Variable length tandem repeat polyglutamine sequences in the flexible tether region of the Tsr chemotaxis receptor of *Escherichia coli*

Sijana H. Dzinic, Meghna Shukla, Ilir Mandija, Tammy S. Ram and Jeffrey L. Ram

Department of Physiology, Wayne State University, Detroit, MI 48201, USA

Correspondence

Jeffrey L. Ram

jeffram@med.wayne.edu

Methyl-accepting chemotaxis proteins (MCPs) are receptors that play an important role in bacterial chemotaxis. Methylation of Tsr, the MCP that mediates chemotaxis towards serine in *Escherichia coli*, is thought to be facilitated by binding of the methyltransferase to a flexible tether region at the C-terminal end of Tsr. This study analysed natural length variants of the tether that occur in *E. coli* due to genetic instability in tandem repeat DNA sequences that code for glutamyl (Q) residues, creating polyQ sequences of variable lengths in the tether region. The *tsr* gene of *E. coli* K-12 (strain MG1655) codes for 4Q at the beginning of its 35 aa tether region. The tether varies in length from 35 to 47 residues among pathogenic and non-pathogenic strains of *Escherichia*, *Shigella* spp., *Salmonella*, *Yersinia* and *Photobacterium*. Among previous sequences, *Escherichia* and *Shigella* mostly have 4Q and 7Q variants, and one strain (*E. coli* HS) has 10Q. In *E. coli* isolated from 50 humans and 75 animals (dogs, cats, horses, birds, etc.), polyQ up to 13Q (44 aa tether) were identified (6 strains); relative frequencies were 7Q (~77 % of the total) >4Q (14 %) >13Q (5 %) >10Q (4 %). Phylogenetic analysis revealed that *E. coli* strains with 10Q or 13Q largely fell within two clusters. Serine chemotaxis was not significantly different among 7Q, 10Q and 13Q strains, and was comparable to chemotaxis in the frequently studied K-12 strain. These results are consistent with models indicating that polyQ sequences from 7Q to 13Q are flexible, and that longer tethers, within this range, would not change the precision of adaptation mediated by methylation. Studies of this naturally variable polyQ region in *E. coli* may also have relevance to mechanisms that mediate polyQ instability in human genetic diseases.

Received 2 January 2008

Revised 18 April 2008

Accepted 23 April 2008

INTRODUCTION

The receptors that mediate one form of chemotactic behaviour in bacteria have been an important focus of study since Adler and colleagues first described them in *Escherichia coli* (Adler, 1969; Springer *et al.*, 1979). *E. coli* generally have five different types of receptors (Tar, Tsr, Trg, Tap and Aer) that differ in their sensitivities to different attractants and repellents, and form mixed clusters of homodimers (Studdert & Parkinson, 2005) at the cell poles (Maddock & Shapiro, 1993) in proportion to their relative abundance in the cells (Studdert & Parkinson, 2004; Gosink *et al.*, 2006). As reviewed recently (Hazelbauer *et al.*, 2008), binding of an attractant chemical

to the periplasmic sensing domain of a chemoreceptor molecule decreases the activity of a histidine kinase (CheA) coupled, via the CheW protein, to its cytoplasmic signalling domain. Decreased CheA activity lowers the phosphorylation state of the flagellar motor regulator, CheY, thereby increasing the probability of counter-clockwise flagellar rotation, which mediates forward swimming. The signalling properties of the chemoreceptors, except for Aer, are also regulated by methylation and demethylation of several specific glutamyl residues in their cytoplasmic domain, mediated by the activity of the methyltransferase CheR and the methylesterase/deamidase CheB. Hence, the receptors for bacterial chemotaxis are generically known as methyl-accepting chemotaxis proteins (MCPs). Methylation shifts MCP molecules toward a low attractant affinity kinase-on state; demethylation shifts them toward a high attractant affinity kinase-off state. This sensory adaptation system enables *E. coli* to mediate chemotactic responses over a broad range of attractant concentrations.

Abbreviations: MCP, methyl-accepting chemotaxis protein; UTI, urinary tract infection.

The GenBank accession numbers for the sequences reported in this paper are EU658753–EU658757 and EU637473–EU637575.

Supplementary sequence data and two supplementary figures are available with the online version of this paper.

