The cryptic ushA gene (ushA<sup>c</sup>) in natural isolates of Salmonella enterica (serotype Typhimurium) has been inactivated by a single missense mutation

David Innes, Ifor R. Beacham, Carie-Anne Beven, Meaghan Douglas, Michael W. Laird, John C. Joly and Dennis M. Burns

Two mutational mechanisms, both supported by experimental studies, have been proposed for the evolution of new or improved enzyme specificities in bacteria. One mechanism involves point mutation(s) in a gene conferring novel substrate specificity with partial or complete loss of the original (wild-type) activity of the encoded product. The second mechanism involves gene duplication followed by silencing (inactivation) of one of these duplicates. Some of these ‘silent genes’ may still be transcribed and translated but produce greatly reduced levels of functional protein; gene silencing, in this context, is distinct from the more common associations with bacterial partitioning sequences, and with genes which are no longer transcribed or translated. Whereas most Salmonella enterica strains are ushA<sup>M</sup>, encoding an active 5'-nucleotidase (UDP-sugar hydrolase), some natural isolates, including most genetically related strains of serotype Typhimurium, have an ushA<sup>c</sup> allele (designated ushA<sup>c</sup>) which produces a protein with, comparatively, very low 5'-nucleotidase activity. Previous sequence analysis of cloned ushA<sup>M</sup> and ushA<sup>c</sup> genes from serotype Typhimurium strain LT2 and Escherichia coli, respectively, did not reveal any changes which might account for the significantly different 5'-nucleotidase activities. The mechanism responsible for this reduced activity of UshA<sup>c</sup> has hitherto not been known. Sequence analysis of Salmonella ushA<sup>c</sup> and ushA<sup>M</sup> alleles indicated that the relative inactivity of UshA<sup>c</sup> may be due to one, or more, of four amino acid substitutions. One of these changes (S139Y) is in a sequence motif that is conserved in 5'-nucleotidases across a range of diverse prokaryotic and eukaryotic species. Site-directed mutagenesis confirmed that a Tyr substitution of Ser-139 in Salmonella UshA<sup>M</sup> was solely responsible for loss of 5'-nucleotidase activity. It is concluded that the corresponding single missense mutation is the cause of the UshA<sup>c</sup> phenotype. This is the first reported instance of gene inactivation in natural isolates of bacteria via a missense mutation. These results support a model of evolution of new enzymes involving a ‘silent gene’ which produces an inactive, or relatively inactive, product, and are also consistent with the evolution of a novel, but unknown, enzyme specificity by a single amino acid change.

Keywords: UDP-sugar hydrolase, 5'-nucleotidase, evolution, silent gene, cryptic gene
intermediate’). Further evolution of this ‘silent gene’, initially silenced by a mutation causing reduced transcription or a reduced level of functional product, which is maintained in the population for a finite period of time, then ensues via accumulation of a variety of additional mutations. Eventually such a gene may be lost by deletion but, before this occurs, it may instead acquire a functional, albeit modified, enzymic form that will persist if of selective value to the cell (Koch, 1972; Hartley, 1974, 1979, 1980, 1984; Rigby et al., 1974; Beacham, 1987). The term ‘silent gene’ has been previously used to include the possibility that, in one allele of the duplicates, lack of a functional gene product may also result from a missense mutation(s) causing, for example, aberrant folding of the encoded protein (Hartley, 1974; Rigby et al., 1974). The high level of gene paralogy in prokaryotic genomes indicates that gene duplication is a major mechanism for generating biochemical diversity (Labeledan & Riley, 1995; Riley & Labeledan, 1997).

Cryptic genes are also not usually expressed, i.e. are phenotypically silent at the level of transcription or translation, but do not have permanent loss of normal function. They can be periodically activated in a few members of a population by decryptification (reversion) under appropriate environmental circumstances to provide a selective advantage. Thus, cryptic genetic systems, involving cycles of inactivation and reactivation, represent a form of long-term gene regulation of rarely utilized functions operating at the level of bacterial populations (Hall et al., 1983; Hall, 1989).

Since the term ‘silencing’ has recently become more commonly associated with genes which are transcriptionally or translationally inactive or with partitioning sequences (Rine, 1999), use of the term ‘silent gene’ may imply lack of expression at the protein level. Because this is not necessarily always the case, we suggest instead that cryptic genes may, in principle, be of two types: those which encode a known function which is not normally expressed (as in the case of cryptic gene regulatory systems), and those which are expressed but whose function is unknown or whose protein product is simply inactive due to, for example, misfolding; the latter may be intermediates in the evolution of new genes (see above; Hartley, 1974; Rigby et al., 1974). Where lack of a protein product is known, the corresponding gene designation is superscript ‘c’, as in the case of known cryptic gene regulatory systems. Where the mechanism of gene inactivation is not known, or where the expressed protein has lost its original activity but may, or may not, have acquired novel substrate specificity, we suggest a gene designation of superscript ‘c’.

