IDENTIFICATION OF GENES

The present invention relates to methods for the identification of genes involved in the adaptation of a microorganism to its environment, particularly the identification of genes responsible for the virulence of a pathogenic microorganism.

Background to the invention

Antibiotic resistance in bacterial and other pathogens is becoming increasingly important. It is therefore important to find new therapeutic approaches to attack pathogenic microorganisms.

Pathogenic microorganisms have to evade the host's defence mechanisms and be able to grow in a poor nutritional environment to establish an infection. To do so a number of "virulence" genes of the microorganism are required.

Virulence genes have been detected using classical genetics and a variety of approaches have been used to exploit transposon mutagenesis for the identification of bacterial virulence genes. For example, mutants have been screened for defined physiological defects, such as the loss of iron regulated proteins (Holland *et al*, 1992), or in assays to study the penetration of epithelial cells (Finlay *et al*, 1988) and survival within macrophages (Fields *et al*, 1989; Miller *et al*, 1989a; Groisman *et al*, 1989). Transposon mutants have also been tested for altered virulence in live animal models of infection (Miller *et al*, 1989b). This approach has the advantage that genes can be identified which are important during different stages of infection, but is severely limited by the need to test a wide range of mutants individually for alterations to virulence. Miller *et*

al (1989b) used groups of 8 to 10 mice and infected orally 95 separate groups with a different mutant thereby using between 760 and 950 mice. Because of the extremely large numbers of animals required, comprehensive screening of a bacterial genome for virulence genes has not been feasible.

Recently a genetic system (*in vivo* expression technology [IVET]) was described which positively selects for *Salmonella* genes that are specifically induced during infection (Mahan *et al*, 1993). The technique will identify genes that are expressed at a particular stage in the infection process. However, it will not identify virulence genes that are regulated posttranscriptionally, and more importantly, will not provide information on whether the gene(s) which have been identified are actually required for, or contribute to, the infection process.

Lee & Falkow (1994) *Methods Enzymol.* **236**, 531-545 describe a method of identifying factors influencing the invasion of *Salmonella* into mammalian cells *in vitro* by isolating hyperinvasive mutants.

Walsh and Cepko (1992) *Science* **255**, 434-440 describe a method of tracking the spatial location of cerebral cortical progenitor cells during the development of the cerebral cortex in the rat. The Walsh and Cepko method uses a tag that contains a unique nucleic acid sequence and the lacZ gene but there is no indication that useful mutants or genes could be detected by their method.

WO 94/26933 and Smith *et al* (1995) *Proc. Natl. Acad. Sci. USA* **92**, 6479-6483 describe methods aimed at the identification of the functional regions of a known gene, or at least of a DNA molecule for which some sequence information is available.

Groisman *et al* (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1033-1037 describes the molecular, functional and evolutionary analysis of sequences specific to *Salmonella*.

Some virulence genes are already known for pathogenic microorganisms such as *Escherichia coli*, *Salmonella typhimurium*, *Salmonella typhi*, *Vibrio cholerae*, *Clostridium botulinum*, *Yersinia pestis*, *Shigella flexneri* and *Listeria monocytogenes* but in all cases only a relatively small number of the total have been identified.

The disease which *Salmonella typhimurium* causes in mice provides a good experimental model of typhoid fever (Carter & Collins, 1974). Approximately forty two genes affecting *Salmonella* virulence have been identified to date (Groisman & Ochman, 1994). These represent approximately one third of the total number of predicted virulence genes (Groisman and Saier, 1990).

The object of the present invention is to identify genes involved in the adaptation of a microorganism to its environment, particularly to identify further virulence genes in pathogenic microorganisms, with increased efficiency. A further object is to reduce the number of experimental animals used in identifying virulence genes. Still further objects of the invention provide vaccines, and methods for screening for drugs which reduce virulence.

Summary of the invention

A first aspect of the invention provides a method for identifying a microorganism having a reduced adaptation to a particular environment comprising the steps of:

This Sheet Has Been Replaced

(1) providing a plurality of microorganisms each of which is independently mutated by the insertional inactivation of a gene with a nucleic acid comprising a unique marker sequence so that each mutant contains a different marker sequence, or clones of the said microorganism;

(2) providing individually a stored sample of each mutant produced by step (1) and providing individually stored nucleic acid comprising the unique marker sequence from each individual mutant;

(3) introducing a plurality of mutants produced by step (1) into the said particular environment and allowing those microorganisms which are able to do so to grow in the said environment;

(4) retrieving microorganisms from the said environment or a selected part thereof and isolating the nucleic acid from the retrieved microorganisms;

(5) comparing any marker sequences in the nucleic acid isolated in step (4) to the unique marker sequence of each individual mutant stored as in step (2); and

(6) selecting an individual mutant which does not contain any of the marker sequences as isolated in step (4).

Thus, the method uses negative selection to identify microorganisms with reduced capacity to proliferate in the environment.

A microorganism can live in a number of different environments and it is known that particular genes and their products allow the microorganism to adapt to a particular environment. For example, in order for a pathogenic microorganism, such as a pathogenic bacterium or pathogenic fungus, to survive in its host the product of one or more virulence genes is required. Thus, in a preferred embodiment of the invention a gene of a microorganism which allows the microorganism to adapt to a particular environment is a virulence gene.

Conveniently, the particular environment is a differentiated multicellular organism such as a plant or animal. Many bacteria and fungi are known to infect plants and they are able to survive within the plant and cause disease because of the presence of and expression from virulence genes. Suitable microorganisms when the particular environment is a plant include the bacteria *Agrobacterium tumefaciens* which forms tumours (galls) particularly in grape; *Erwinia amylovara*; *Pseudomonas solanacearum* which causes wilt in a wide range of plants; *Rhizobium leguminosarum* which causes disease in beans; *Xanthomonas campestris* p.v. *citri* which causes canker in citrus fruits; and include the fungi *Magnaporthe grisea* which causes rice blast disease; *Fusarium* spp. which cause a variety of plant diseases; *Erisyphe* spp.; *Colletotrichum gloeosporiodes*; *Gaeumannomyces graminis* which causes root and crown diseases in cereals and grasses; *Glomus* spp., *Laccaria* spp.; *Leptosphaeria maculans*; *Phoma tracheiphila*; *Phytophthora* spp., *Pyrenophora teres*; *Verticillium alboatrum* and *V. dahliae*; and *Mycosphaerella musicola* and *M. fijiensis*. As described in more detail below, when the microorganism is a fungus a haploid phase to its life cycle is required.

Similarly, many microorganisms including bacteria, fungi, protozoa and trypanosomes are known to infect animals, particularly mammals including humans. Survival of the microorganism within the animal and the ability of the microorganism to cause disease is due in large part to the presence of and expression from virulence genes. Suitable bacteria include *Bordetella* spp. particularly *B. pertussis*, *Campylobacter* spp. particularly *C. jejuni*, *Clostridium* spp. particularly *C. botulinum*, *Enterococcus* spp. particularly *E. faecalis*, *Escherichia* spp. particularly *E. coli*, *Haemophilus* spp. particularly *H. ducreyi* and *H. influenzae*, *Helicobacter* spp. particularly *H. pylori*, *Klebsiella* spp. particularly *K. pneumoniae*, *Legionella* spp. particularly *L. pneumophila*, *Listeria* spp. particularly *L.*

monocytogenes, *Mycobacterium* spp. particularly *M. smegmatis* and *M. tuberculosis*, *Neisseria* spp. particularly *N. gonorrhoeae* and *N. meningitidis*, *Pseudomonas* spp., particularly *Ps. aeruginosa*, *Salmonella* spp., *Shigella* spp., *Staphylococcus* spp. particularly *S. aureus*, *Streptococcus* spp. particularly *S. pyogenes* and *pneumoniae*, *Vibrio* spp. and *Yersinia* spp. particularly *Y. pestis*. All of these bacteria cause disease in man and also there are animal models of the disease. Thus, when these bacteria are used in the method of the invention, the particular environment is an animal which they can infect and in which they cause disease. For example, when *Salmonella typhimurium* is used to infect a mouse the mouse develops a disease which serves as a model for typhoid fever in man. *Staphylococcus aureus* causes bacteraemia and renal abscess formation in mice (Albus *et al* (1991) *Infect. Immun.* **59**, 1008-1014) and endocarditis in rabbits (Perlman & Freedman (1971) *Yale J. Biol. Med.* **44**, 206-213).