In *E. coli*, the *tsr* gene encodes the serine-binding receptor, Tsr. The transmembrane Tsr molecule is a homodimer with a number of discrete structure–function regions (Fig. 1). The HAMP domain at the cytoplasmic side of the membrane converts ligand-induced conformational changes in the periplasmic region of the receptor to conformational changes in the cytoplasmic portion of the molecule that regulate its methylation and flagellar signalling behaviour. The cytoplasmic tip of the receptor interacts with CheW and CheA to form the ternary signalling complexes. The methylation sites for sensory adaptation lie between the HAMP domain and the signalling region. The activities of both CheR (Wu *et al.*, 1996) and CheB (Barnakov *et al.*, 1999) are thought to be facilitated by their binding to a conserved sequence, -NWETF, located at the C terminus of each receptor subunit. The C-terminal CheR/CheB-binding site is joined to the signalling region by a flexible segment that acts as a tether, thereby constraining the location of the adaptation enzymes to the vicinity of the methylation sites in the same and neighbouring MCP molecules (Lai & Hazelbauer, 2005). The flexible tether is a relatively recently evolved structure in Tsr and occurs only in proteobacteria, e.g. *E. coli*, and a few other species (Alexander & Zhulin, 2007). Crystallographic studies (Kim *et al.*, 1999) and secondary-structure analyses (Le Moual & Koshland, 1996) indicate

that the methylation and signalling regions form a dimeric four-helix supercoiled bundle.

The flexible tether region is not as well conserved in methyl-accepting chemotaxis receptors as the methylation and signalling regions. As noted above, some types of MCPs lack the tether region altogether. Even among various bacterial species whose Tsr receptors have the tether, its length varies, from as short as 35 aa in *Shigella boydii* (counting I517 as the beginning of the tether) to as long as 47 aa in *Photobacterium luminescens* (see Supplementary Fig. S1, available with the online version of this paper). In the common laboratory strain of *E. coli* K-12 (strain MG1655), in which chemotaxis has been most extensively studied, the tether region is 35 aa and one might wonder whether this relatively short sequence is a special adaptation of *E. coli* and *Shigella*. However, as this paper will show, longer tether lengths also occur in *E. coli*, most likely due to genetic instability caused by the occurrence of a tandem repeat sequence coding for multiple glutaminyl (Q) residues in the tether region. In *E. coli* K-12 MG1655, *Shigella boydii* and *Salmonella typhimurium*, Tsr has a sequence of 4Q near the beginning of the 35 aa tether. The present study analysed variation in this polyQ region, comparing previously sequenced genomes, then identifying variants that have longer polyQ regions in undomesticated strains of *E. coli* isolated from faeces. This study also investigated whether chemotaxis works well in strains of *E. coli* with tethers up to 44 aa residues in length.

Fig. 1. Secondary structure of bacterial Tsr methyl-accepting chemotaxis receptors, modified from Kim *et al.* (1999), Le Moual & Koshland (1996) and Chalah & Weis (2005). This block diagram shows one monomer of the homodimer structure, with locations of predicted alpha-helical regions as rectangles, and turns, coils and flexible strands as curves and lines. Methylation sites are represented by triangles, and the relative locations of the -QQQQ- and -NWETF sequences in the tether region are indicated. Methylation sites Q297 and Q311 are initially expressed as glutaminyl residues but deamidated to glutamyl residues by CheB (Rice & Dahlquist, 1991).

METHODS

Bacterial strains. This study used animal and human faecal strains of *E. coli* that we previously isolated (Ram *et al.*, 2004, 2007). Each strain was isolated from a different individual human or animal. After culture, streaking and selection of isolated colonies, strains were verified as *E. coli* in Colilert-18 (Hach) and by sequencing various genes, particularly the β -glucuronidase gene, *uidA* (Ram *et al.*, 2004, 2007). *E. coli* HS was provided by James Nataro (University of Maryland); it is 'a human commensal isolate (Levine *et al.*, 1978) [that colonizes] the human gastrointestinal tract with no overt signs of disease' [GenBank annotation for ZP_00708242 (NCBI Microbial Genomes Annotation Project, 2005)]. The accession numbers for sequences that are indexed in GenBank that are shown in Fig. 2 and Supplementary Fig. S1 are given in Supplementary sequence data. Strain MG1655 was obtained from Harry L.T. Mobley (Department of Microbiology and Immunology, University of Michigan). Strain UU1624 (Δ *tsr*-7028 Δ *tap*-3654 Δ *trg*-100 Δ *aer*-1) was obtained from John S. Parkinson (Department of Biology, University of Utah). All strains were stored at -80°C in Colilert-18, supplemented with 14% (v/v) glycerol.