In the course of studies on the ush genes in Salmonella enterica and Escherichia coli, we have identified two inactive alleles, originally designated ushA\textsuperscript{c} and ushB\textsuperscript{c}, respectively (Burns et al., 1983; Burns & Beacham, 1986a, b; Edwards et al., 1993). E. coli K-12 and S. enterica serotype Typhimurium (synonym of Salmonella typhimurium) LT2 synthesize biochemically and genetically distinct UDP-sugar hydrolases, encoded by the ushA and ushB genes, respectively. The E. coli ushA gene encodes a periplasmic enzyme which is bi-functional, also possessing 5'-nucleotidase activity. This enzyme belongs to a superfamily of phosphoesterases and displays substantial sequence homology to 5'-nucleotidases from a variety of sources, many of which also have UDP-sugar hydrolase activity (Zimmermann, 1992; Koonin, 1994). In contrast, the ushB-encoded UDP-sugar hydrolase from Salmonella does not possess 5'-nucleotidase activity, is membrane-bound, and is unrelated immunologically and in primary sequence (Garrett et al., 1989; Jones et al., 1993).

It has previously been documented that S. enterica serotype Typhimurium strain LT2 does not possess a 5'-nucleotidase (Glaser et al., 1967; Neu, 1968; see also Burns & Beacham, 1986a). However, when the cloned ushA gene from E. coli was used to probe the genomes of strain LT2 and natural isolates of S. enterica serotype Typhimurium using Southern hybridization, a homologous sequence was identified (Burns & Beacham, 1986a). The LT2 homologue was cloned and characterized, and subsequently designated ushA\textsuperscript{c} (Burns & Beacham, 1986a); in this paper, we now refer to this allele as ushA\textsuperscript{c} since an inactive protein product is made (see Results). Further, most natural isolates of S. enterica serotype Typhimurium from a variety of sources, as well as strains of serotypes Gallinarum and Pullorum, have either no detectable, or very low, levels of 5'-nucleotidase activity (Neu, 1968; Burns & Beacham, 1986a; Edwards et al., 1993) but, like strain LT2, were shown by Southern analysis to contain homologues using an ushA\textsuperscript{c}-specific probe (Edwards et al., 1993; D. Innes, unpublished data). In contrast, all other natural isolates of S. enterica examined, representing all seven Salmonella subgroups, were found to contain active 5'-nucleotidases (Edwards et al., 1993).

The cloned ushA and ushA\textsuperscript{c} genes were shown to be located in homologous regions of their respective genomes, and to be 80% identical at the DNA sequence level and 87% identical at the amino acid level, the latter equating to 71 residue differences among the 550 residues. The ushA\textsuperscript{c} protein was shown to be immunologically related to UsHA and, like its counterpart in E. coli, to be located in the periplasm (Burns & Beacham, 1986a). Examination of the LT2 ushA\textsuperscript{c} sequence did not reveal any premature stop codons, internal deletions, nor any other detrimental alterations which might account for its greatly reduced 5'-nucleotidase activity in comparison with its E. coli homologue. This is in contrast to other well-known E. coli–Salmonella differences (e.g. lac, phoA) where the lack of functional Salmonella genes is due to the absence of the corresponding chromosomal segments (Riley & Sanderson, 1990). We have previously suggested that the ushA\textsuperscript{c} allele is not a component of a cryptic regulatory system, but rather has been very recently ‘silenced’, and may be destined for eventual deletion from the genome (Burns & Beacham, 1986a; Beacham, 1987; Edwards et al., 1993).
The mechanism of ushA\(^+\) inactivation, however, has hitherto not been known. In this work, we provide evidence that ushA\(^+\) alleles contain a missense mutation that results in a non-conserved amino acid substitution within a region which is generally conserved amongst similar enzymes from diverse prokaryotic and eukaryotic sources. We conclude that the corresponding single missense mutation is the cause of the UshA\(^+\) phenotype. This is the first reported instance of gene inactivation in natural isolates of bacteria via a missense mutation. This work supports a model of evolution of new enzymes involving a ‘silent gene’ which produces an inactive, or relatively inactive, product, and is also consistent with the evolution of a novel, but unknown, enzyme specificity by a single amino acid change.

