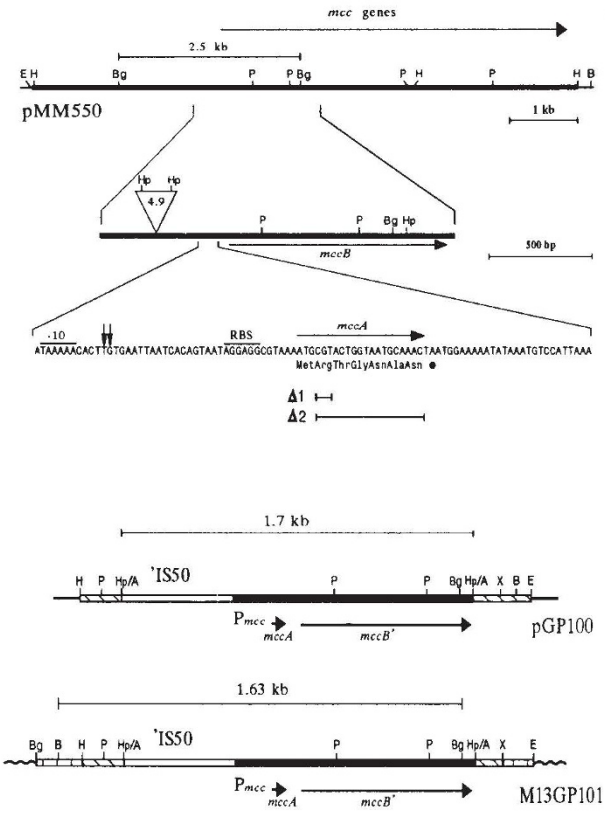
The smallest known gene

SIR — Although commonly found in nucleotide sequences, very short open reading frames (ORFs) are usually thought to have no biological significance. But our results, described below, show that such ORFs deserve more careful attention. Microcin C7 (MccC7), a modified linear heptapeptide, inhibits protein synthesis in Enterobacteriaceae[1]. Chemical sequencing indicated that MccC7 consists of Acetyl–Met–Arg–Thr–Gly–Asn–Ala–Asp–X, where X symbolizes an unidentified, acid-labile group substituting for the C-terminal aspartate residue[1]. NMR and mass spectrometry studies indicate that the M_r of MccC7 is 1,177 ± 1 daltons, the Met residue is not acetylated but formylated, and X includes a nucleotide (I. Guijarro *et al.*, manuscript in preparation). We show here that the heptapeptide forming part of MccC7 is synthesized from a 21 base-pair (bp) gene called *mccA*.

MccC7 is produced by resting *Escherichia coli* cells harbouring the 43-kilobase (kb) single-copy plasmid pMccC7(ref. 2). Like other microcins[3,4], the biosynthesis, maturation and secretion of MccC7 requires the activity of several clustered genes. We have entirely sequenced on both strands the 5.2-kb DNA region where these genes lie. It contains four ORFs that overlap the genes *mccB–E* previously identified by functional tests[2] (and our manuscript in preparation). Another ORF, located 77 bp upstream of *mccB*, is 21 bp long and its deduced amino-acid sequence matches that of the peptide present in mature microcin, except for the C-terminal residue which is aspartate by chemical determination but asparagine according to the nucleotide sequence. Given this almost perfect match, we assume that this 21-nucleotide ORF (*mccA*) encodes the microcin heptapeptide (see figure).



Plasmid pMM550 is a pBR322 derivative that contains the five genes, *mccA-E*, required to produce active extracellular MccC7. Inverted triangle, Tn5 insertion 4.9 (ref.2), which does not affect microcin production. Vertical arrows, start of *mccA-mccB* transcripts from Pmcc promoter. The −10 box of P_{mcc}, the RBS of *mccA* and the extent of deletions Δ1 and Δ2 are indicated. Filled boxes, pMccC7 DNA; open boxes, IS50 DNA; diagonally hatched boxes, pUC18 polylinker; single line, pUC18 DNA; vertically hatched boxes, M13tg130 polylinker; wavy line, M13tg130 DNA. Restriction sites: A, *Acc*I; B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; Hp, *Hpa*II; P, *Pst*I; X, *Xba*I. Methods are available from F. M. by request.