It is required that a fungus or higher eukaryotic parasite is haploid for the relevant parts of its life (such as growth in the environment). Preferably, a DNA-mediated integrative transformation system is available and, when the microorganism is a human pathogen, conveniently an animal model of the human disease is available. Suitable fungi pathogenic to humans include certain *Aspergillus* spp. (for example *A. fumigatus*), *Cryptococcus neoformans* and *Histoplasma capsulatum*. Clearly the above-mentioned fungi have a haploid phase and a DNA-mediated integrative transformation systems are available for them. *Toxoplasma* may also be used, being a parasite with a haploid phase during infection. Bacteria have a haploid genome.

Animal models of human disease are often available in which the animal is a mouse, rat, rabbit, dog or monkey. It is preferred if the animal is a

mouse. Virulence genes detected by the method of the invention using an animal model of a human disease are clearly very likely to be genes which determine the virulence of the microorganism in man.

Particularly preferred microorganisms for use in the methods of the invention are *Salmonella typhimurium*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Enterococcus faecalis*, *Pseudomonas aeruginosa* and *Aspergillus fumigatus*.

A preferred embodiment of the invention is now described.

A nucleic acid comprising a unique marker sequence is made as follows. A complex pool of double stranded DNA sequence "tags" is generated using oligonucleotide synthesis and a polymerase chain reaction (PCR). Each DNA "tag" has a unique sequence of between about 20 and 80 bp, preferably about 40 bp which is flanked by "arms" of about 15 to 30 bp, preferably about 20 bp, which are common to all "tags". The number of bp in the unique sequence is sufficient to allow large numbers (for example $> 10^{10}$) of unique sequences to be generated by random oligonucleotide synthesis but not too large to allow the formation of secondary structures which may interfere with a PCR. Similarly, the length of the arms should be sufficient to allow efficient priming of oligonucleotides in a PCR.

It is well known that the sequence at the 5′ end of the oligonucleotide need not match the target sequence to be amplified.

It is usual that the PCR primers do not contain any complementary structures with each other longer than 2 bases, especially at their 3′ ends, as this feature may promote the formation of an artifactual product called

"primer dimer". When the 3' ends of the two primers hybridize, they form a "primed template" complex, and primer extension results in a short duplex product called "primer dimer".

Internal secondary structure should be avoided in primers. For symmetric PCR, a 40-60% G+C content is often recommended for both primers, with no long stretches of any one base. The classical melting temperature calculations used in conjunction with DNA probe hybridization studies often predict that a given primer should anneal at a specific temperature or that the 72°C extension temperature will dissociate the primer/template hybrid prematurely. In practice, the hybrids are more effective in the PCR process than generally predicted by simple T_m calculations.

Optimum annealing temperatures may be determined empirically and may be higher than predicted. *Taq* DNA polymerase does have activity in the 37-55°C region, so primer extension will occur during the annealing step and the hybrid will be stabilized. The concentrations of the primers are equal in conventional (symmetric) PCR and, typically, within 0.1- to 1-μM range.

The "tags" are ligated into a transposon or transposon-like element to form the nucleic acid comprising a unique marker sequence. Conveniently, the transposon is carried on a suicide vector which is maintained as a plasmid in a "helper" organism, but which is lost after transfer to the microorganism of the method of the invention. For example, the "helper" organism may be a strain of *Escherichia coli*, the microorganism of the method may be *Salmonella* and the transfer is a conjugal transfer. Although the transposon can be lost after transfer, in a proportion of cells it undergoes a transposition event through which it integrates at random, along with its unique tag, into the genome of the

microorganism used in the method. It is most preferred if the transposon or transposon-like element can be selected. For example, in the case of *Salmonella*, a kanamycin resistance gene may be present in the transposon and exconjugants are selected on medium containing kanamycin. It is also possible to complement an auxotrophic marker in the recipient cell with a functional gene in the nucleic acid comprising the unique marker. This method is particularly convenient when fungi are used in the method.

Preferably the complementing functional gene is not derived from the same species as the recipient microorganism, otherwise non-random integration events may occur.

It is also particularly convenient if the transposon or transposon-like element is carried on a vector which is maintained episomally (ie not as part of the chromosome) in the microorganism used in the method of the first aspect of the invention when in a first given condition whereas, upon changing that condition to a second given condition, the episome cannot be maintained permitting the selection of a cell in which the transposon or transposon-like element has undergone a transposition event through which it integrates at random, along with its unique tag, into the genome of the microorganism used in the method. This particularly convenient embodiment is advantageous because, once a microorganism carrying the episomal vector is made, then each time the transposition event is selected for or induced by changing the condition of the microorganism (or a clone thereof) from a first given condition to a second given condition, the transposon can integrate at a different site in the genome of the microorganism. Thus, once a master collection of microorganisms are made, each member of which contains a unique tag sequence in the transposon or transposon-like element carried on the episomal vector (when in the first given condition), it can be used repeatedly to generate

pools of random insertional mutants, each of which contains a different tag sequence (ie unique within the pool). This embodiment is particularly useful because (a) it reduces the number and complexity of manipulations required to generate the plurality ("pool") of independently mutated microorganisms in step (1) of the method; and (b) the number of different tags need only be the same as the number of microorganisms in the plurality of microorganisms in step (1) of the method. Point (a) makes the method easier to use in organisms for which transposon mutagenesis is more difficult to perform (for example, *Staphylococcus aureus*) and point (b) means that tag sequences with particularly good hybridisation characteristics can be selected therefore making quality control easier. As is described in more detail below, the "pool" size is conveniently about 100 or 200 independently-mutated microorganisms and, therefore the master collection of microorganisms is conveniently stored in one or two 96-well microtitre plates.

In a particularly preferred embodiment the first given condition is a first particular temperature or temperature range such as 25°C to 32°C, most preferably about 30°C and the second given condition is a second particular temperature or temperature range such as 35°C to 45°C, most preferably 42°C. In further preferred embodiments the first given condition is the presence of an antibiotic, such as streptomycin, and the second given condition is the absence of the said antibiotic; or the first given condition is the absence of an antibiotic and the second given condition is the presence of the said antibiotic.

Transposons suitable for integration into the genome of Gram negative bacteria include Tn5, Tn10 and derivatives thereof. Transposons suitable for integration into the genome of Gram positive bacteria include Tn916 and derivatives or analogues thereof. Transposons particularly suited for

use with *Staphylococcus aureus* include Tn917 (Cheung *et al* (1992) *Proc. Natl. Acad. Sci. USA* **89**, 6462-6466) and Tn918 (Albus *et al* (1991) *Infect. Immun.* **59**, 1008-1014).