**METHODS**

**Bacterial strains.** The *Salmonella* strains used in this study were taken from DNA subgroup I as this is the only group that has members with ushA\(^+\) alleles (Edwards *et al*., 1993). All strains used are serotypes of *S. enterica* and the designations for each strain have been abbreviated to their serotype names for simplicity. Where known, the electrophoretic type (ET) for each strain is also indicated (see Edwards *et al*., 1993; Reeves *et al*., 1989). The strains used were as follows: S. Typhimurium LT2, S. Agona CDC 3531-75 (ET9), S. Cholerae-suis CDC 290-71 (ET11), S. Gallinarum CDC 2950-79 (ET13), S. Hadar CDC 585-86 (ET8), S. Heidelberg CDC 4689-70 (ET6), S. Heidelberg 10540 \(\phi\) J R, S. Pullorum CDC 2385-70 (ET14), S. Typhimurium CDC 137-85 (ET5), S. Typhimurium CDC 331-86 (ET3), S. Typhimurium CDC 655-84 (ET3) and S. Typhimurium CDC 1119-83 (ET3). Sources of these organisms are as previously described (Edwards *et al*., 1993; Reeves *et al*., 1989). To create an *E. coli* chromosomal deletion of ushA, regions S\(^5\) and S\(^3\) were amplified using PCR and cloned into the R6K\(^+\)oriR plasmid, pS1080 (Bass *et al*., 1996), creating pM1L. This plasmid was then recombined into the chromosome of strain 46C5 [DE3(lac)K7 X74 \(\Delta\) phoA532 rep-71 phn[EcoB]/F\(^{pOX38}\)::Tn10-11] by M13 transduction and carbencillin selection (Steed & Wanner, 1993). P1 lysates were made on the resulting carbencillin-resistant chromosomal integrates and transduced into strain W3110 (Hill & Harnish, 1981). Carbencillin-resistant transductants were patched onto sucrose media (Bass *et al*., 1996) at 25 °C for selection of plasmid resolvents. Sucrose-resistant, carbencillin-sensitive colonies were screened for the deleted ushA\(^+\) gene by DNA amplification using oligonucleotides flanking the locus. One W3110 ushA\(^-\) isolate was stored as strain 51G1.

**Plasmids.** All plasmids used in this study were pBluescript SK\((+\)) (Stratagene) derivatives. pSM1 was generated by inserting the ushA\(^+\)-encoding 2.62 kb Msxl fragment from pDBS1 (Burns & Beacham, 1986a) into EcoRV-cut vector. pDBS1 was generated by inserting a 2.4 kb PCR-amplified fragment containing the ushA\(^+\) gene from *S. enterica* serotype Typhimurium CDC 137-85 into Kpn1/HindIII-cut vector. The amplification primer pair used corresponds to nucleotide positions 53–75 and 206–2048 (Burns & Beacham, 1986a), and incorporated add-on sequences to generate these unique restriction sites in the amplified product.

**PCR and DNA sequencing.** The ushA\(^+\) and ushA\(^-\) alleles were amplified from the *Salmonella* strains using oligonucleotides designed to anneal 90 bases upstream of the sequence encoding the initiating methionine codon and 207 nucleotides downstream of the sequence for the ushA\(^+\) stop codon (Burns & Beacham, 1986a). Amplification reactions (50 µl) consisted of 1.5 mmol MgCl\(_2\), 50 pmol of each primer, 12.5 µmol dNTP mix (25 µm each of dGTP, dATP, dTTP, dCTP) and 5 µl of a suspension of a single colony in 100 µl distilled water. Reactions were placed in a thermal cycler (preheated to 95 °C) for 7 min, prior to the addition of 2 units of Taq polymerase (Promega). The thermal cycling programme comprised 35 cycles of 95 °C for 30 s, 50 °C for 90 s and 72 °C for 60 s. PCR products were purified from 0.7% agarose gels using QiaEX II resin (Qiagen) and directly sequenced using the *Prism* Dye Terminator Cycle Sequencing reagents (Applied Biosystems). Sequencing primers were designed to anneal to the following nucleotide positions of the ushA\(^-\) coding strand (Burns & Beacham, 1986a): 319–333, 571–585, 835–849, 1091–1105, 1279–1299, 1484–1498 and 1691–1709. A similar set of primers was designed for determining the complementary sequence. These primers annealed to the following positions within the ushA\(^+\) sequence: 307–321, 546–560, 796–810, 1048–1062 and 1301–1315. Sequencing reactions were electrophoresed on an Applied Biosystems 373A automated sequencer.

**Immunoblotting.** Immunoblots were performed as described by Ausubel *et al*., (1987). Protein samples were transferred to PVDF membrane using a Bio-Rad Transblotting cell. The primary antibody was a polyclonal UshA antibody from rabbit serum (Beacham & Wilson, 1982). Bands were detected with a goat anti-rabbit–horse radish peroxidase conjugate and 3’,3’-diaminobenzidine/NiCl\(_2\) colour development solution.