PCR with primers flanking the polyQ region. In order to detect sequence variants differing in length by as few as 9 bases on agarose gels, primers flanking the polyQ region of the *tsr* gene were designed using Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), to give amplicons of <200 bp in length. Sequences of these 'flanking primers' were *tsr*1437F, 5'-ATGGACCGGTAACTCAACA-3'; and *tsr*1584R, 5'-TTACCACAGCCGATGTTTCA-3'. PCR conditions were initial denaturation, 3 min at 94°C ; 30 cycles of denaturation (30 s, 94°C), primer annealing (30 s, 60°C) and

Strain or species	DNA sequence	no. of Q
<i>E. coli</i> K12 MG1655	TTCCGGATTTCAGCAACAGCAG-----CCTGAACA	4
<i>E. coli</i> O157:H7	TTCCGGATTTCAGCAACAGCAGCAACAGCAG-----CCTGAACA	7
<i>E. coli</i> CFT073	TTCCGGATTTCAGCAACAGCAGCAACAGCAG-----CCTGAACA	7
<i>E. coli</i> UTI89	TTCCGGATTTCAGCAACAGCAG-----CCTGAACA	4
<i>E. coli</i> 536	TTCCGGATTTCAGCAACAGCAGCAACAGCAG-----CCTGAACA	7
<i>S. boydii</i>	TTCCGGATTTCAGCAACAGCAG-----CCTGAACA	4
<i>S. flexneri</i>	TTCCGGATTTCAGCAACAGCAGCAACAGCAG-----CCTGAACA	7
<i>E. coli</i> HS	TTCCGGATTTCAGCAACAGCAGCAACAGCAGCAACAGCAG-----CCTGAACA	10
<i>E. coli</i> RamLab137	TTCCGGATTTCAGCAACAGCAG-----CCTGAACA	4
<i>E. coli</i> RamLab378	TTCCGGATTTCAGCAACAGCAGCAACAGCAG-----CCTGAACA	7
<i>E. coli</i> RamLab133	TTCCGGATTTCAGCAACAGCAGCAACAGCAG-----CCTGAACA	6
<i>E. coli</i> RamLab366	TTCCGGATTTCAGCAACAGCAGCAACAGCAGCAACAGCAG-----CCTGAACA	10
<i>E. coli</i> RamLab564	TTCCGGATTTCAGCAACAGCAGCAACAGCAGCAACAGCAGCAACAGCAGCTGAACA	13

Amino acid sequence: F R I Q Q Q Q Q Q Q Q Q Q Q Q R E T

Fig. 2. DNA sequences coding for the polyQ region and neighbouring three residues of Tsr in various strains of *E. coli* and *Shigella* spp., based on sequences in GenBank (top eight sequences), and on sequences of *E. coli* strains isolated in the Ram laboratory. The longest sequence is for *E. coli* RamLab strain 564, which codes for a sequence of 13 glutamyl residues, shown in the protein sequence below the DNA sequences. Changes in amino acid sequence due to single nucleotide polymorphisms, underlined in the DNA sequence of some strains, are shown in parentheses.

elongation (45 s, 72 °C); and final elongation of 10 min at 72 °C. PCR products were analysed on 3% agarose gels.

Sequencing. The polyQ and adjacent coding regions of Tsr were sequenced either by Polymorphic DNA, Inc., by the Wayne State University sequencing facility or by Laragen. To prepare samples for Polymorphic DNA, Inc., fresh cultures were grown in T-broth overnight at 37 °C; DNA was extracted with the Wizard SV 96 Genomic DNA Purification System, 1 × 96 (Promega). Polymorphic DNA, Inc. analysed sequences of 20 strains from bases 795 to 1656 of the *tsr* gene. For the Wayne State University sequencing facility, DNA was amplified by PCR with primers *tsr*855F (5'-CACCGGCAATAACGATCTCT-3') and *tsr*1654R (5'-AAAATGTTTCCCAGTTCCTCT-3') and then sequenced in forward and reverse directions using the same primers. Similarly, Laragen bidirectionally sequenced PCR products amplified with primers *tsr*210F (5'-GGGTATCCGCTACATGATGG-3') and *tsr*840R (5'-ACCGCTATAGATGGCCATTGG-3').

