CLNTK after K256	This study
pL198	pSS2192 with 15 bp insertion encoding FKQRL after L198	This study
pL221	pSS2192 with 15 bp insertion encoding CLNIL after L221	This study
pM-ΔNACRS	Derivative of pMRns5 minus the nucleotides encoding the NACRS linker region	This study
pMAL-c2	ColE1 replicon, MBP fusion protein expression vector	New England Biolabs
pM-C102	pMRns5 containing the 15 bp insertion from pC102	This study
pM-Q227	pMRns5 containing the 15 bp insertion from pQ227	This study
pM-RhaSLinker	Derivative of pMRns5 with the DNA sequence corresponding to the NACRS linker region of Rns replaced with nucleotides encoding the LENSASR linker of the RhaS protein	This study
pMRns5	<i>rns</i> ORF inserted into pMAL-c2 between <i>XmnI</i> and <i>BamHI</i> sites	This study
pQ227	pSS2192 with 15 bp insertion encoding CLNNQ after Q227	This study
pQE-30	ColE1 replicon, expression vector, T5 promoter/ <i>lac</i> operator element; 5' six-His tag coding sequence	Qiagen
pRare	p15A replicon, encodes rare tRNAs	Novagen
pRnsLacZ-2	<i>rns</i> promoter region inserted into pRW50	This study
pRW50	RK2 replicon, low-copy-number vector carrying a promoterless <i>lacZ</i> gene	Lodge <i>et al.</i> (1992)
pS239	pSS2192 derivative with 15 bp insertion encoding CLNTS after S239	This study
pS239Δ	pSS2192 derivative with 15 bp insertion encoding VOchTSS after S239*	This study
pSS2192	2 kb <i>XbaI/PstI</i> fragment from pCS200 containing <i>rns</i> gene and upstream area inserted into pCL1920	This study; Boylan <i>et al.</i> (1988)
pV253Δ	pSS2192 derivative with 15 bp insertion encoding VOchTGV after V253	This study
pY242	pSS2192 derivative with 15 bp insertion encoding CLNTY after Y242	This study
pZEP08	ColE1 replicon, encodes promoterless <i>gfp</i> downstream of MCS	Hautefort <i>et al.</i> (2003)

*Och, ochre stop codon.

Chemiluminescent EMSA kit (Pierce). To generate DNA probes for EMSAs, a 195 bp region of the *coo* promoter was amplified by PCR from plasmid pCooGFP-2 and a 434 bp region of the *rns* promoter was amplified by PCR from plasmid pSS2192. The primers in these PCRs had been biotinylated at their 5'-ends; therefore, they produced biotin-end-labelled PCR products. The PCR products were gel purified. Approximately 20 pg biotinylated DNA probe was incubated with a range of protein concentrations at 37 °C for 10 min in a

reaction buffer containing 10 mM Tris (pH 7.5), 50 mM KCl, 1 mM dithiothreitol, 2 ng poly(dI.dC) μl⁻¹, 0.1 μg BSA μl⁻¹, 0.025 % (v/v) Nonidet P-40 and 6.5 % (v/v) glycerol. Binding reactions were separated on 5 % non-denaturing polyacrylamide gels at 100 V for 65 min and then transferred to a nylon membrane at 80 V for 1 h. After the transfer, DNA was cross-linked to the membrane at 120 mJ cm⁻² using a UV-light cross-linker (Uvitec). The biotinylated DNA was detected with a streptavidin-horseradish peroxidase conjugate

and a chemiluminescent substrate as directed by the LightShift Chemiluminescent EMSA kit protocol (Pierce).

3D modelling of Rns. 3D models of the N terminus and C terminus of Rns were generated using PHYRE and Swiss-model, respectively (Bordoli *et al.*, 2009; Kelley & Sternberg, 2009).

RESULTS

Generation of a panel of Rns mutants

To characterize the structure–function relationship of Rns, the GPS-LS linker scanning system (NEB) was used to mutagenize the *rns*-encoding plasmid pSS2192. Mutant derivatives of this plasmid that displayed reduced activation of the *coo::gfp* fusion in plasmid pCooGFP (as defined by lower fluorescence) were screened for and sequenced (Table 2). The jpred and PROF programmes (Cuff *et al.*, 1998; Rost *et al.*, 2004) were used to obtain a consensus predicted secondary structure of Rns to better understand the regions where the insertions had occurred and to relate this to the effects of the insertions on Rns activity (Fig. 1).