**Site-directed mutagenesis.** The mutants described in this study were generated using the methods of Eckstein *et al*. (Taylor *et al*., 1985; Nakamaye & Eckstein, 1986) and Kunkel (1985), using reagents supplied by Amersham and Bio-Rad, respectively. The change of Phe-463 (nt positions 1532–1534) to a Leu residue was introduced using a 39-mer annealing to the region corresponding to nt positions 1535–1573 of ushA\(^-\), and the change of Ser-139 (nt positions 580–582) to a Tyr residue was introduced using a 21-mer annealing to the region corresponding to nt positions 571–591 of ushA\(^+\) (see Burns & Beacham, 1986a).

**Enzyme assay.** Assays of 5’-nucleotidase were performed as previously described using periplasmic extracts (Edwards *et al*., 1993). For each culture, an equal volume of periplasmic extract was added into three reaction mixes; each of these triplicate assays was sampled at four time points during the reaction providing a mean activity with less than 3% variance. A comparison of the measured activities at least three independently grown cultures of each isolate revealed a similarly low level of variance. Periplasmic extracts were prepared from 40 ml R-broth cultures grown aerobically at 37 °C to OD\(_{590}\) 0.8; R-broth contains (per litre) 10 g tryptone, 5 g NaCl, 1 g yeast extract, 1 g glucose. The cells were harvested by centrifugation and the cells resuspended in 1.8 ml of a solution containing 15% sucrose, 50 mM Tris/HCl, pH 8.0. Lysozyme (140 µl of a 2 mg ml\(^{-1}\) stock) and 40 µl 0.1M EDTA were added and the mixture was incubated at 25 °C for 10 min. Spheroplasts were removed by centrifugation at 8000 r.p.m. (Sorvall SS34 rotor) and the supernatant was carefully transferred to a fresh tube. The centrifugation step was repeated and supernatants (periplasmic fractions) were aliquoted into fresh tubes and assayed immediately, or stored at −80 °C until required. Specific activity was expressed as units (mg protein\(^{-1}\)), where a unit is defined as nmol P, min\(^{-1}\).

Assays of UDP-sugar hydrolase were also performed as previously described (Edwards *et al*., 1993) except that activities were measured in periplasmic extracts. Specific activity was expressed as units (mg protein\(^{-1}\)) where a unit is defined as (µmol UDP-glucose hydrolysed) min\(^{-1}\).
**Sequence analysis.** 5’-Nucleotidase sequences were identified following a blast search (Altschul et al., 1990) of GenBank, PDB, SWISS-PROT, PIR and PRF databases. Fully annotated sequences resulting from the search were extracted and edited, leaving only the mature portions of each primary sequence. These sequences were aligned using clustal x (Thompson et al., 1994) and analysed using Prosite (Bairoch, 1992) and Blocks (Henikoff & Henikoff, 1992) databases.

**RESULTS**

**Enzymic analysis of ushA and ushA+ strains of Salmonella**

We have previously shown that ushA alleles show mainly a clonal distribution amongst a set of isolates representing the seven DNA subgroups of *Salmonella*, whose genetic relationship was determined from multilocus enzyme electrophoresis (Reeves et al., 1989). Accordingly, for this study, a set of ushA and ushA+ strains representing DNA subgroup I of *Salmonella* (see Reeves et al., 1989; Edwards et al., 1993) was chosen for analysis. The specific activities of ushA-encoded periplasmic proteins could not be readily compared with those for chromosomally encoded 5’-nucleotidase in corresponding *Salmonella* isolates. The former was found to have a specific UDP-sugar hydrolase activity of 1-22 units mg⁻¹ whereas the latter had undetectable activity (<2.75 × 10⁻³ units mg⁻¹). Irrespective of which of the two enzymic activities are considered, the allocated phenotypes as UshA⁺ or UshA⁻ are consistent.

**Immunoblot analysis of ushA and ushA+ strains of Salmonella**

To determine whether the ushA+ allele is expressed at the translational level in natural *S. enterica* isolates, the presence of the protein in cell extracts was examined using immunoblotting. The results clearly indicated the presence of a cross-reactive polypeptide of molecular mass indistinguishable from native 5’-nucleotidase (Fig. 1). The presence of a second cross-reactive polypeptide of higher molecular mass (~85 kDa; see Fig. 1) has been previously observed (Burns & Beacham, 1986a) but its identity is not known. From this experiment we conclude that the ushA+ allele is transcribed and translated, producing a mature protein.