Chemotaxis. Chemotaxis-dependent behaviour was measured in soft agar motility plates containing minimal medium (10 mM potassium phosphate, 1 mM MgSO₄, 1 mM (NH₄)₂SO₄, 0.25% agar), attractant (0.1 mM serine, 0.1 mM aspartate or no attractant for the control) and 0.009% glycerol as an additional carbon source (method according to G.L. Hazelbauer, personal communication). *E. coli* strains were initially grown overnight at 35 °C in Colilert-18, then inoculated into 5 ml T-broth medium and incubated for 5 h in a shaking water bath at 35 °C. After incubation, bacterial optical density at 600 nm was measured and adjusted to 0.3; duplicate soft agar plates were inoculated in the centre with 10 µl bacterial suspension and incubated at 35 °C for 13–15 h, after which the diameters of the chemotactic growth rings were measured. The minimum growth ring size was 5 mm (diameter of the inoculum), while the maximum growth ring size was 80 mm (diameter of the plate).

Statistics. ANOVA, descriptive, chi squared and other statistics were performed using Jandel Scientific Sigma Stat 2.1. Ninety-five per cent confidence intervals of proportions were estimated using a web-based tool at http://www.dimensionresearch.com/resources/calculators/conf_prop.html.

RESULTS

Previously sequenced *Escherichia* and *Shigella* genomes have 4Q and 7Q *tsr* variants

The 4Q region of *E. coli* K-12 MG1655 *tsr* is encoded by the sequence -CAGCAACAGCAG-. However, DNA sequences of several previously sequenced pathogenic *E. coli* strains (O157:H7 and the CFT073 urinary tract infection strain) have an additional 9 base repeat of -CAACAGCAG- and thus code for 7Q (Fig. 2). Although this might suggest that pathogenic strains of *E. coli* are more prone to polyQ expansion, comparison of other pathogenic strains of *E. coli* indicates that this is not necessarily the case. Other urinary tract infection (UTI) strains code for either 4Q (strain UTI89) or 7Q (strain 536). Moreover, *Shigella* spp., which are pathogenic subclades of *E. coli* (Pupo *et al.*, 2000), include both 4Q and 7Q strains (Fig. 2). *E. coli* HS, a non-pathogenic strain of *E. coli* originally isolated from humans (DuPont *et al.*, 1971; Levine *et al.*, 1978), has a Tsr with 10Q which has the addition of another -CAACAGCAG- tandem repeat (NCBI Microbial Genomes Annotation Project, 2005). The rest of the tether region of *tsr* is identical in length. Thus, previous genomic sequence data indicate that the polyQ region of *E. coli* *tsr* is variable in length.

7Q strains are most common, but >7Q also occurs

Primers that flank the tandem repeats were used to screen undomesticated strains of *E. coli* isolated from faecal samples of 50 different human subjects (representative strains shown in Fig. 3a) and from 75 stool samples of birds, farm animals, pets and other animals (representative strains shown in Fig. 3b). The animal strains were chosen from among strains having different alleles of the gene *uidA* (Ram *et al.*, 2004, 2007), and therefore known not to be clones of each other. Overall, the frequency of *tsr* variants in the human strains was 4Q, 5 (10%); 7Q, 43 (86%); 10Q, 1 (2%); and 13Q, 1 (2%). The frequency in the animal strains was 4Q, 13 (17%); 7Q, 53 (71%); 10Q, 4 (5%); and 13Q, 5 (7%). The frequencies of the polyQ variants did not differ significantly between animal and human strains (χ^2 , $P=0.2$). The 95% confidence intervals for samples of this size are approximately $\pm 10\%$, indicating that the frequency of strains that have 7Q is significantly larger than for other polyQ sizes.

In order to verify the sequences, the polyQ and adjacent region of *tsr* was sequenced in a subset of strains of each variant. The polyQ sequences of both 13Q (strains 149, 364, 404, 553, 564 and 565) and 10Q (strains 134, 351, 366 and HS) strains were derived by expansion involving the addition of another -CAACAGCAG- repeat, as illustrated by representative sequences in Fig. 2. The nucleotide sequences coding for 4Q ($n=2$) and 7Q ($n=17$) strains were the same in K-12 MG1655 and O157:H7 strains, respectively; in this region, however, single nucleotide polymorphisms occurred in some of the adjacent sequences of these strains (data not shown). In addition, at least one