It is particularly preferred if the transposons have the properties of the Tn917 derivatives described by Camilli *et al* (1990) *J. Bacteriol.* **172**, 3738-3744, and are carried by a temperature-sensitive vector such as pE194Ts (Villafane *et al* (1987) *J. Bacteriol.* **169**, 4822-4829).

It will be appreciated that although transposons are convenient for insertionally inactivating a gene, any other known method, or method developed in the future may be used. A further convenient method of insertionally inactivating a gene, particularly in certain bacteria such as *Streptococcus*, is using insertion-duplication mutagenesis such as that described in Morrison *et al* (1984) *J.Bacteriol* **159**, 870 with respect to *S.pneumoniae*. The general method may also be applied to other microorganisms, especially bacteria.

For fungi, insertional mutations are created by transformation using DNA fragments or plasmids carrying the "tags" and, preferably, selectable markers encoding, for example, resistance to hygromycin B or phleomycin (see Smith *et al* (1994) *Infect. Immunol.* **62**, 5247-5254). Random, single integration of DNA fragments encoding hygromycin B resistance into the genome of filamentous fungi, using restriction enzyme mediated integration (REMI; Schiestl & Petes (1991); Lu *et al* (1994) *Proc. Natl. Acad. Sci. USA* **91**, 12649-12653) are known.

A simple insertional mutagenesis technique for a fungus is described in Schiestl & Petes (1994) incorporated herein by reference, and include, for example, the use of Ty elements and ribosomal DNA in yeast.

Random integration of the transposon or other DNA sequence allows isolation of a plurality of independently mutated microorganisms wherein a different gene is insertionally inactivated in each mutant and each mutant contains a different marker sequence.

A library of such insertion mutants is arrayed in welled microtitre dishes so that each well contains a different mutant microorganism. DNA comprising the unique marker sequence from each individual mutant microorganism (conveniently, the total DNA from the clone is used) is stored. Conveniently, this is done by removing a sample of the microorganism from the microtitre dish, spotting it onto a nucleic acid hybridisation membrane (such as nitrocellulose or nylon membranes), lysing the microorganism in alkali and fixing the nucleic acid to the membrane. Thus, a replica of the contents of the welled microtitre dishes is made.

Pools of the microorganisms from the welled microtitre dish are made and DNA is extracted. This DNA is used as a target for a PCR using primers that anneal to the common "arms" flanking the "tags" and the amplified DNA is labelled, for example with ^{32}P. The product of the PCR is used to probe the DNA stored from each individual mutant to provide a reference hybridisation pattern for the replicas of the welled microtitre dishes. This is a check that each of the individual microorganisms does, in fact, contain a marker sequence and that the marker sequence can be amplified and labelled efficiently.

Pools of transposon mutants are made to introduce into the particular environment. Conveniently, 96-well microtitre dishes are used and the pool contains 96 transposon mutants. However, the lower limit for the pool is two mutants: there is no theoretical upper limit to the size of the

pool but, as discussed below, the upper limit may be determined in relation to the environment into which the mutants are introduced.

Once the microorganisms are introduced into the said particular environment those microorganisms which are able to do so are allowed to grow in the environment. The length of time the microorganisms are left in the environment is determined by the nature of the microorganism and the environment. After a suitable length of time, the microorganisms are recovered from the environment, DNA is extracted and the DNA is used as a template for a PCR using primers that anneal to the "arms" flanking the "tags". The PCR product is labelled, for example with ^{32}P, and is used to probe the DNA stored from each individual mutant replicated from the welled microtitre dish. Stored DNA are identified which hybridise weakly or not at all with the probe generated from the DNA isolated from the microorganisms recovered from environment. These non-hybridising DNAs correspond to mutants whose adaptation to the particular environment has been attenuated by insertion of the transposon or other DNA sequence.

In a particularly preferred embodiment the "arms" have no, or very little, label compared to the "tags". For example, the PCR primers are suitably designed to contain no, or a single, G residue, the ^{32}P-labelled nucleotide is dCTP and, in this case, no or one radiolabelled C residue is incorporated in each "arm" but a greater number of radiolabelled C residues are incorporated in the "tag". It is preferred if the "tag" has at least ten-fold more label incorporated than the "arms"; preferably twenty-fold or more; more preferably fifty-fold or more. Conveniently the "arms" can be removed from the "tag" using a suitable restriction enzyme, a site for which may be incorporated in the primer design.

As discussed above, a particularly preferred embodiment of the invention is when the microorganism is a pathogenic microorganism and the particular environment is an animal. In this embodiment, the size of the pool of mutants introduced into the animal is determined by (a) the number of cells of each mutant that are likely to survive in the animal (assuming a virulence gene has not been inactivated) and (b) the total inoculum of the microorganism. If the number in (a) is too low then false positive results may arise and if the number in (b) is too high then the animal may die before enough mutants have had a chance to grow in the desired way. The number of cells in (a) can be determined for each microorganism used but it is preferably more than 50, more preferably more than 100.

The number of different mutants that can be introduced into a single animal is preferably between 50 and 500, conveniently about 100. It is preferred if the total inoculum does not exceed 10^6 cells (and it is preferably 10^5 cells) although the size of the inoculum may be varied above or below this amount depending on the microorganism and the animal.

In a particularly convenient method an inoculum of 10^5 is used containing 1000 cells each of 100 different mutants for a single animal. It will be appreciated that in this method one animal can be used to screen 100 mutants compared to prior art methods which require at least 100 animals to screen 100 mutants.

However, it is convenient to inoculate three animals with the same pool of mutants so that at least two can be investigated (one as a replica to check the reliability of the method), whilst the third is kept as a back-up. Nevertheless, the method still provides a greater than 30-fold saving in the

number of animals used.

The time between the pool of mutants being introduced into the animal and the microorganisms being recovered may vary with the microorganism and animal used. For example, when the animal is a mouse and the microorganism is *Salmonella typhimurium*, the time between inoculation and recovery is about three days.

In one embodiment of the invention microorganisms are retrieved from the environment in step (5) at a site remote from the site of introduction in step (4), so that the virulence genes being investigated include those involved in the spread of the microorganism between the two sites.

For example, in a plant the microorganism may be introduced in a lesion in the stem or at one site on a leaf and the microorganism retrieved from another site on the leaf where a disease state is indicated.

In the case of an animal, the microorganism may be introduced orally, intraperitoneally, intravenously or intranasally and is retrieved at a later time from an internal organ such as the spleen. It may be useful to compare the virulence genes identified by oral administration and those identified by intraperitoneal administration as some genes may be required to establish infection by one route but not by the other. It is preferred if *Salmonella* is introduced intraperitoneally.

Other preferred environments which may be used to identify virulence genes are animal cells in culture (particularly macrophages and epithelial cells) and plant cells in culture. Although using cells in culture will be useful in its own right, it will also complement the use of the whole animal or plant, as the case may be, as the environment.

It is also preferred if the environment is a part of the animal body. Within a given host-parasite interaction, a number of different environments are possible, including different organs and tissues, and parts thereof such as the Peyer's patch.

The number of individual microorganisms (ie cells) recovered from the environment should be at least twice, preferably at least ten times, more preferably 100-times the number of different mutants introduced into the environment. For example, when an animal is inoculated with 100 different mutants around 10,000 individual microorganisms should be retrieved and their marker DNA isolated.