**Comparative nucleotide sequence analysis of ushA and ushA+ alleles**

To investigate the mechanism of inactivation of UshA⁻, we have sequenced the ushA+ and ushA⁻ alleles from all of the *Salmonella* strains listed in Table 1. The 5’-untranslated sequences, up to 62 nt from the initiation codon, are invariant amongst these strains (data not shown) and are the same as the ushA+ allele from strain LT2 (see Burns & Beacham, 1986a). They include the Shine–Dalgarno region, which is the same as in the ushA+ gene of *E. coli*, and the −10 and −35 regions, identified on the basis of transcription mapping of the ushA+ allele from *E. coli* (Burns & Beacham, 1986b). There is an identical single nucleotide change in their −10 regions when compared with *E. coli* (Burns & Beacham, 1986a). All of these observations are in accord with the fact that the ushA+ gene encodes a protein product.

We hence considered it likely that an amino acid change(s) in the UshA polypeptide results in a relatively inactive 5’-nucleotidase (see Table 1). Two approaches were used to identify those amino acid residues which merited further investigation.

First, the coding sequences of the *Salmonella* ushA+ and ushA⁻ genes were examined for nucleotide substitutions that occurred in at least three of the sequences at any
Missense mutation in the ushA<sup>c</sup> gene

**Fig. 1.** Immunoblot of UshA and UshA<sup>c</sup> proteins in cell extracts of *Salmonella* strains. Molecular mass markers are indicated to the left of the panel (kDa). Lanes: A, *E. coli*; B, *S. Heidelberg* CDC 4689-70; C, *S. Heidelberg* 10540 J90; D, *S. Agona* CDC 3531-75; E, *S. Hadar* CDC 585-86; F, *S. Cholerae-suis* CDC 290-71; G, *S. Typhimurium* CDC 137-85; H, *S. Typhimurium* CDC 179-87; I, *S. Typhimurium* LT2; J, *S. Typhimurium* CDC 1119-83; K, *S. Typhimurium* CDC 655-84; L, *S. Typhimurium* CDC 331-86; M, *S. Gallinarum* CDC 2950-79; N, *S. Pullorum* CDC 2385-70; O, purified Ush (200 ng).

**Table 2.** Amino acid substitutions in ushA<sup>c</sup> and ushA<sup>+</sup> strains of *Salmonella*

<table>
<thead>
<tr>
<th>Strain</th>
<th>ushA&lt;sup&gt;+&lt;/sup&gt;/ushA&lt;sup&gt;c&lt;/sup&gt;</th>
<th>ET&lt;sup&gt;*&lt;/sup&gt;</th>
<th>Amino acid codon:†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>139</td>
</tr>
<tr>
<td><em>S. Heidelberg</em> CDC 4689-70</td>
<td>ushA&lt;sup&gt;+&lt;/sup&gt;</td>
<td>6</td>
<td>Ser</td>
</tr>
<tr>
<td><em>S. Heidelberg</em> 10540 J90</td>
<td>ushA&lt;sup&gt;+&lt;/sup&gt;</td>
<td>–</td>
<td>Ser</td>
</tr>
<tr>
<td><em>S. Cholerae-suis</em> CDC 290-71</td>
<td>ushA&lt;sup&gt;+&lt;/sup&gt;</td>
<td>11</td>
<td>Ser</td>
</tr>
<tr>
<td><em>S. Agona</em> CDC 3531-75</td>
<td>ushA&lt;sup&gt;+&lt;/sup&gt;</td>
<td>9</td>
<td>Ser</td>
</tr>
<tr>
<td><em>S. Hadar</em> CDC 585-86</td>
<td>ushA&lt;sup&gt;+&lt;/sup&gt;</td>
<td>8</td>
<td>Ser</td>
</tr>
<tr>
<td><em>S. Typhimurium</em> CDC 137-85</td>
<td>ushA&lt;sup&gt;+&lt;/sup&gt;</td>
<td>5</td>
<td>Ser</td>
</tr>
<tr>
<td><em>S. Typhimurium</em> LT2</td>
<td>ushA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>–</td>
<td>Tyr</td>
</tr>
<tr>
<td><em>S. Typhimurium</em> CDC 1119-83</td>
<td>ushA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3</td>
<td>Tyr</td>
</tr>
<tr>
<td><em>S. Typhimurium</em> CDC 331-86</td>
<td>ushA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3</td>
<td>Tyr</td>
</tr>
<tr>
<td><em>S. Typhimurium</em> CDC 655-84</td>
<td>ushA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3</td>
<td>Tyr</td>
</tr>
<tr>
<td><em>S. Gallinarum</em> CDC 2950-79</td>
<td>ushA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>13</td>
<td>Ser</td>
</tr>
<tr>
<td><em>S. Pullorum</em> CDC 2385-70</td>
<td>ushA&lt;sup&gt;c&lt;/sup&gt;</td>
<td>14</td>
<td>Ser</td>
</tr>
</tbody>
</table>