Fig. 3. PCR analysis to detect various sizes of polyQ-coding regions in the tether section of *E. coli* *tsr*. Primers flanking the polyQ region gave different-sized products depending on the polyQ coding length. The lengths of the corresponding polyQ sequences, based on comparisons to the DNA ladders (first and last lanes; sizes in bp) and previously sequenced positive controls, are shown above each gel. (a) Strains isolated from human faecal samples (strains 1639–1841), shown with positive controls K-12 (originally derived from a human strain), 564 (isolated from a horse) and 378 (isolated from a seagull). (b) Strains isolated from animals. Strains 110, 131, 134, 149, 151 and 157 were isolated from dog faeces; strain 576 is from a cat; and strain 564 is from a horse.

7Q strain (strain 133, Fig. 2) and one 10Q strain (strain 399, not shown) had a histidinyl in place of one glutaminyl residue, due to a single nucleotide difference in the tandem ‘repeat’ sequence. Overall, however, we conclude that 7Q strains are the most common.

Expanded polyQ regions occur in two clusters on a phylogenetic tree

A genetic context in which larger polyQ regions are more likely to occur might be revealed by examining whether 10Q and 13Q strains cluster together on a phylogenetic tree. In order to examine a variable region not including the 10Q and 13Q sequences themselves and yet not be so distant that their relationship might be obscured by intervening recombination, we sequenced 570 bases of the less-conserved N-terminal end of *tsr*. In the resultant phylogenetic tree (Supplementary Fig. S2), all of the 13Q strains and half of the 10Q strains appear in two clusters. Cluster A, which has completely identical sequences in this relatively non-conserved region of the gene, nevertheless had all of the polyQ sizes (i.e. 4Q, 7Q, 10Q and 13Q), suggesting that polyQ variants in this cluster have arisen relatively recently. Cluster B strains have several single nucleotide differences in this part of the sequence and therefore form a group of closely related branches. 4Q and 7Q strains are broadly spread throughout the tree.

Expansion of polyQ to 13Q has no significant effect on chemotaxis

Chemotactic behaviour of representative strains having various polyQ sizes was tested in motility assays on soft agar plates (Fig. 4). The *E. coli* K-12 MG1655 strain showed typical soft agar plate behaviour, with growth rings in response to serine or aspartate in the presence of glycerol averaging about 35 mm larger than with glycerol alone. *E. coli* UU1624, which has Tar receptors but no Tsr receptors, similarly showed a larger growth ring with aspartate but approximately the same size ring for serine as with glycerol alone. Strains with 7Q ($n=7$), 10Q ($n=4$) and 13Q ($n=6$) all showed significant chemotaxis in response to serine and aspartate in the presence of glycerol compared to glycerol alone (one way repeated measures ANOVA, $P<0.05$, followed by Dunnett’s multiple comparison test to glycerol alone); however, there was no significant difference in the responses of the 10Q and 13Q strains from the responses of the 7Q strains (two way repeated measures ANOVA, $P=0.986$; as the most common variant, 7Q strains are considered the ‘control’). The chemotactic behaviour of the various groups of polyQ strains was similar to that of *E. coli* K-12 MG1655 and did not differ significantly between the groups. Thus, at least up to the size of 13Q, increasing the length of the tether sequence of Tsr has no significant effects on motility in response to serine in soft agar plates.

DISCUSSION

The demonstration of serine chemotaxis in natural variants of Tsr which have varying numbers of glutaminyl residues

Fig. 4. Comparison of chemotactic behaviour on soft agar plates of *E. coli* strains having various sizes of polyQ in Tsr. Bars represent diameters (mean $\pm$ SEM) of chemotactic rings after 14 h incubation at 35 °C with 0.009% glycerol (control), glycerol plus 0.1 mM aspartate, or glycerol plus 0.1 mM serine for seven 7Q strains, four 10Q strains and six 13Q strains. Also shown are the means (bars) and the individual values (●) of duplicate assays of an *E. coli* K-12 strain (a typical 4Q strain) and *E. coli* UU1624 (a K-12 derivative strain that lacks the Tsr receptor).

inserted into their tether regions provides supporting evidence that the tether region is a flexible peptide that can mediate chemotaxis over a range of sequence lengths. The presence of a core polyQ sequence of 4Q has provided a site at which genetic instability leading to polyQ expansion has occurred in at least two clusters of naturally occurring *E. coli* strains. These variations may be a worthwhile model for the study of genetic instability mechanisms underlying expansions of polyQ sequences such as those that occur in human genetic diseases.