A number of the mutated *rns* ORFs encoded truncated forms of Rns (indicated by Δ) as the 15 bp insertion generated by the GPS-LS mutagenesis system introduces a premature TAA stop codon in two of the six possible reading frames. Over 90 % of the protein was missing in truncates RnsI17 Δ and RnsH20 Δ , while most of the predicted DNA-binding domain had been lost in truncates

RnsI192 Δ and RnsI195 Δ . These truncated variants were not studied further.

Transactivation of the *coo* promoter by Rns is reduced by mutations in, or in the vicinity of, the predicted HTH motifs

The fluorescence levels in cultures encoding Rns derivatives with pentapeptide insertions in, or in the vicinity of, the protein's predicted HTH motifs were similar to those of cultures harbouring the empty vector pCL1920, i.e. less than 5 % of that detected in the presence of wild-type Rns (Fig. 2). This was also true for the two truncated variants tested, RnsS239 Δ and RnsV253 Δ . The mutant protein RnsC102 retained an intermediate level of activity at the *coo* promoter. Bacteria expressing this protein were 12 % as fluorescent as those expressing wild-type Rns.

The intracellular levels of the Rns mutant proteins were examined by Western immunoblotting with anti-Rns antiserum. While Rns protein was observed in lysates of *E. coli* strain XL-1 containing wild-type plasmid pSS2192, it was not detected in lysates prepared from cells harbouring the mutant derivatives of plasmid pSS2192. However, RT-PCR analysis confirmed that these mutants were at least transcribed (results not shown). Rns is expressed from its native, positively autoregulated promoter in pSS2192-based plasmids. The insertions may have prevented the mutants from auto-activating the *rns* promoter to produce the levels of protein necessary for detection by Western immunoblotting.

Therefore, a selection of the mutant *rns* genes were subcloned downstream of the T5 promoter in the His-tagged Rns expression vector pHisRns. His-tagged forms of mutants C102, L198, F205, Q227, Y242 and V253 Δ were detectable (by SDS-PAGE and/or Western immunoblotting), thus ruling out the possibility that the lack of *coo* promoter activation observed in the presence of these mutants was due to their mutations preventing expression of the encoded Rns protein (see supplementary Fig. S1, available with the online version of this paper). Accordingly, the ability of these Rns mutants to activate a *coo::gfp* fusion was assessed (Fig. 3). Again, RnsC102 exhibited a reduced ability to activate the *coo* promoter. Cultures of *E. coli* XL-1 containing plasmids pCooGFP-2 and pHisC102 were 27 % as fluorescent as those containing wild-type pHisRns. As before, the levels of expression from the *coo* promoter in the presence of the truncate RnsV253 Δ or the mutants with insertions in or near the predicted HTH motifs of Rns were similar to those found in the presence of the *rns*-free control.

Pentapeptide insertions also affect Rns autoregulation

Two pentapeptide insertion mutants were chosen for further analysis. RnsQ227 was selected as a representative of the mutants with insertions in the vicinity of the

Table 2. Non-functional insertion mutants resulting from mutagenesis of pSS2192

Mutant*	Pentapeptide insertion	Predicted structure at insertion site†
RnsI17 Δ ‡	IVOchTI§	Unstructured, N-terminal domain
RnsH20 Δ	KYVOchT	Unstructured, N-terminal domain
RnsC102	RCLNS	Unstructured, N-terminal domain
RnsI192 Δ	VOchTEI	Recognition helix of HTH1
RnsI195 Δ	NVOchTR	Recognition helix of HTH1
RnsL198	FKQRL	Recognition helix of HTH1
RnsF205	MFKHF	α -Helix between HTH1 and 2
RnsL221	CLNIL	α -Helix between HTH1 and 2
RnsQ227	CLNNQ	First helix of HTH2
RnsS239	CLNTS	Turn of HTH2
RnsS239 Δ	VOchTSS	Turn of HTH2
RnsY242	CLNTY	Recognition helix of HTH2
RnsV253 Δ	VOchTGV	α -Helix downstream of HTH2
RnsK256	CLNTK	α -Helix downstream of HTH2

*Mutant Rns proteins are named based on the amino acid residue to the N-terminal side of the insertion position.