* ET, electrophoretic type (see Edwards et al., 1993; Reeves et al., 1989). –, Unknown.
† The nucleotide positions (see Burns & Beacham, 1986a) corresponding to these codons are: 580–582 (codon 139), 1204–1206 (codon 347), 1510–1512 (codon 449), 1552–1554 (codon 463) and 1795–1797 (codon 544).

given nucleotide position, which would potentially reveal amino acid substitutions for which a possible correlation to phenotype could be established. Through such an analysis, we discovered 13 nucleotide differences, of which 8 occur at the 3' base of codons (data not shown) and therefore do not result in amino acid substitutions. The remaining five nucleotide differences were examined to see whether they result in conservative or non-conservative amino acid substitutions, and for correlation to phenotype. These nucleotide changes have led to four amino acid substitutions, of which two are conservative (at positions 463 and 544; Table 2) and two non-conservative (at positions 139 and 449; Table 2). We conclude that one or more of these alterations are responsible for the loss of 5'-nucleotidase activity of the ushA<sup>c</sup> gene product.

The second approach involved ordered pairwise comparisons of the UshA and UshA<sup>c</sup> amino acid sequences to identify all residue differences. Depending on the two sequences examined, the numbers of substitutions ranged from zero to seven residues (data not shown). All amino acid changes were then analysed for possible correlation to phenotype. In addition to the four residues already identified using the first approach, another candidate residue (at position 347; Table 2) was identified. However, this residue is of possible interest only in relation to the UshA<sup>c</sup> phenotype of serotypes Gallinarum and Pullorum.

**Comparative amino acid sequence analysis of 5’-nucleotidases from diverse sources**

Notwithstanding the availability of the three-dimensional structure of *E. coli* 5’-nucleotidase (Knöfel & Sträter, 1999), we sought to gain insight into which of these residues might be important for the catalytic function of UshA. This was done by aligning amino acid sequences of 5’-nucleotidases from diverse sources.
Fig. 2. Alignment of mature 5'-nucleotidase sequences. The one-letter code for amino acids is used. The sequence of the S. enterica serotype Typhimurium strain LT2 5'-nucleotidase is shown on the top line. The other sequences are of active 5'-nucleotidases from various sources: E.coli, E. coli K-12 (Burns & Beacham, 1986b); Vibrio, Vibrio parahaemolyticus (Tamao et al., 1991); Human, Homo sapiens (Murray et al., 1990a); Bovine, Bos taurus (Suzuki et al., 1993); Rattus, Rattus norvegicus (Mishima et al., 1990b); Mus, Mus musculus (Resta et al., 1993); Discopyge, Discopyge ommata (Volkan et al., 1991); Treponema, Treponema pallidum (Volkan et al., 1998); Boophilus, Boophilus microplus (Liyou et al., 1994). Numbers indicate the position of each residue at the end of the FPN motifs shown, with respect to the sequences of corresponding precursor proteins. The 'y' residue on the top line is Tyr-139 (see text). Boxed residues show sequence identity in at least 7 of the 10 sequences shown; upper-case unboxed residues show strong sequence similarity while the lower-case italicized residues show weaker similarity, as determined by weights assigned in the BLOSUM62 matrix (Henikoff & Henikoff, 1992) and the CLUSTAL program (Thompson et al., 1994).

This change by creating a C→A substitution at the second position of codon 139 in the active ushA<sup>c</sup> gene from S. enterica Typhimurium strain CDC 137-85; pSM1 contains ushA<sup>c</sup> from S. Typhimurium LT22; p1387C-T and pCA-D are site-directed mutants derived from p2.4, as indicated.

<table>
<thead>
<tr>
<th>Plasmid*</th>
<th>Amino acid codon:</th>
<th>Specific activity†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>139</td>
</tr>
<tr>
<td>p2.4</td>
<td>Ser</td>
<td>893</td>
</tr>
<tr>
<td>p1387C-T</td>
<td>Ser</td>
<td>1.7 × 10&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>pSM1</td>
<td>Tyr</td>
<td>87</td>
</tr>
<tr>
<td>pCA-D</td>
<td>Tyr</td>
<td>111</td>
</tr>
</tbody>
</table>

* All strains are 51G1 (ΔushA<sup>c</sup>) containing the indicated plasmid, or no plasmid. Plasmid p2.4 contains the ushA<sup>c</sup> gene from S. Typhimurium CDC 137-85; pSM1 contains ushA<sup>c</sup> from S. Typhimurium LT22; p1387C-T and pCA-D are site-directed mutants derived from p2.4, as indicated.

† Specific activity is expressed as nmol P<sub>1</sub> min<sup>-1</sup> mg<sup>-1</sup>.