A further preferred embodiment comprises the steps:

(1A) removing auxotrophs from the plurality of mutants produced in step (1); or

(6A) determining whether the mutant selected in step (6) is an auxotroph; or

both (1A) and (6A).

It is desirable to distinguish an auxotroph (that is a mutant microorganism which requires growth factors not needed by the wild type or by prototrophs) and a mutant microorganism wherein a gene allowing the microorganism to adapt to a particular environment is inactivated. Conveniently, this is done between steps (1) and (2) or after step (6).

Preferably auxotrophs are not removed when virulence genes are being identified.

A second aspect of the invention provides a method of identifying a gene which allows a microorganism to adapt to a particular environment, the method comprising the method of the first aspect of the invention, followed by the additional step:

(7) isolating the insertionally-inactivated gene or part thereof from the individual mutant selected in step (6).

Methods for isolating a gene containing a unique marker are well known in the art of molecular biology.

A further preferred embodiment comprises the further additional step:

(8) isolating from a wild-type microorganism the corresponding wild-type gene using the insertionally-inactivated gene isolated in step (7) or part thereof as a probe.

Methods for gene probing are well known in the art of molecular biology.

Molecular biological methods suitable for use in the practice of the present invention are disclosed in Sambrook *et al* (1989) incorporated herein by reference.

When the microorganism is a microorganism pathogenic to an animal and the gene is a virulence gene and a transposon has been used to insertionally inactivate the gene, it is convenient for the virulence genes to be cloned by digesting genomic DNA from the individual mutant selected in step (6) with a restriction enzyme which cuts outside the transposon, ligating size-fractionated DNA containing the transposon into a plasmid, and selecting plasmid recombinants on the basis of antibiotic

resistance conferred by the transposon and not by the plasmid. The microorganism genomic DNA adjacent to the transposon is sequenced using two primers which anneal to the terminal regions of the transposon, and two primers which anneal close to the polylinker sequences of the plasmid. The sequences may be subjected to DNA database searches to determine if the transposon has interrupted a known virulence gene. Thus, conveniently, sequence obtained by this method is compared against the sequences present in the publicly available databases such as EMBL and GenBank. Finally, if the interrupted sequence appears to be in a new virulence gene, the mutation is transferred to a new genetic background (for example by phage P22-mediated transduction in the case of *Salmonella*) and the LD_{50} of the mutant strain is determined to confirm that the avirulent phenotype is due to the transposition event and not a secondary mutation.

The number of individual mutants screened in order to detect all of the virulence genes in a microorganism depends on the number of genes in the genome of the microorganism. For example, it is likely that 3000-5000 mutants of *Salmonella typhimurium* need to be screened in order to detect the majority of virulence genes whereas for *Aspergillus* spp., which has a larger genome than *Salmonella*, around 20 000 mutants are screened. Approximately 4% of non-essential *S. typhimurium* genes are thought to be required for virulence (Grossman & Saier, 1990) and, if so, the *S. typhimurium* genome contains approximately 150 virulence genes. However, the methods of the invention provide a faster, more convenient and much more practicable route to identifying virulence genes.

A third aspect of the invention provides a microorganism obtained using the method of the first aspect of the invention.

Such microorganisms are useful because they have the property of not being adapted to survive in a particular environment.

In a preferred embodiment, a pathogenic microorganism is not adapted to survive in a host organism (environment) and, in the case of microorganisms that are pathogenic to animals, particularly mammals, more particularly humans, the mutant obtained by the method of the invention may be used in a vaccine. The mutant is avirulent, and therefore expected to be suitable for administration to a patient, but it is expected to be antigenic and give rise to a protective immune response.

In a further preferred embodiment the pathogenic microorganism not adapted to survive in a host organism, obtained by the methods of the invention, is modified, preferably by the introduction of a suitable DNA sequence, to express an antigenic epitope from another pathogen. This modified microorganism can act as a vaccine for that other pathogen.

A fourth aspect of the invention provides a microorganism comprising a mutation in a gene identified using the method of the second aspect of the invention.

Thus, although the microorganism of the third aspect of the invention is useful, it is preferred if a mutation is specifically introduced into the identified gene. In a preferred embodiment, particularly when the microorganism is to be used in a vaccine, the mutation in the gene is a deletion or a frameshift mutation or any other mutation which is substantially incapable of reverting. Such gene-specific mutations can be made using standard procedures such as introducing into the microorganism a copy of the mutant gene on an autonomous replicon (such as a plasmid or viral genome) and relying on homologous

recombination to introduce the mutation into the copy of the gene in the genome of the microorganism.

Fifth and sixth aspects of the invention provide a suitable microorganism for use in a vaccine and a vaccine comprising a suitable microorganism and a pharmaceutically-acceptable carrier.

The suitable microorganism is the aforementioned avirulent mutant.

Active immunisation of the patient is preferred. In this approach, one or more mutant microorganisms are prepared in an immunogenic formulation containing suitable adjuvants and carriers and administered to the patient in known ways. Suitable adjuvants include Freund's complete or incomplete adjuvant, muramyl dipeptide, the "Iscoms" of EP 109 942, EP 180 564 and EP 231 039, aluminium hydroxide, saponin, DEAE-dextran, neutral oils (such as miglyol), vegetable oils (such as arachis oil), liposomes, Pluronic polyols or the Ribi adjuvant system (see, for example GB-A-2 189 141). "Pluronic" is a Registered Trade Mark. The patient to be immunised is a patient requiring to be protected from the disease caused by the virulent form of the microorganism.

The aforementioned avirulent microorganisms of the invention or a formulation thereof may be administered by any conventional method including oral and parenteral (eg subcutaneous or intramuscular) injection. The treatment may consist of a single dose or a plurality of doses over a period of time.

Whilst it is possible for an avirulent microorganism of the invention to be administered alone, it is preferable to present it as a pharmaceutical formulation, together with one or more acceptable carriers. The carrier(s)

must be "acceptable" in the sense of being compatible with the avirulent microorganism of the invention and not deleterious to the recipients thereof. Typically, the carriers will be water or saline which will be sterile and pyrogen free.

It will be appreciated that the vaccine of the invention, depending on its microorganism component, may be useful in the fields of human medicine and veterinary medicine.

Diseases caused by microorganisms are known in many animals, such as domestic animals. The vaccines of the invention, when containing an appropriate avirulent microorganism, particularly avirulent bacterium, are useful in man but also in, for example, cows, sheep, pigs, horses, dogs and cats, and in poultry such as chickens, turkeys, ducks and geese.

Seventh and eighth aspects of the invention provide a gene obtained by the method of the second aspect of the invention, and a polypeptide encoded thereby.

By "gene" we include not only the regions of DNA that code for a polypeptide but also regulatory regions of DNA such as regions of DNA that regulate transcription, translation and, for some microorganisms, splicing of RNA. Thus, the gene includes promoters, transcription terminators, ribosome-binding sequences and for some organisms introns and splice recognition sites.

Typically, sequence information of the inactivated gene obtained in step 7 is derived. Conveniently, sequences close to the ends of the transposon are used as the hybridisation site of a sequencing primer. The derived sequence or a DNA restriction fragment adjacent to the inactivated gene

itself is used to make a hybridisation probe with which to identify, and isolate from a wild-type organism, the corresponding wild type gene.

It is preferred if the hybridisation probing is done under stringent conditions to ensure that the gene, and not a relative, is obtained. By "stringent" we mean that the gene hybridises to the probe when the gene is immobilised on a membrane and the probe (which, in this case is >200 nucleotides in length) is in solution and the immobilised gene/hybridised probe is washed in 0.1 x SSC at 65°C for 10 min. SSC is 0.15 M NaCl/0.015 M Na citrate.