To demonstrate that *mccA* encodes the peptide chain of MccC7, we generated two deletions within the gene and examined their effect on MccC7 production. Both deletions were constructed on the recombinant phage M13GP101 by the uracil-substitution mutagenesis procedure of Kunkel[5]. This phage contains *mccA*, a large part of *mccB* and the Pmcc promoter driving transcription of both genes. One of the deletions (△2) spanned the six last codons of *mcc*A, while the other (△1) eliminated only the second one (CGT). In both cases the reading frame of MccA was maintained. When either of these deleted *mccA* genes was replaced for the wild-type *mccA* gene in pMM550, the ability of this plasmid to direct microcin production was completely lost. The production was recovered by introducing pDG250, a single-copy plasmid compatible with pMM550, containing Pmcc, wild-type *mccA* and the 5′ half of *mccB*, but lacking other *mcc* genes[6], to the same cells.

We believe that the microcin oligopeptide does not derive from a larger translation product because: (1) There is a canonical ribosome binding site (RBS), AGGAGG, six nucleotides 5′ of the first codon of *mccA*; (2) the amino methionine residue of microcin is formylated — it is well known that internal methionines of polypeptides are never formylated and that protein translation in *E. coli* begins by *N*-formyl-methionine, which usually is later deformylated; and (3) the reading from the unique alternative putative translation initiation codon in frame with *mccA*, and ATG located at position −84, is interrupted by six stop codons before reaching *mccA*. Thus the *mccA* gene encodes the peptidic chain of MccC7. To our knowledge, *mccA* is the smallest gene so far reported.

Ribosomally encoded peptide antibiotics are usually synthesized as a prepropeptide consisting of two parts: a propeptide which after suffering postranslational modification(s) yields the mature antibiotic, and an N-terminal leader sequence thought to anchor the propeptide so it can be modified. Antibiotic maturation is achieved by cleavage of the leader peptide before or during secretion from cells[3,7]. In contrast with these antibiotics, our results show that the MccC7 heptapeptide does not derive from a larger translation product, it being directly synthesized as a propeptide.

As well as *mccA*, a handful of small genes have been identified. We know of *iad*, the 22-codon gene from *Enterococcus faecalis* encoding the precursor of a competitive inhibitor of the sex pheromone cAD1[8], and of two of 26 codons, one encoding the cytolytic toxin δ-lysin from *Staphylococcus aureus*[9] and the other the morphogenetic product SPOVM from *Bacillus subtilis*[10]. The existence of these small genes means that small ORFs should not be ignored when they are preceded by a consensus RBS, or important biological functions might be overlooked.

José E. González-Pastor
José L. San Millán
Felipe Moreno*
Unidad de Genética Molecular,
Hospital Ramón y Cajal,
Carretera de Colmenar km 9.1.,
Madrid 28034, Spain

*Corresponding author.

1. García-Bustos, J. F., Pezzi, N. & Méndez, E. *Antimicrob. Agents Chemother.* **27**, 791–797 (1985).
2. Novoa, M. A., Díaz-Guerra, L., San Millán, J. L. & Moreno, F. *J. Bact.* **168**, 1384–1391 (1986).
3. Kolter, R. & Moreno, F. *A. Rev. Microbiol.* **46**, 141–163 (1992).
4. Moreno, F., San Millán, J. L., Hernánez–Chico, C. & Kolter, R. in *Genetics and Biochemistry of Antibiotic Production* (eds Vining, L.C. & Stuttard, C.) 71–85 (Butterworth–Heinemann, Boston, 1993).
5. Kunkel, T.A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488 (1985).
6. Díaz-Guerra, L., Moreno, F. & San Millán, J. L. *J. Bact.* **171**, 2906–2908 (1989).
7. Jung, G. *Angew. Chem. Int. Ed.* **30**, 1051–1068 (1991).
8. Clewell, D.B., Pontius, L. T., An, F. Y., Ike, Y., Suzuki, A. & Nakayama, J. *Plasmid* **24**, 156–161 (1990).
9. Janzon, L. & Arvidson, S. *EMBO J.* **9**, 1391–1399 (1990).
10. Levin, P.A., Fan, N., Ricca, E., Driks, A., Losick, R. & Cutting, S. *Molec. Microbiol.* **9**, 761–771 (1993).