**Table 3. 5'-Nucleotidase activity in site-directed mutants of ushA<sup>c</sup>**
Missense mutation in the \textit{ushA}® gene

(FPN) motif. This amino acid substitution was due to a single C→A nucleotide change. In the three-dimensional structure of \textit{E. coli} UshA (Knöfel & Sträte, 1999), this residue is located between β-sheets 4 and 5 and does not appear to be within the active site. Computer simulation of this substitution predicts no significant effect on tertiary structure (data not shown). Nevertheless, the location of this residue within the conserved FPN motif suggests an important role in structure and/or function.

Fig. 3 summarizes the amino acid substitution data presented in Table 2 in the context of depicting how the postulated mutational change (see below) has led to the inactivation of the \textit{ushA}® gene in natural isolates of \textit{S. enterica} serotype Typhimurium. The independent mutation(s) responsible for the inactivation of the \textit{ushA}® alleles in serotypes Gallinarum and Pullorum are not included in this figure but are discussed below.

This study is the first report of gene inactivation by missense mutation occurring in natural isolates of bacteria, leading to an expressed yet relatively inactive (5'-nucleotidase) enzyme. This is quite distinct from gene inactivation (or deletion) as a consequence of laboratory cultivation in rich medium, and confirms the importance of examining natural isolates (Delorme et al., 1993; Burns et al., 1995).

\textit{ushA}® in other \textit{Salmonella} serotypes

All other \textit{Salmonella} isolates examined are UshA+ except for two additional strains, one Gallinarum and one Pullorum (ET13 and ET14, respectively), which were found to be phenotypically UshA® (Edwards et al., 1993; Table 1). As in the case of CDC 137-85 (serotype Typhimurium, ET3, UshA+), these strains are also relatively distantly related to those of ET3 and do not share the inactivating mutation in codon 139. Hence, their UshA® phenotype is consistent with independent mutational event(s) and with the clonal origin of the \textit{ushA}® Typhimurium strains. The only consistent amino acid residue differences between all UshA+ proteins examined and these two UshA® sequences were identified as S347L, A449R and F463L (Table 2). Of these, only A449R is a non-conservative substitution. However, the identity of the inactivating mutations responsible for the UshA® phenotype in the Gallinarum and Pullorum strains remain to be determined.

The origin of new genes

Aside from past or present combinatorial mechanisms, there are two mutational mechanisms whereby, in principle, new or improved enzyme specificities may arise. First, there is a finite probability that ‘beneficial’ mutations may be acquired, resulting in one or more amino acid substitutions which do not significantly alter expression but change the properties of the resulting polypeptide. Such changes will be ‘fixed’ if of selective value; this may occur with partial or complete loss of the original enzymic activity. There is substantial evidence for this mechanism from studies of experimental evolution (Clarke, 1974; Hall, 1981; Hartley, 1984). In the context of \textit{ushA}®, it is possible that a new, unknown, enzymic specificity has arisen concomitantly with loss of 5'-nucleotidase (UDP-sugar hydrolase) activity.

Second, the analysis of bacterial genomes clearly indicates that the evolution of biochemical diversity has involved substantial levels of gene duplication; at least 37% of \textit{E. coli} genes, for example, have paralogous counterparts in the genome (Labedan & Riley, 1995; Riley & Labedan, 1997). It has previously been proposed that enzyme evolution is more efficient if one duplicate is rendered non-functional at the level of transcription or translation, later to revert to an active allele (Koch, 1972). Similarly, it has been suggested that the evolution of an enzyme family may involve ‘mutation of one copy to a ‘silent gene’ that produces an inactive product that cannot fold correctly’ (Rigby et al., 1974); following genetic divergence, reversion, at the original or at another site, may restore the ability of the protein to fold correctly (Hartley, 1974; Rigby et al., 1974). If they do not revert to a new form, such ‘silent genes’ will inevitably be eventually deleted from the genome; their
presence in extant genomes is therefore likely to be rare, as recent analyses of sequenced bacterial genomes confirms (Fraser et al., 2000), and in contrast to higher eukaryotic genomes, in which the presence of pseudogenes is common (Mighell et al., 2000). Whether or not inactive alleles are retained in the genomes of prokaryotes is of significance with respect to this model for the evolution of new or improved enzyme specificities in bacteria. Our observations with the ushA/ushA′ system are consistent with the hypothesized ‘silent gene intermediate’, although the ushA′ gene does not appear to be a parologue of an active ushA gene. However, the unrelated ushB gene encoding an active UDP-sugar hydrolase, and present in all Salmonella strains examined (see Edwards et al., 1993), is a functional duplicate of ushA and provides a possible rationale for the inactivation of ushA′. If the resident ‘ushA′ gene in Salmonella became functionally a duplicate of ushB, which had been acquired more recently, then the former would be expected to have become inactivated (ushA′) if the acquired gene (ushB) has a similar role and a greater selective value. UshB is a more active UDP-sugar hydrolase, and present in all salmonellae. Our observations with the inactivation of ushA′, and present in all salmonellae, are consistent with the hypothesized ‘silent gene intermediate’, although the ushA′ gene does not appear to be a parologue of an active ushA gene. However, the unrelated ushB gene encoding an active UDP-sugar hydrolase, and present in all Salmonella strains examined (see Edwards et al., 1993), is a functional duplicate of ushA and provides a possible rationale for the inactivation of ushA′. If the resident ‘ushA′ gene in Salmonella became functionally a duplicate of ushB, which had been acquired more recently, then the former would be expected to have become inactivated (ushA′) if the acquired gene (ushB) has a similar role and a greater selective value. UshB is a more active UDP-sugar hydrolase than the Salmonella UshA′ enzyme (see Results) and the E. coli UshA enzyme (based on cellular specific activities and measured K_m; Glaser et al., 1967). UshB also has broader specificity for nucleoside-diphosphate sugars than the E. coli UshA enzyme (Glaser et al., 1967). The fact that ushA′ is not a tandem duplication of ushB may account for the persistence of ushA′ in the Salmonella genome since homologous recombination between them would not occur.