Preferred probe sequences for cloning *Salmonella* virulence genes are shown in Figures 5 and 6 and described in Example 2.

In a particularly preferred embodiment the *Salmonella* virulence genes comprise the sequence shown in Figures 5 and 6 and described in Example 2.

In further preference the genes are those contained within, or at least part of which is contained within, the sequences shown in Figures 11 and 12 and which have been identified by the method of the second aspect of the invention. The sequences shown in Figures 11 and 12 are part of a gene cluster from *Salmonella typhimurium* which I have called virulence gene cluster 2 (VGC2). The position of transposon insertions are indicated within the sequence, and these transposon insertions inactivate a virulence determinant of the organism. As is discussed more fully below, and in particular in Example 4, when the method of the second aspect of the invention is used to identify virulence genes in *Salmonella typhimurium*, many of the nucleic acid insertions (and therefore genes identified) are clustered in a relatively small part of the genome. This region, VGC2,

contains other virulence genes which, as described below, form part of the invention.

The gene isolated by the method of the invention can be expressed in a suitable host cell. Thus, the gene (DNA) may be used in accordance with known techniques, appropriately modified in view of the teachings contained herein, to construct an expression vector, which is then used to transform an appropriate host cell for the expression and production of the polypeptide of the invention. Such techniques include those disclosed in US Patent Nos. 4,440,859 issued 3 April 1984 to Rutter *et al*, 4,530,901 issued 23 July 1985 to Weissman, 4,582,800 issued 15 April 1986 to Crowl, 4,677,063 issued 30 June 1987 to Mark *et al*, 4,678,751 issued 7 July 1987 to Goeddel, 4,704,362 issued 3 November 1987 to Itakura *et al*, 4,710,463 issued 1 December 1987 to Murray, 4,757,006 issued 12 July 1988 to Toole, Jr. *et al*, 4,766,075 issued 23 August 1988 to Goeddel *et al* and 4,810,648 issued 7 March 1989 to Stalker, all of which are incorporated herein by reference.

The DNA encoding the polypeptide constituting the compound of the invention may be joined to a wide variety of other DNA sequences for introduction into an appropriate host. The companion DNA will depend upon the nature of the host, the manner of the introduction of the DNA into the host, and whether episomal maintenance or integration is desired.

Generally, the DNA is inserted into an expression vector, such as a plasmid, in proper orientation and correct reading frame for expression. If necessary, the DNA may be linked to the appropriate transcriptional and translational regulatory control nucleotide sequences recognised by the desired host, although such controls are generally available in the expression vector. The vector is then introduced into the host through

standard techniques. Generally, not all of the hosts will be transformed by the vector. Therefore, it will be necessary to select for transformed host cells. One selection technique involves incorporating into the expression vector a DNA sequence, with any necessary control elements, that codes for a selectable trait in the transformed cell, such as antibiotic resistance. Alternatively, the gene for such selectable trait can be on another vector, which is used to co-transform the desired host cell.

Host cells that have been transformed by the recombinant DNA of the invention are then cultured for a sufficient time and under appropriate conditions known to those skilled in the art in view of the teachings disclosed herein to permit the expression of the polypeptide, which can then be recovered.

Many expression systems are known, including bacteria (for example *E. coli* and *Bacillus subtilis*), yeasts (for example *Saccharomyces cerevisiae*), filamentous fungi (for example *Aspergillus*), plant cells, animal cells and insect cells.

The vectors include a prokaryotic replicon, such as the ColE1 *ori*, for propagation in a prokaryote, even if the vector is to be used for expression in other, non-prokaryotic, cell types. The vectors can also include an appropriate promoter such as a prokaryotic promoter capable of directing the expression (transcription and translation) of the genes in a bacterial host cell, such as *E. coli*, transformed therewith.

A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with exemplary bacterial hosts are typically provided in plasmid vectors containing convenient restriction sites

for insertion of a DNA segment of the present invention.

Typical prokaryotic vector plasmids are pUC18, pUC19, pBR322 and pBR329 available from Biorad Laboratories, (Richmond, CA, USA) and p*Trc*99A and pKK223-3 available from Pharmacia, Piscataway, NJ, USA.

A typical mammalian cell vector plasmid is pSVL available from Pharmacia, Piscataway, NJ, USA. This vector uses the SV40 late promoter to drive expression of cloned genes, the highest level of expression being found in T antigen-producing cells, such as COS-1 cells.

An example of an inducible mammalian expression vector is pMSG, also available from Pharmacia. This vector uses the glucocorticoid-inducible promoter of the mouse mammary tumour virus long terminal repeat to drive expression of the cloned gene.

Useful yeast plasmid vectors are pRS403-406 and pRS413-416 and are generally available from Stratagene Cloning Systems, La Jolla, CA 92037, USA. Plasmids pRS403, pRS404, pRS405 and pRS406 are Yeast Integrating plasmids (YIps) and incorporate the yeast selectable markers *HIS3*, *TRP1*, *LEU2* and *URA3*. Plasmids pRS413-416 are Yeast Centromere plasmids (YCps)

A variety of methods have been developed to operably link DNA to vectors via complementary cohesive termini. For instance, complementary homopolymer tracts can be added to the DNA segment to be inserted to the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules.

Synthetic linkers containing one or more restriction sites provide an alternative method of joining the DNA segment to vectors. The DNA segment, generated by endonuclease restriction digestion as described earlier, is treated with bacteriophage T4 DNA polymerase or *E. coli* DNA polymerase I, enzymes that remove protruding, 3'-single-stranded termini with their 3'-5'-exonucleolytic activities, and fill in recessed 3'-ends with their polymerizing activities.

The combination of these activities therefore generates blunt-ended DNA segments. The blunt-ended segments are then incubated with a large molar excess of linker molecules in the presence of an enzyme that is able to catalyze the ligation of blunt-ended DNA molecules, such as bacteriophage T4 DNA ligase. Thus, the products of the reaction are DNA segments carrying polymeric linker sequences at their ends. These DNA segments are then cleaved with the appropriate restriction enzyme and ligated to an expression vector that has been cleaved with an enzyme that produces termini compatible with those of the DNA segment.

Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including International Biotechnologies Inc, New Haven, CN, USA.

A desirable way to modify the DNA encoding the polypeptide of the invention is to use the polymerase chain reaction as disclosed by Saiki *et al* (1988) *Science* **239**, 487-491.

In this method the DNA to be enzymatically amplified is flanked by two specific oligonucleotide primers which themselves become incorporated into the amplified DNA. The said specific primers may contain restriction endonuclease recognition sites which can be used for cloning into

expression vectors using methods known in the art.

Variants of the genes also form part of the invention. It is preferred if the variant has at least 70% sequence identity, more preferably at least 85% sequence identity, most preferably at least 95% sequence identity with the genes isolated by the method of the invention. Of course, replacements, deletions and insertions may be tolerated. The degree of similarity between one nucleic acid sequence and another can be determined using the GAP program of the University of Wisconsin Computer Group.

Similarly, variants of proteins encoded by the genes are included.

By "variants" we include insertions, deletions and substitutions, either conservative or non-conservative, where such changes do not substantially alter the normal function of the protein.

By "conservative substitutions" is intended combinations such as Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr.

Such variants may be made using the well known methods of protein engineering and site-directed mutagenesis.