†Structure predictions from jpred and PROF.

‡Truncated proteins are indicated by Δ .

§Och, ochre stop codon.

Fig. 1. Predicted secondary structure of the Rns protein. The numbered amino acid sequence of Rns is shown on the top line. Secondary structure predictions from jpred and PROF are shown on the second and third lines, respectively. H, α -helix; E, β -strand. The consensus prediction is depicted on the bottom line. α , α -helix; β , β -strand. Helices predicted to be involved in the HTH motifs are shown in grey. There was no consensus prediction for the extreme N terminus of Rns but subsequent analysis revealed that this area is predicted to be unstructured or disordered.

predicted HTH motifs of Rns and RnsC102 was selected due to its unique status of retaining a partial level of activity at the *coo* promoter. DNA fragments containing the 15 bp insertions encoding the relevant mutations were subcloned into the MBP::Rns fusion expression vector pMRns5 to create the plasmids pM-Q227 and pM-C102. Fusion of MBP to the amino terminus of Rns does not affect the activity of the protein *in vivo* or *in vitro*, but does improve its solubility such that it can be overexpressed and purified (Munson & Scott, 1999).

The pRnsLacZ-2 plasmid contains a transcriptional fusion of the *rns* promoter to the *lacZ* reporter gene. This plasmid, along with plasmid pMAL-c2, pMRns5, pM-C102 or pM-Q227, was transformed into the ETEC *lac* deletion strain LMC10 to assess the ability of the mutants to activate the *rns* promoter. The levels of β -galactosidase activity in cells containing the positive control plasmid pMRns5 were consistently 1.6-fold higher than those in cells containing the negative empty vector control pMAL-c2 (Fig. 4). This is due to positive regulation of the *rns* promoter by the MBP::Rns fusion protein. The RnsC102 insertion mutant also activated transcription from the *rns* promoter, but to a level lower than wild-type Rns. The β -galactosidase expression levels in cells containing MBP::RnsC102 were approximately 1.3-fold less than those in cells containing

the MBP::Rns fusion protein ($P < 0.005$). The RnsQ227 mutant did not activate *rns* transcription. β -Galactosidase activity levels in cells containing plasmid pM-Q227 were similar to those in cells containing plasmid pMAL-c2. Therefore, RnsC102 and RnsQ227 displayed similar levels of activity at the *rns* promoter as they had at the *coo* promoter.

A pentapeptide insertion in the vicinity of HTH2 of Rns eliminates DNA binding by the protein

EMSAs were performed to study the interaction of Rns and RnsQ227 with the *coo* and *rns* promoters and thus determine if the inactivity of the mutant was due to the pentapeptide insertion adversely affecting its ability to bind to DNA. A 195 bp region of the *coo* promoter (positions -196 to -2 relative to the transcription start site) that includes the two Rns-binding sites previously identified by DNase I footprinting (Munson & Scott, 1999) was used as the *coo* promoter DNA probe. Three major protein-DNA complexes were detected when EMSAs were performed with this probe and a range of concentrations of the wild-type MBP::Rns fusion protein (Fig. 5a). The highest mobility complex results from Rns initially occupying binding site I or binding site II alone. The next highest

Fig. 2. Transactivation of the *coo* promoter by Rns and mutant derivatives. Fluorescence levels of *E. coli* strain XL-1 cultures harbouring pCooGFP and the empty vector negative control pCL1920 or a pSS2192 derivative expressing the indicated mutant Rns were calculated as percentages of the level of fluorescence of cultures expressing wild-type Rns. The data represent averages of duplicate cultures each measured in triplicate. Error bars indicate SD. Measurements were performed independently at least twice; a representative dataset is shown.