The main objective in the soft agar plate experiments was to determine whether longer lengths of the Tsr tether in several natural *E. coli* strains disrupted serine-chemotaxis-dependent behaviour. Although the genetic background varied among these *E. coli* strains, it is clear that lengthening the tether region in Tsr caused no major disruption in serine chemotaxis and produced results similar to the K-12 laboratory strain. The accompanying control experiment with a Δ tsr strain demonstrated the efficacy of the assay in detecting decreases in the function of Tsr. If quantitative changes in responses on soft agar plates had occurred in these experiments, testing the variant genes in an isogenic background by transfer into a K-12 strain might have been useful to rule out differences due to the genetic diversity of the strains; this was unnecessary, since no differences were observed. An interesting comparison is with aspartate chemotaxis, which was tested in a set of strains having Tar constructs with short tethers in an isogenic K-12-related background (Li & Hazelbauer, 2006). Strains with short tethers exhibited smaller growth rings in soft agar plate assays. Accompanying capillary chemotaxis assay results correlated well with the soft agar plate results, corroborating the good relationship between soft agar plate and capillary chemotaxis methods in testing tether length variants.

Tether length would be expected to affect the size of MCP assistance neighbourhoods, a variable that has been investigated in a recent computational model of bacterial chemotaxis receptors by Endres & Wingreen (2006). The model shows that precise adaptation of Tsr responses in mixed clusters of Tsr and Tar receptors occurs with methylation/demethylation assistance neighbourhoods of six MCP dimers, approximately the same size that actually occurs in *E. coli* K-12 (Li & Hazelbauer, 2005). Models with shorter tethers (i.e. smaller assistance neighbourhoods) had decreased adaptation, in agreement with the decreased chemotaxis in experimental data (Li & Hazelbauer, 2006). Models with assistance neighbourhoods of nine MCPs, as might be expected with longer tethers, exhibited no difference in function from the six-MCP neighbourhood model. Thus, the normal serine chemotaxis in natural variants that have longer tethers (shown in this paper) is consistent with no change in adaptation in models of MCPs that have larger assistance neighbourhoods.

The new sequence data in this paper, as well as bioinformatic analysis of previous genome data, highlight the flexibility of the tether region in both length and amino acid composition. The tether region of Tsr in *E. coli* varies from 35 (e.g. in *E. coli* K-12) to 44 residues (in 13Q strains) (Fig. 2) and is 47 residues long in *P. luminescens* (Duchaud *et al.*, 2003). The sequence of the tether in *E. coli* K-12 MG1655 (-IQQQQRETSADVVKTVTPAAPRKMAVADSEE-, between the alpha-helical methylation region and the terminal -NWETF) is composed predominantly (21 out of 30 amino acids) of A, R, G, Q, S, P and E, which are known to promote protein disorder (Dunker *et al.*, 2001). Calculations of the predicted flexibility of Tsr sequences based on X-ray crystallographic B values (PROFbval, <http://cubic.bioc.columbia.edu/services/profbval/>) and other criteria (e.g. DRIP-PRED, spritz and PSIPRED disorder predictors, accessed via the DisProt homepage, <http://www.disprot.org/>) all predict the tether region to be more mobile than the adjacent methylation region (data not shown). Insertions of additional Q residues in the tether sequence extend the calculated length of the flexible region. Thus, both comparative analyses of Tsr proteins in various strains and species of bacteria, as well as the biochemical characteristics of the sequences, support the flexible nature of the tether region.

Flexibility of the tether enables CheR to methylate multiple glutamyl residues in nearby, but not identical, positions (Fig. 1) and to mediate binding of CheB (Barnakov *et al.*, 1999) for which the spatial requirements may be different from CheR. This flexibility also enables bound CheR and CheB to methylate and demethylate neighbouring MCPs (Li & Hazelbauer, 2005; Wu *et al.*, 1996). However, despite the functional, comparative, computational and biochemical evidence for flexibility of the tether region, there is no direct physical evidence for its flexibility. X-ray crystallographic data on the cytoplasmic region of Tsr (Kim *et al.*, 1999) have thus far been obtained only for amino acid residues 286 to 526 (thus including only a few residues from the tether) and the data showed, for residues 521 to 526 only, that they had large B values that prevented their being assigned identifiable structural positions. Ideally, future studies should consider physical methods of directly assessing the extent, movement and flexibility of the tether region in mediating its hypothesized tether function.