A ninth aspect of the invention provides a method of identifying a compound which reduces the ability of a microorganism to adapt to a particular environment comprising the steps of selecting a compound which interferes with the function of (1) a gene obtained by the method of the second aspect of the invention or of (2) a polypeptide encoded by such a gene.

Pairwise screens for compounds which affect the wild type cell but not a

cell overproducing a gene isolated by the methods of the invention form part of this aspect of the invention.

For example, in one embodiment one cell is a wild type cell and a second cell is the Salmonella which is made to overexpress the gene isolated by the method of the invention. The viability and/or growth of each cell in the particular environment is determined in the presence of a compound to be tested to identify which compound reduces the viability or growth of the wild type cell but not the cell overexpressing the said gene.

It is preferred if the gene is a virulence gene.

For example, in one embodiment the microorganism (such as *S. typhimurium*) is made to over-express the virulence gene identified by the method of the first aspect of the invention. Each of (a) the "over-expressing" microorganism and (b) an equivalent microorganism (which does not over-express the virulence gene) are used to infect cells in culture. Whether a particular test compound will selectively inhibit the virulence gene function is determined by assessing the amount of the test compound which is required to prevent infection of the host cells by (a) the over-expressing microorganism and (b) the equivalent microorganism (at least for some virulence gene products it is envisaged that the test compound will inactivate them, and itself be inactivated, by binding to the virulence gene product). If more of the compound is required to prevent infection by the (a) than (b) then this suggests that the compound is selective. It is preferred if the microorganisms (such as Salmonella) are destroyed extracellularly by a mild antibiotic such as gentamicin (which does not penetrate host cells) and that the effect of the test compound in preventing infection of the cell by the microorganism is by lysing the said cell and determining how many microorganisms are present (for example

by plating on agar).

Pairwise screens and other screens for compounds are generally disclosed in Kirsch & Di Domenico (1993) in "The Discovery of Natural Products with a Therapeutic Potential" (Ed, V.P. Gallo), Chapter 6, pages 177-221, Butterworths, V.K. (incorporated herein by reference).

Pairwise screens can be designed in a number of related formats in which the relative sensitivity to a compound is compared using two genetically related strains. If the strains differ at a single locus, then a compound specific for that target can be identified by comparing each strain's sensitivity to the inhibitor. For example, inhibitors specific to the target will be more active against a super-sensitive test strain when compared to an otherwise isogenic sister strain. In an agar diffusion format, this is determined by measuring the size of the zone of inhibition surrounding the disc or well carrying the compound. Because of diffusion, a continuous concentration gradient of compound is set up, and the strain's sensitivity to inhibitors is proportional to the distance from the disc or well to the edge of the zone. General antimicrobials, or antimicrobials with modes of action other than the desired one are generally observed as having similar activities against the two strains.

Another type of molecular genetic screen, involving pairs of strains where a cloned gene product is overexpressed in one strain compared to a control strain. The rationale behind this type of assay is that the strain containing an elevated quantity of the target protein should be more resistant to inhibitors specific to the cloned gene product than an isogenic strain, containing normal amounts of the target protein. In an agar diffusion assay, the zone size surrounding a specific compound is expected to be smaller in the strain overexpressing the target protein compared to an

otherwise isogenic strain.

Additionally or alternatively selection of a compound is achieved in the following steps:

1. A mutant microorganism obtained using the method of the first aspect of the invention is used as a control (it has a given phenotype, for example, avirulence).

2. A compound to be tested is mixed with the wild-type microorganism.

3. The wild-type microorganism is introduced into the environment (with or without the test compound).

4. If the wild-type microorganism is unable to adapt to the environment (following treatment by, or in the presence of, the compound), the compound is one which reduces the ability of the microorganism to adapt to, or survive in, the particular environment.

When the environment is an animal body and the microorganism is a pathogenic microorganism, the compound identified by this method can be used in a medicament to prevent or ameliorate infection with the microorganism.

A tenth aspect of the invention therefore provides a compound identifiable by the method of the ninth aspect.

It will be appreciated that uses of the compound of the tenth aspect are related to the method by which it can be identified, and in particular in

relation to the host of a pathogenic microorganism. For example, if the compound is identifiable by a method which uses a virulence gene, or polypeptide encoded thereby, from a bacterium which infects a mammal, the compound may be useful in treating infection of a mammal by that bacterium.

Similarly, if the compound is identifiable by a method which uses a virulence gene, or polypeptide encoded thereby, from a fungus which infects a plant, the compound may be useful in treating infection of a plant by that fungus.

An eleventh aspect of the invention provides a molecule which selectively interacts with, and substantially inhibits the function of, a gene of the seventh aspect of the invention or a nucleic acid product thereof.

By "nucleic acid product thereof" we include any RNA, especially mRNA, transcribed from the gene.

Preferably a molecule which selectively interacts with. and substantially inhibits the function of, said gene or said nucleic acid product is an antisense nucleic acid or nucleic acid derivative.

More preferably, said molecule is an antisense oligonucleotide.

Antisense oligonucleotides are single-stranded nucleic acid, which can specifically bind to a complementary nucleic acid sequence. By binding to the appropriate target sequence, an RNA-RNA, a DNA-DNA, or RNA-DNA duplex is formed. These nucleic acids are often termed "antisense" because they are complementary to the sense or coding strand of the gene. Recently, formation of a triple helix has proven possible where the

oligonucleotide is bound to a DNA duplex. It was found that oligonucleotides could recognise sequences in the major groove of the DNA double helix. A triple helix was formed thereby. This suggests that it is possible to synthesise sequence-specific molecules which specifically bind double-stranded DNA via recognition of major groove hydrogen binding sites.

Clearly, the sequence of the antisense nucleic acid or oligonucleotide can readily be determined by reference to the nucleotide sequence of the gene in question. For example, antisense nucleic acid or oligonucleotides can be designed which are complementary to a part of the sequence shown in Figures 11 or 12, especially to sequences which form a part of a virulence gene.

Oligonucleotides are subject to being degraded or inactivated by cellular endogenous nucleases. To counter this problem, it is possible to use modified oligonucleotides, eg having altered internucleotide linkages, in which the naturally occurring phosphodiester linkages have been replaced with another linkage. For example, Agrawal *et al* (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7079-7083 showed increased inhibition in tissue culture of HIV-1 using oligonucleotide phosphoramidates and phosphorothioates. Sarin *et al* (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7448-7451 demonstrated increased inhibition of HIV-1 using oligonucleotide methylphosphonates. Agrawal *et al* (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7790-7794 showed inhibition of HIV-1 replication in both early-infected and chronically infected cell cultures, using nucleotide sequence-specific oligonucleotide phosphorothioates. Leither *et al* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3430-3434 report inhibition in tissue culture of influenza virus replication by oligonucleotide phosphorothioates.

Oligonucleotides having artificial linkages have been shown to be resistant to degradation *in vivo*. For example, Shaw *et al* (1991) in *Nucleic Acids Res.* **19**, 747-750, report that otherwise unmodified oligonucleotides become more resistant to nucleases *in vivo* when they are blocked at the 3′ end by certain capping structures and that uncapped oligonucleotide phosphorothioates are not degraded *in vivo*.