Fig. 3. Transactivation of the *coo* promoter by His-tagged Rns and mutant derivatives. Fluorescence levels of *E. coli* strain XL-1 cultures harbouring pCooGFP-2 and the empty vector control pQE-30 or a pHisRns derivative expressing the indicated mutant Rns were calculated as percentages of the level of fluorescence of cultures containing wild-type His-tagged Rns. The data represent averages of duplicate cultures measured in triplicate. Error bars indicate SD. Measurements were performed independently at least twice; a representative dataset is shown.

mobility complex is due to simultaneous occupation of both of these sites at the *coo* promoter (V. Mahon and S. G. J. Smith, unpublished observations). It is likely that the third, low mobility complex is due to non-specific interactions that occur in the presence of high protein concentrations. RnsQ227 did not bind specifically to the *coo* promoter probe (Fig. 5a). As determined by densitometry, the amount of *coo* DNA that remained unbound in the presence of 171 nM MBP::RnsQ227 fusion protein was equal to the amount of probe present in the protein-free control. At the highest concentration of protein tested – 514 nM MBP::RnsQ227 fusion protein – only the non-specific protein–DNA complex was formed.

A 434 bp sequence of the *rns* promoter (positions –305 to +129 relative to the transcription start site), which includes the three sites shown by DNase I footprinting to be bound by Rns (Munson & Scott, 2000), was used as the *rns* promoter DNA probe. Three protein–DNA complexes were observed when EMSAs were performed with this probe and a range of concentrations of wild-type MBP::Rns fusion protein (Fig. 5b). It can be postulated that the complexes are a result of Rns binding to one site initially, then at higher protein concentrations to two sites and finally at even higher protein concentrations binding to all three of its sites at the *rns* promoter. Protein–DNA complexes were not detected when EMSAs were conducted with MBP::RnsQ227 fusion protein and *rns* promoter

Fig. 4. Transactivation of the *rns* promoter by MBP::Rns fusion protein and mutant derivatives. The β-galactosidase activities of ETEC strain LMC10 harbouring plasmid pRnsLacZ-2 and the MBP::Rns-expressing plasmid pMRns5, the empty vector control plasmid pMAL-c2 or a plasmid pMRns5 derivative expressing the indicated mutant were measured. The data represent averages of duplicate assays on duplicate cultures. Error bars indicate SD. Measurements were performed independently at least twice; a representative dataset is shown. Statistical significance is indicated by **, where $P < 0.005$.

Fig. 5. Interaction of MBP::Rns and MBP::RnsQ227 fusion proteins with DNA. EMSAs were carried out with approximately 20 pg of (a) a biotinylated *coo* promoter DNA probe or (b) a biotinylated *rns* promoter DNA probe and a range of concentrations of MBP::Rns or MBP::RnsQ227 fusion proteins as indicated at the top of each panel. The unbound DNA fragments and protein-DNA complexes are indicated on the left.

DNA even in the presence of 171 nM MBP::RnsQ227 fusion protein (Fig. 5b). Thus, the pentapeptide insertion after position Q227 of Rns results in a mutant protein that has lost the ability to bind to DNA and, therefore, cannot activate transcription from the *coo* and *rns* promoters.

A predicted area of disorder around residue C102 of Rns is essential for full activity

The uncharacterized area of Rns around residue C102 is not predicted to be of α -helical or β -sheet structure and the pentapeptide insertion within RnsC102 is not predicted to introduce a novel α -helix or β -sheet in the protein. To better understand why this insertion had reduced but not abolished protein activity, the effects of more subtle point mutations in the region were assessed. Site-directed mutagenesis was employed to individually alter the C102 residue itself and the neighbouring residue R103 to alanine (there is already a naturally occurring alanine residue at position 101 of Rns). Neither of these alanine substitutions affected the ability of Rns to transactivate the *coo* promoter.

As secondary structure predictions could not provide a reason for the effects of the mutations in the area around

the C102 residue of Rns, alternative prediction programmes were employed. GlobPlot (<http://globplot.embl.de>) is a web-based tool that plots the tendency of protein sequences to be ordered or disordered. GlobPlot analysis of the primary sequence of Rns revealed that the extreme N- and C-termini of Rns and the region encompassing residues 100–104 are predicted to be disordered. DRIP-PRED (<http://www.sbc.su.se/~maccallr/disorder>) and Spritz (<http://distill.ucd.ie/spritz>), two other computational methods for protein disorder prediction, also predict that an area of disorder is present around residues 100–104 of Rns (Vullo *et al.*, 2006). However, when the GlobPlot analysis was performed on the sequence of Rns including the pentapeptide insertion after position C102, the disorder in this region was no longer predicted to be present. In contrast, the single amino acid changes of C102A and R103A were not predicted to disrupt the disorder in this area. It is possible that to be fully functional as a transcriptional activator, it may be necessary for residues 100–104 of Rns to be disordered.