The occurrence of polyQ variation in *E. coli* Tsr is a notable exception to the general rarity of tandem repeats in *E. coli* in comparison to other species of bacteria. Compared to species such as *Haemophilus influenzae* and *Neisseria meningitidis*, relatively few tandem repeat sequences are found in *E. coli* K-12 (Bichara *et al.*, 2006). Indeed, Mrazek *et al.* (2007) have observed that large tandem repeats (such as the 9 base repeat described in the present paper) probably expand in some genomes but not in others due to (i) mutational bias promoting expansion, and (ii) lack of strong negative selection against the expansion. Since serine chemotaxis seems unaffected by expansion of the tandem repeats, and the expansions seem clustered in the

dendrogram of *E. coli* (Fig. S2), both conditions may be met in the clusters showing polyQ expansion in *E. coli*.

The bacterial strains described here may have use as a natural bacterial model for studying polyQ genetic instability, a mechanism found in several human genetic diseases, such as Huntington's disease. Synthetic polyQ-coding constructs that have been previously used in *E. coli* to analyse mechanisms mediating polyQ tandem repeats (e.g. Jaworski *et al.*, 1995) vary greatly in their levels of instability depending on whether they were constructed in plasmids (Kim *et al.*, 2006b) or inserted into the *E. coli* genome (Kim *et al.*, 2006a). In contrast, the present study describes a polyQ tandem repeat that is a natural phenomenon of a set of *E. coli* strains. The strains described in this paper may be useful in investigating natural mechanisms involved in regulating expansions of polyQ-coding sequences.

ACKNOWLEDGEMENTS

Wayne State University provided research funding. James Nataro (University of Maryland) contributed the *E. coli* HS strain. Harry L. T. Mobley (Department of Microbiology and Immunology, University of Michigan) provided the MG1655 strain. John S. Parkinson (Department of Biology, University of Utah) contributed the *E. coli* UU1624 strain and gave helpful suggestions on the manuscript. A. W. Black (Wayne State University) assisted with tether flexibility calculations described in the Discussion.