A detailed description of the H-phosphonate approach to synthesizing oligonucleoside phosphorothioates is provided in Agrawal and Tang (1990) *Tetrahedron Letters* **31**, 7541-7544, the teachings of which are hereby incorporated herein by reference. Syntheses of oligonucleoside methylphosphonates, phosphorodithioates, phosphoramidates, phosphate esters, bridged phosphoramidates and bridge phosphorothioates are known in the art. See, for example, Agrawal and Goodchild (1987) *Tetrahedron Letters* **28**, 3539; Nielsen *et al* (1988) *Tetrahedron Letters* **29**, 2911; Jager *et al* (1988) *Biochemistry* **27**, 7237; Uznanski *et al* (1987) *Tetrahedron Letters* **28**, 3401; Bannwarth (1988) *Helv. Chim. Acta.* **71**, 1517; Crosstick and Vyle (1989) *Tetrahedron Letters* **30**, 4693; Agrawal *et al* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1401-1405, the teachings of which are incorporated herein by reference. Other methods for synthesis or production also are possible. In a preferred embodiment the oligonucleotide is a deoxyribonucleic acid (DNA), although ribonucleic acid (RNA) sequences may also be synthesized and applied.

The oligonucleotides useful in the invention preferably are designed to resist degradation by endogenous nucleolytic enzymes. *In vivo* degradation of oligonucleotides produces oligonucleotide breakdown products of reduced length. Such breakdown products are more likely to engage in non-specific hybridization and are less likely to be effective, relative to their full-length counterparts. Thus, it is desirable to use

oligonucleotides that are resistant to degradation in the body and which are able to reach the targeted cells. The present oligonucleotides can be rendered more resistant to degradation *in vivo* by substituting one or more internal artificial internucleotide linkages for the native phosphodiester linkages, for example, by replacing phosphate with sulphur in the linkage. Examples of linkages that may be used include phosphorothioates, methylphosphonates, sulphone, sulphate, ketyl, phosphorodithioates, various phosphoramidates, phosphate esters, bridged phosphorothioates and bridged phosphoramidates. Such examples are illustrative, rather than limiting, since other internucleotide linkages are known in the art. See, for example, Cohen, (1990) *Trends in Biotechnology*. The synthesis of oligonucleotides having one or more of these linkages substituted for the phosphodiester internucleotide linkages is well known in the art, including synthetic pathways for producing oligonucleotides having mixed internucleotide linkages.

Oligonucleotides can be made resistant to extension by endogenous enzymes by "capping" or incorporating similar groups on the 5′ or 3′ terminal nucleotides. A reagent for capping is commercially available as Amino-Link II™ from Applied BioSystems Inc, Foster City, CA. Methods for capping are described, for example, by Shaw *et al* (1991) *Nucleic Acids Res.* **19**, 747-750 and Agrawal *et al* (1991) *Proc. Natl. Acad. Sci. USA* **88(17)**, 7595-7599, the teachings of which are hereby incorporated herein by reference.

A further method of making oligonucleotides resistant to nuclease attack is for them to be "self-stabilized" as described by Tang *et al* (1993) *Nucl. Acids Res.* **21**, 2729-2735 incorporated herein by reference. Self-stabilized oligonucleotides have hairpin loop structures at their 3′ ends, and show increased resistance to degradation by snake venom

phosphodiesterase, DNA polymerase I and fetal bovine serum. The self-stabilized region of the oligonucleotide does not interfere in hybridization with complementary nucleic acids, and pharmacokinetic and stability studies in mice have shown increased *in vivo* persistence of self-stabilized oligonucleotides with respect to their linear counterparts.

In accordance with the invention, the inherent binding specificity of antisense oligonucleotides characteristic of base pairing is enhanced by limiting the availability of the antisense compound to its intend locus *in vivo*, permitting lower dosages to be used and minimizing systemic effects. Thus, oligonucleotides are applied locally to achieve the desired effect. The concentration of the oligonucleotides at the desired locus is much higher than if the oligonucleotides were administered systemically, and the therapeutic effect can be achieved using a significantly lower total amount. The local high concentration of oligonucleotides enhances penetration of the targeted cells and effectively blocks translation of the target nucleic acid sequences.

The oligonucleotides can be delivered to the locus by any means appropriate for localized administration of a drug. For example, a solution of the oligonucleotides can be injected directly to the site or can be delivered by infusion using an infusion pump. The oligonucleotides also can be incorporated into an implantable device which when placed at the desired site, permits the oligonucleotides to be released into the surrounding locus.

The oligonucleotides are most preferably administered via a hydrogel material. The hydrogel is noninflammatory and biodegradable. Many such materials now are known, including those made from natural and synthetic polymers. In a preferred embodiment, the method exploits a

hydrogel which is liquid below body temperature but gels to form a shape-retaining semisolid hydrogel at or near body temperature. Preferred hydrogel are polymers of ethylene oxide-propylene oxide repeating units. The properties of the polymer are dependent on the molecular weight of the polymer and the relative percentage of polyethylene oxide and polypropylene oxide in the polymer. Preferred hydrogels contain from about 10 to about 80% by weight ethylene oxide and from about 20 to about 90% by weight propylene oxide. A particularly preferred hydrogel contains about 70% polyethylene oxide and 30% polypropylene oxide. Hydrogels which can be used are available, for example, from BASF Corp., Parsippany, NJ, under the tradename Pluronic^{R}.

In this embodiment, the hydrogel is cooled to a liquid state and the oligonucleotides are admixed into the liquid to a concentration of about 1 mg oligonucleotide per gram of hydrogel. The resulting mixture then is applied onto the surface to be treated, for example by spraying or painting during surgery or using a catheter or endoscopic procedures. As the polymer warms, it solidifies to form a gel, and the oligonucleotides diffuse out of the gel into the surrounding cells over a period of time defined by the exact composition of the gel.

The oligonucleotides can be administered by means of other implants that are commercially available or described in the scientific literature, including liposomes, microcapsules and implantable devices. For example, implants made of biodegradable materials such as polyanhydrides, polyorthoesters, polylactic acid and polyglycolic acid and copolymers thereof, collagen, and protein polymers, or non-biodegradable materials such as ethylenevinyl acetate (EVAc), polyvinyl acetate, ethylene vinyl alcohol, and derivatives thereof can be used to locally deliver the oligonucleotides. The oligonucleotides can be incorporated into the

material as it is polymerized or solidified, using melt or solvent evaporation techniques, or mechanically mixed with the material. In one embodiment, the oligonucleotides are mixed into or applied onto coatings for implantable devices such as dextran coated silica beads, stents, or catheters.

The dose of oligonucleotides is dependent on the size of the oligonucleotides and the purpose for which is it administered. In general, the range is calculated based on the surface area of tissue to be treated. The effective dose of oligonucleotide is somewhat dependent on the length and chemical composition of the oligonucleotide but is generally in the range of about 30 to 3000 μg per square centimetre of tissue surface area.

The oligonucleotides may be administered to the patient systemically for both therapeutic and prophylactic purposes. The oligonucleotides may be administered by any effective method, for example, parenterally (eg intravenously, subcutaneously, intramuscularly) or by oral, nasal or other means which permit the oligonucleotides to access and circulate in the patient's bloodstream. Oligonucleotides administered systemically preferably are given in addition to locally administered oligonucleotides, but also have utility in the absence of local administration. A dosage in the range of from about 0.1 to about 10 grams per administration to an adult human generally will be effective for this purpose.

It will be appreciated that the molecules of this aspect of the invention are useful in treating or preventing any infection caused by the microorganism from which the said gene has been isolated, or a close relative of said microorganism. Thus, the said molecule is an antibiotic.

Thus, a twelfth aspect of the invention provides a molecule of the eleventh

aspect of the invention for use in medicine.