Despite the lack of evidence for gene duplication preceding inactivation of ushA′, we suggest that the existence of ushA′ also supports the notion of ‘silent gene intermediates’ in the evolution of new enzyme activities.

More ‘cryptic’ genes in sequenced genomes?

There are ORFs present in sequenced genomes to which a function has been assigned, on the basis of sequence similarity to known genes, but for which the corresponding enzyme activity is not apparent. These ORFs have been suggested to be possibly ‘remnant’ genes, which have not yet been eliminated from the genome (Pollack, 1997; Cordwell, 1999). If they encode a protein product, their original enzymic activity may have been lost by virtue of missense mutation(s), as in the case of ushA′. Alternatively, these genes may code for proteins which have unknown functions, despite the sequence similarity of their putative encoded proteins to those of known function (Pollack, 1997; Cordwell, 1999).

Is ushA′ a ‘remnant’ gene?

Cryptic genes have been previously defined as those which are postulated to have a role in long-term gene regulation. They are normally not expressed but undergo cycles of inactivation and activation involving periodic selection, which may allow expression of the gene (i.e. reversion via single mutational events) under selective conditions (Hall et al., 1983; Hall, 1989). The question arises as to whether the ushA′ gene should be interpreted in this way. Clearly, the single C→A substitution at nt position 581, responsible for ushA′ inactivation (loss of 5′-nucleotidase activity) in Typhimurium isolates, may be presumed to be potentially reversible, although this has not yet been observed (Burns & Beacham, 1986a). In cryptic gene regulatory systems, however, the great majority of natural isolates contain the cryptic (non-expressed) allele (Hall & Betts, 1987; Burns et al., 1995). In contrast, the great majority of S. enterica isolates are ushA′ (Edwards et al., 1993). Those relatively few isolates which are ushA′ can simply be explained by a recent C→A change which occurred in an ancestor of most strains of serotype Typhimurium (see Edwards et al., 1993). In summary, there is no compelling evidence that ushA′ is part of a cryptic gene regulatory system.

Instead, ushA′ may exemplify a second category of cryptic gene, expressed at the protein level. The E. coli serotype O157:H7 uidA gene may be another such example (Feng & Lampel, 1994). Such genes may have evolved new (unknown) specificities.

The possible acquisition of altered enzyme specificity for the ushA′ gene product is difficult to exclude experimentally in the absence of any notion of which substrates might be tested. Alternatively, such a cryptic gene may represent a ‘remnant’ gene which has become redundant. This, in turn, may be due to the presence of a more active UDP-sugar hydrolase encoded by ushB (see above). It is also possible that loss of its 5′-nucleotidase activity may have adaptive value in that it may possibly be related to the facultative intracellular existence of Salmonella. A role for the enzyme in affecting host cell function has been suggested in the case of invasive strains of E. coli due to the inhibition of host protein kinases (Berger et al., 1996). Although over-expression of 5′-nucleotidase increases intracellular survival in vitro, it is possible that the alteration of intracellular function is maladaptive. In either case, the ushA′ gene may have accumulated neutral or deleterious mutations and may eventually be deleted from the genome as a non-essential or maladaptive gene.

ACKNOWLEDGEMENTS

D. Innes gratefully acknowledges the receipt of an Australian Postgraduate Research Award.

REFERENCES


Neu, H. C. (1968). The S'-nucleotidases and cyclic phospho-...


Received 29 December 2000; revised 6 March 2001; accepted 22 March 2001.