REFERENCES

- Adler, J. (1969). Chemoreceptors in bacteria. *Science* **166**, 1588–1597.
- Alexander, R. P. & Zhulin, I. B. (2007). Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. *Proc Natl Acad Sci U S A* **104**, 2885–2890.
- Barnakov, A. N., Barnakova, L. A. & Hazelbauer, G. L. (1999). Efficient adaptational demethylation of chemoreceptors requires the same enzyme-docking site as efficient methylation. *Proc Natl Acad Sci U S A* **96**, 10667–10672.
- Bichara, M., Wagner, J. & Lambert, I. B. (2006). Mechanisms of tandem repeat instability in bacteria. *Mutat Res* **598**, 144–163.
- Chalah, A. & Weis, R. M. (2005). Site-specific and synergistic stimulation of methylation on the bacterial chemotaxis receptor Tsr by serine and CheW. *BMC Microbiol* **5**, 12. doi:10.1186/1471-2180-5-12
- Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., Bocs, S., Boursaux-Eude, C., Chandler, M. & other authors (2003). The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat Biotechnol* **21**, 1307–1313.
- Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M. & other authors (2001). Intrinsically disordered protein. *J Mol Graph Model* **19**, 26–59.
- DuPont, H. L., Formal, S. B., Hornick, R. B., Snyder, M. J., Libonati, J. P., Sheahan, D. G., LaBrec, E. H. & Kalas, J. P. (1971). Pathogenesis of *Escherichia coli* diarrhea. *N Engl J Med* **285**, 1–9.
- Endres, R. G. & Wingreen, N. S. (2006). Precise adaptation in bacterial chemotaxis through “assistance neighborhoods”. *Proc Natl Acad Sci U S A* **103**, 13040–13044.
- Gosink, K. K., Burón-Barral, M. & Parkinson, J. S. (2006). Signaling interactions between the aerotaxis transducer Aer and heterologous chemoreceptors in *Escherichia coli*. *J Bacteriol* **188**, 3487–3493.
- Hazelbauer, G. L., Falke, J. J. & Parkinson, J. S. (2008). Bacterial chemoreceptors: high-performance signaling in networked arrays. *Trends Biochem Sci* **33**, 9–19.
- Jaworski, A., Rosche, W. A., Gellibolian, R., Kang, S., Shimizu, M., Bowater, R. P., Sinden, R. R. & Wells, R. D. (1995). Mismatch repair in *Escherichia coli* enhances instability of (CTG)_n triplet repeats from human hereditary diseases. *Proc Natl Acad Sci U S A* **92**, 11019–11023.
- Kim, K. K., Yokota, H. & Kim, S.-H. (1999). Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. *Nature* **400**, 787–792.
- Kim, S.-H., Pytlos, M. J., Rosche, W. A. & Sinden, R. R. (2006a). (CAG)/(CTG) repeats associated with neurodegenerative diseases are stable in the *Escherichia coli* chromosome. *J Biol Chem* **281**, 27950–27955.
- Kim, S. H., Pytlos, M. J. & Sinden, R. R. (2006b). Replication restart: a pathway for (CTG)(CAG) repeat deletion in *Escherichia coli*. *Mutat Res* **595**, 5–22.
- Lai, W.-C. & Hazelbauer, G. L. (2005). Carboxyl-terminal extensions beyond the conserved pentapeptide reduce rates of chemoreceptor adaptational modification. *J Bacteriol* **187**, 5115–5121.
- Le Moual, H. & Koshland, D. E., Jr (1996). Molecular evolution of the C-terminal cytoplasmic domain of a superfamily of bacterial receptors involved in taxis. *J Mol Biol* **261**, 568–585.
- Levine, M. M., Bergquist, E. J., Nalin, D. R., Waterman, D. H., Hornick, R. B., Young, C. R., Sotman, S. & Rowe, B. (1978). *Escherichia coli* strains that cause diarrhoea but do not produce heat-labile or heat-stable enterotoxins and are non-invasive. *Lancet* **1**, 1119–1122.
- Li, M. & Hazelbauer, G. L. (2005). Adaptational assistance in clusters of bacterial chemoreceptors. *Mol Microbiol* **56**, 1617–1626.
- Li, M. & Hazelbauer, G. L. (2006). The carboxyl-terminal linker is important for chemoreceptor function. *Mol Microbiol* **60**, 469–479.
- Maddock, J. R. & Shapiro, L. (1993). Polar location of the chemoreceptor complex in the *Escherichia coli* cell. *Science* **259**, 1717–1723.
- Mrazek, J., Guo, X. & Shah, A. (2007). Simple sequence repeats in prokaryotic genomes. *Proc Natl Acad Sci U S A* **104**, 8472–8477.
- NCBI Microbial Genomes Annotation Project (2005). COG0840: Methyl-accepting chemotaxis protein [*Escherichia coli* HS]. Genbank Accession: ZP_00708242.1 GI:75198172.
- Pupo, G. M., Lan, R. & Reeves, P. R. (2000). Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci U S A* **97**, 10567–10572.
- Ram, J. L., Ritchie, R. P., Fang, J., Gonzales, F. & Selegue, J. P. (2004). Sequence-based source tracking of *Escherichia coli* based on genetic diversity of beta-glucuronidase. *J Environ Qual* **33**, 1024–1033.
- Ram, J. L., Thompson, B., Turner, C., Nechvatal, J. M., Sheehan, H. & Bobrin, J. (2007). Identification of pets and raccoons as sources of bacterial contamination of urban storm sewers using a sequence-based bacterial source tracking method. *Water Res* **41**, 3605–3614.
- Rice, M. S. & Dahlquist, F. W. (1991). Sites of deamidation and methylation, a bacterial chemotaxis sensory transducer. *J Biol Chem* **266**, 9746–9753.
- Springer, M. S., Goy, M. F. & Adler, J. (1979). Protein methylation in behavioural control mechanisms and in signal transduction. *Nature* **280**, 279–284.

Studdert, C. A. & Parkinson, J. S. (2004). Crosslinking snapshots of bacterial chemoreceptor squads. *Proc Natl Acad Sci U S A* **101**, 2117–2122.

Studdert, C. A. & Parkinson, J. S. (2005). Insights into the organization and dynamics of bacterial chemoreceptor clusters through *in vivo* crosslinking studies. *Proc Natl Acad Sci U S A* **102**, 15623–15628.

Wu, J., Li, J., Li, G., Long, D. G. & Weis, R. M. (1996). The receptor binding site for the methyltransferase of bacterial chemotaxis is distinct from the sites of methylation. *Biochemistry* **35**, 4984–4993.

Edited by: M. F. Hynes