To determine the importance of predicted protein disorder for Rns activity a mutant protein, Rns Δ NACRS, was produced in which the amino acid sequence that comprises

the predicted disordered region between positions 100 and 104 was deleted. To test the possibility that the identity of the amino acids in the potentially disordered region is not critical, another Rns derivative was created in which LENSASR, the seven-residue linker sequence of the AraC family member RhaS of *E. coli*, was substituted for Rns residues 100–104. The RhaS linker was selected because in its native setting, the LENSASR sequence has been suggested to function only to flexibly link the two functional domains of RhaS. Furthermore, a derivative of the related RhaR protein containing the RhaS linker in place of its own linker sequence was functional with only small deficits in its activity (Kolin *et al.*, 2007).

GlobPlot analysis revealed that replacing the NACRS residues of Rns with the LENSASR linker sequence restored the predicted disorder in this region. The fluorescence levels in cultures of *E. coli* XL-1/pCooGFP-2 harbouring the deletion mutant encoded by pM- Δ NACRS were similar to those of cultures harbouring the empty vector pMAL-c2 ($21 \pm 1\%$ and $17 \pm 3\%$ of the level of fluorescence in the presence of wild-type Rns, respectively). Thus, deletion of

the potentially disordered area of Rns eliminated the ability of the protein to activate the *coo* promoter. Cultures of *E. coli* strain XL-1 harbouring plasmids pCooGFP-2 and pM-RhaSLinker were as fluorescent as those containing pMRns5 (the relative fluorescence levels of the cultures were $108 \pm 3\%$ and $100 \pm 3\%$, respectively). Therefore, the Rns derivative containing the RhaS linker appeared to be fully functional at the *coo* promoter.

EMSAs were performed to study the binding of purified MBP fusions of these Rns derivatives to the *coo* promoter (Fig. 6a). The RnsC102 insertion mutant bound to *coo* promoter DNA. However, it did so with less affinity than wild-type Rns. Approximately 20% (as determined by densitometry) of the *coo* promoter probe remained unbound in the presence of 171 nM MBP::RnsC102 fusion protein, while at the same concentration of the MBP::Rns fusion, no free probe remained. MBP::RnsC102 fusion protein also displayed a greater tendency than wild-type Rns to form the low-mobility non-specific protein–DNA complex with the *coo* promoter. In contrast, the MBP::Rns Δ NACRS fusion protein did not specifically

Fig. 6. Interaction of MBP::Rns or its mutant derivatives with *coo* and *rns* promoter DNA. EMSAs were carried out with approximately 20 pg of (a) a biotinylated *coo* promoter DNA probe or (b) a biotinylated *rns* promoter DNA probe and a range of concentrations of MBP::Rns or its mutant derivatives as indicated at the top of each panel. The unbound DNA fragments and protein–DNA complexes are indicated on the left.

bind the *coo* promoter DNA probe. Even at the highest concentration of the deletion mutant tested, the majority of the probe remained unbound and only a small amount of the non-specific complex had been formed. However, MBP::Rns with RhaS linker fusion protein bound the *coo* promoter DNA probe and formed specific protein–DNA complexes at a similar rate to wild-type MBP::Rns fusion protein.

Similar results were observed when the EMSAs were repeated with the *rns* promoter probe (Fig. 6b). The MBP::RnsC102 fusion protein bound to *rns* promoter DNA but with less affinity than wild-type MBP::Rns fusion protein. Unlike wild-type MBP::Rns fusion protein, in the presence of 17 nM MBP::RnsC102 fusion protein only the first protein–DNA complex was formed and at 68 nM MBP::RnsC102 fusion protein, approximately 40 % (as determined by densitometry) of the *rns* probe remained unbound. While the MBP::Rns with RhaS linker fusion protein was found to complex with the *rns* promoter DNA in a manner similar to that observed for MBP::Rns fusion, the MBP::Rns Δ NACRS fusion protein did not form protein–DNA complexes.

Therefore, the EMSAs revealed that disruption of the NACRS sequence of Rns reduces the affinity of the protein for DNA, while elimination of the sequence results in a mutant protein that is incapable of binding to DNA and consequently cannot activate transcription. However, replacement of the NACRS sequence with an alternative putative disordered region produced a form of Rns that displayed wild-type activity at both the *coo* and *rns* promoters. This supports the theory that Rns activity is dependent on the presence of an area of disorder in the vicinity of residues 100–104.

Homology modelling of Rns was attempted to determine the structure of the region in the vicinity of aa 100–104. Using both PHYRE and Swiss-model, we were unable to model the entire structure of Rns. In particular, no structure could be determined for aa 73–154. However, we could generate models of the N terminus and C terminus with these tools, respectively. The N terminus from aa 15–72 was predicted to have a Cupin-like fold (see Supplementary Figs S2 and S3, available with the online version of this paper). The C terminus (from amino acids 155–264) was modelled on the ToxT protein of *Vibrio cholerae* (Lowden *et al.*, 2010) and broadly agrees with our secondary structure predictions (see Supplementary Figs S4 and S5, available with the online version of this paper).

DISCUSSION

In the absence of a 3D structure, mutagenesis was employed to gain insight into the structure–function relationship of Rns. The C-terminal pentapeptide insertions had a severe negative effect on the activity of Rns at the *coo* promoter, signifying that the affected areas (two predicted HTH motifs, a long α -helix between these motifs

and the C-terminal α -helix of the protein) are vital for Rns activity. It is postulated that these areas are involved in DNA binding. This was confirmed for RnsQ227, which contains an insertion immediately upstream of the predicted HTH2 and did not bind specifically to either the *coo* or the *rns* promoter DNA probes. The affected areas may participate directly in DNA binding, i.e. form specific interactions with nucleotides in Rns-binding sites. This is most likely for the two predicted HTHs.

Therefore, Rns may use both HTHs to bind DNA as has been proposed or demonstrated for the AraC family members ToxT (Childers *et al.*, 2007), VirF (Porter & Dorman, 2002) and PerA (Porter *et al.*, 2004). The other C-terminal α -helices of Rns may not be directly involved in DNA binding. It is proposed that, like the α -helix observed between the two HTHs in the crystal structure of the AraC family protein Rob (Kwon *et al.*, 2000), they correctly orientate or stabilize the HTHs to ensure they are in the configuration necessary to interact with DNA.

All truncated variants were inactive. This is unsurprising in mutants lacking over 90 % of the protein (RnsI17 Δ and RnsH20 Δ) or most of the predicted DNA-binding domain (RnsI192 Δ and RnsI195 Δ). That the RnsS239 Δ variant was inactive is further proof that HTH2 is critical for Rns function. However, the loss of 26 aa may disrupt the folding of the protein's entire C-terminal domain. Only the last 12 residues of Rns are missing in the RnsV253 Δ mutant. The inactivity of this truncated variant supports the theory that the final predicted α -helix of Rns plays an essential role.

The sole mutant with an N-terminal pentapeptide insertion, RnsC102, had a reduced ability to activate the *coo* and *rns* promoters and a decreased affinity for DNA binding. The insertion after residue C102 had eliminated an area of predicted disorder within Rns. The effects of further mutagenesis and deletion of this potentially disordered area, the NACRS amino acid sequence, indicate that to be fully functional, Rns requires a disordered region between its N- and C-terminal domains. It is postulated that this area of Rns forms, or is part of, a flexible interdomain linker that is necessary to correctly orientate the N- and C-terminal regions of the protein to enable them to participate in their functional roles. The NACRS sequence of Rns has some features in common with the flexible linkers of the AraC, RhaR and RhaS proteins (Eustance & Schleif, 1996; Kolin *et al.*, 2007). The activity of other AraC-like proteins may also depend on the presence of a central disordered/flexible connecting sequence. Residues 100–131 of Rns have been previously suggested to probably function as a flexible linker (Basturea *et al.*, 2008). This study has specifically identified the NACRS sequence that may comprise the minimal linker region of Rns or at least a segment within a larger linker in which disorder or flexibility is vital.

In summary, this analysis has defined regions critical for the activity of Rns and serves as a basis for future research on this protein and other AraC family members.

ACKNOWLEDGEMENTS

This work was supported by an Irish Research Council for Science, Engineering and Technology (IRCSET) Studentship and an Enterprise Ireland Basic Research Grant (SC-2002-535).

REFERENCES

- Anantha, R. P., McVeigh, A. L., Lee, L. H., Agnew, M. K., Cassels, F. J., Scott, D. A., Whittam, T. S. & Savarino, S. J. (2004). Evolutionary and functional relationships of colonization factor antigen I and other class 5 adhesive fimbriae of enterotoxigenic *Escherichia coli*. *Infect Immun* **72**, 7190–7201.
- Basturea, G. N., Bodero, M. D., Moreno, M. E. & Munson, G. P. (2008). Residues near the amino terminus of Rns are essential for positive autoregulation and DNA binding. *J Bacteriol* **190**, 2279–2285.
- Bodero, M. D., Pilonieta, M. C. & Munson, G. P. (2007). Repression of the inner membrane lipoprotein NlpA by Rns in enterotoxigenic *Escherichia coli*. *J Bacteriol* **189**, 1627–1632.
- Bodero, M. D., Harden, E. A. & Munson, G. P. (2008). Transcriptional regulation of subclass 5b fimbriae. *BMC Microbiol* **8**, 180.
- Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* **4**, 1–13.
- Boylan, M., Coleman, D. C., Scott, J. R. & Smyth, C. J. (1988). Molecular cloning of the plasmid-located determinants for CS1 and CS2 fimbriae of enterotoxigenic *Escherichia coli* of serotype O6:K15:H16 of human origin. *J Gen Microbiol* **134**, 2189–2199.
- Caron, J., Coffield, L. M. & Scott, J. R. (1989). A plasmid-encoded regulatory gene, *rns*, required for expression of the CS1 and CS2 adhesins of enterotoxigenic *Escherichia coli*. *Proc Natl Acad Sci U S A* **86**, 963–967.
- Childers, B. M., Weber, G. G., Prouty, M. G., Castaneda, M. M., Peng, F. & Klose, K. E. (2007). Identification of residues critical for the function of the *Vibrio cholerae* virulence regulator ToxT by scanning alanine mutagenesis. *J Mol Biol* **367**, 1413–1430.
- Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. & Barton, G. J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**, 892–893.
- Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **16**, 6127–6145.
- Eustance, R. J. & Schleif, R. F. (1996). The linker region of AraC protein. *J Bacteriol* **178**, 7025–7030.
- Fagan, R. P., Lambert, M. A. & Smith, S. G. (2008). The Hek outer membrane protein of *Escherichia coli* strain RS218 binds to proteoglycan and utilizes a single extracellular loop for adherence, invasion, and autoaggregation. *Infect Immun* **76**, 1135–1142.
- Froehlich, B., Husmann, L., Caron, J. & Scott, J. R. (1994). Regulation of *rns*, a positive regulatory factor for pili of enterotoxigenic *Escherichia coli*. *J Bacteriol* **176**, 5385–5392.
- Gaastra, W. & Svennerholm, A. M. (1996). Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends Microbiol* **4**, 444–452.
- Gallegos, M. T., Schleif, R., Bairoch, A., Hofmann, K. & Ramos, J. L. (1997). AraC/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev* **61**, 393–410.
- Hautefort, I., Proenca, M. J. & Hinton, J. C. (2003). Single-copy green fluorescent protein gene fusions allow accurate measurement of *Salmonella* gene expression in vitro and during infection of mammalian cells. *Appl Environ Microbiol* **69**, 7480–7491.
- Ibarra, J. A., Pérez-Rueda, E., Segovia, L. & Puente, J. L. (2007). The DNA-binding domain as a functional indicator: the case of the AraC/XylS family of transcription factors. *Genetica* **133**, 65–76.
- Kelley, L. A. & Sternberg, M. J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363–371.
- Kolin, A., Jevtic, V., Swint-Kruse, L. & Egan, S. M. (2007). Linker regions of the RhaS and RhaR proteins. *J Bacteriol* **189**, 269–271.
- Kwon, H. J., Bennik, M. H., Demple, B. & Ellenberger, T. (2000). Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA. *Nat Struct Biol* **7**, 424–430.
- Lerner, C. G. & Inouye, M. (1990). Low copy number plasmids for regulated low-level expression of cloned genes in *Escherichia coli* with blue/white insert screening capability. *Nucleic Acids Res* **18**, 4631.
- Lodge, J., Fear, J., Busby, S., Gunasekaran, P. & Kamini, N. R. (1992). Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. *FEMS Microbiol Lett* **74**, 271–276.
- Lowden, M. J., Skorupski, K., Pellegrini, M., Chiorazzo, M. G., Taylor, R. K. & Kull, F. J. (2010). Structure of *Vibrio cholerae* ToxT reveals a mechanism for fatty acid regulation of virulence genes. *Proc Natl Acad Sci U S A* **107**, 2860–2865.
- Miller, J. H. (1992). *A Short Course in Bacterial Genetics*. Cold Spring Harbour, NY: Cold Spring Harbour Laboratory.
- Munson, G. P. & Scott, J. R. (1999). Binding site recognition by Rns, a virulence regulator in the AraC family. *J Bacteriol* **181**, 2110–2117.
- Munson, G. P. & Scott, J. R. (2000). Rns, a virulence regulator within the AraC family, requires binding sites upstream and downstream of its own promoter to function as an activator. *Mol Microbiol* **36**, 1391–1402.
- Munson, G. P., Holcomb, L. G. & Scott, J. R. (2001). Novel group of virulence activators within the AraC family that are not restricted to upstream binding sites. *Infect Immun* **69**, 186–193.
- Munson, G. P., Holcomb, L. G., Alexander, H. L. & Scott, J. R. (2002). *In vitro* identification of Rns-regulated genes. *J Bacteriol* **184**, 1196–1199.
- Murphree, D., Froehlich, B. & Scott, J. R. (1997). Transcriptional control of genes encoding CS1 pili: negative regulation by a silencer and positive regulation by Rns. *J Bacteriol* **179**, 5736–5743.
- Pilonieta, M. C., Bodero, M. D. & Munson, G. P. (2007). CfaD-dependent expression of a novel extracytoplasmic protein from enterotoxigenic *Escherichia coli*. *J Bacteriol* **189**, 5060–5067.
- Porter, M. E. & Dorman, C. J. (2002). *In vivo* DNA-binding and oligomerization properties of the *Shigella flexneri* AraC-like transcriptional regulator VirF as identified by random and site-specific mutagenesis. *J Bacteriol* **184**, 531–539.
- Porter, M. E., Mitchell, P., Roe, A. J., Free, A., Smith, D. G. E. & Gally, D. L. (2004). Direct and indirect transcriptional activation of virulence genes by an AraC-like protein, PerA from enteropathogenic *Escherichia coli*. *Mol Microbiol* **54**, 1117–1133.
- Qadri, F., Svennerholm, A. M., Faruque, A. S. & Sack, R. B. (2005). Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin Microbiol Rev* **18**, 465–483.
- Rost, B., Yachdav, G. & Liu, J. (2004). The PredictProtein server. *Nucleic Acids Res* **32**, W321–W326.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbour, NY: Cold Spring Harbour Laboratory.

Silber, K. R. & Sauer, R. T. (1994). Deletion of the *prc* (*tsp*) gene provides evidence for additional tail-specific proteolytic activity in *Escherichia coli* K-12. *Mol Gen Genet* **242**, 237–240.

Turner, S. M., Scott-Tucker, A., Cooper, L. M. & Henderson, I. R. (2006). Weapons of mass destruction: virulence factors of the global killer enterotoxigenic *Escherichia coli*. *FEMS Microbiol Lett* **263**, 10–20.

Vullo, A., Bortolami, O., Pollastri, G. & Tosatto, S. C. E. (2006). Spritz: a server for the prediction of intrinsically disordered regions in protein sequences using kernel machines. *Nucleic Acids Res* **34**, W164–W168.

Edited by: U. Dobrindt