A thirteenth aspect of the invention provides a method of treating a host which has, or is susceptible to, an infection with a microorganism, the method comprising administering an effective amount of a molecule according to the eleventh aspect of the invention wherein said gene is present in said microorganisms, or a close relative of said microorganism.

By "effective amount" we mean an amount which substantially prevents or ameliorates the infection. By "host" we include any animal or plant which may be infected by a microorganism.

It will be appreciated that pharmaceutical formulations of the molecule of the eleventh aspect of the invention form part of the invention. Such pharmaceutical formulations comprise the said molecule together with one or more acceptable carriers. The carrier(s) must be "acceptable" in the sense of being compatible with the said molecule of the invention and not deleterious to the recipients thereof. Typically, the carriers will be water or saline which will be sterile and pyrogen free.

As mentioned above, and as described in more detail in Example 4 below, I have found that certain virulence genes are clustered in *Salmonella typhimurium* in a region of the chromosome that I have called VGC2. DNA-DNA hybridisation experiments have determined that sequences homologous to at least part of VGC2 are found in many species and strains of *Salmonella* but are not present in the *E. coli* and *Shigella* strains tested (see Example 4). These sequences almost certainly correspond to conserved genes, at least in *Salmonella*, and at least some of which are virulence genes. It is believed that equivalent genes in other *Salmonella* species and, if present, equivalent genes in other enteric or other bacteria

will also be virulence genes.

Whether a gene within the VGC2 region is a virulence gene is readily determined. For example, those genes within VGC2 which have been identified by the method of the second aspect of the invention (when applied to *Salmonella typhimurium* and wherein the environment is an animal such as a mouse) are virulence genes. Virulence genes are also identified by making a mutation in the gene (preferably a non-polar mutation) and determining whether the mutant strain is avirulent. Methods of making mutations in a selected gene are well known and are described below.

A fourteenth aspect of the invention provides the VGC2 DNA of *Salmonella typhimurium* or a part thereof, or a variant of said DNA or a variant of a part thereof.

The VGC2 DNA of *Salmonella typhimurium* is depicted diagrammatically in Figure 8 and is readily obtainable from *Salmonella typhimurium* ATCC 14028 (available from the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, USA; also deposited at the NCTC, Public Health Laboratory Service, Colindale, UK under accession no. NCTC 12021) using the information provided in Example 4. For example, probes derived from the sequences shown in Figures 11 and 12 may be used to identify λ clones from a *Salmonella typhimurium* genomic library. Standard genome walking methods can be employed to obtain all of the VGC2 DNA. The restriction map shown in Figure 8 can be used to identify and locate DNA fragments from VGC2.

By "part of the VGC2 DNA of *Salmonella typhimurium*" we mean any DNA sequence which comprises at least 10 nucleotides, preferably at least

20 nucleotides, more preferably at least 50 nucleotides, still more preferably at least 100 nucleotides, and most preferably at least 500 nucleotides of VGC2. A particularly preferred part of the VGC2 DNA is the sequence shown in Figure 11, or a part thereof. Another particularly preferred part of the VGC2 DNA is the sequence shown in Figure 12, or a part thereof.

Advantageously, the part of the VGC2 DNA is a gene, or part thereof.

Genes can be identified within the VGC2 region by statistical analysis of the open reading frames using computer programs known in the art. If an open reading frame is greater than about 100 codons it is likely to be a gene (although genes smaller than this are known). Whether an open reading frame corresponds to the polypeptide coding region of a gene can be determined experimentally. For example, a part of the DNA corresponding to the open reading frame may be used as a probe in a northern (RNA) blot to determine whether mRNA is expressed which hybridises to the said DNA; alternatively or additionally a mutation may be introduced into the open reading frame and the effect of the mutation on the phenotype of the microorganism can be determined. If the phenotype is changed then the open reading frame corresponds to a gene. Methods of identifying genes within a DNA sequence are known in the art.

By "variant of said DNA or a variant of a part thereof" we include any variant as defined by the term "variant" in the seventh aspect of the invention.

Thus, variants of VGC2 DNA of *Salmonella typhimurium* include equivalent genes, or parts thereof, from other *Salmonella* species, such as

Salmonella typhi and *Salmonella enterica*, as well as equivalent genes, or parts thereof, from other bacteria such as other enteric bacteria.

By "equivalent gene" we include genes which are functionally equivalent and those in which a mutation leads to a similar phenotype (such as avirulence). It will be appreciated that before the present invention VGC2 or the genes contained therein had not been identified and certainly not implicated in virulence determination.

Thus, further aspects of the invention provide a mutant bacterium wherein if the bacterium normally contains a gene that is the same as or equivalent to a gene in VGC2, said gene is mutated or absent in said mutant bacterium; methods of making a mutant bacterium wherein if the bacterium normally contains a gene that is the same as or equivalent to a gene in VGC2, said gene is mutated or absent in said mutant bacterium. The following is a preferred method to inactivate a VGC2 gene. One first subclones the gene on a DNA fragment from a *Salmonella* λ DNA library or other DNA library using a fragment of VGC2 as a probe in hybridisation experiments, and map the gene with respect to restriction enzyme sites and characterise the gene by DNA sequencing in *Escherichia coli*. Using restriction enzymes, one then introduces into the coding region of the gene a segment of DNA encoding resistance to an antibiotic (for example, kanamycin), possibly after deleting a portion of the coding region of the cloned gene by restriction enzymes. Methods and DNA constructs containing an antibiotic resistance marker are available to ensure that the inactivation of the gene of interest is preferably non-polar, that is to say, does not affect the expression of genes downstream from the gene of interest. The mutant version of the gene is then transferred from *E. coli* to *Salmonella typhimurium* usiing phage P22 transduction and transductants checked by Southern hybridisation for homologous

recombination of the mutant gene into the chromosome.

This approach is commonly used in *Salmonella* (and can be used in *S. typhi*), and further details can be found in many papers, including Galan *et al* (1992) **174**, 4338-4349.

Still further aspects provide a use of said mutant mutant bacterium in a vaccine; pharmaceutical compositions comprising said bacterium and a pharmaceutically acceptable carrier; a polypeptide encoded by VGC2 DNA of *Salmonella typhimurium* or a part thereof, or a variant of a part thereof; a method of identifying a compound which reduces the ability of a bacterium to infect or cause disease in a host; a compound identifiable by said method; a molecule which selectively interacts with, and substantially inhibits the function of, a gene in VGC2 or a nucleic product thereof; and medical uses and pharmaceutical compositions thereof.

The VGC2 DNA contains genes which have been identified by the methods of the first and second aspects of the invention as well as genes which have been identified by their location (although identifiable by the methods of the first and second aspects of the invention). These further aspects of the invention relate closely to the fourth, fifth, sixth, seventh, eighth, ninth, tenth, eleventh, twelfth and thirteenth aspects of the invention and, accordingly, the information given in relation to those aspects, and preferences expressed in relation to those aspects, applies to these further aspects.

It is preferred if the gene is from VGC2 or is an equivalent gene from another species of *Salmonella* such as *S. typhi*. It is preferred if the mutant bacterium is a *S. typhimurium* mutant or a mutant of another species of *Salmonella* such as *S. typhi*.

It is believed that at least some of the genes in VGC2 confer the ability for the bacterium, such as *S. typhimurium*, to enter cells.

The invention will now be described with reference to the following Examples and Figures wherein:

Figure 1 illustrates diagrammatically one particularly preferred method of the invention.

Figure 2 shows a Southern hybridisation analysis of DNA from 12 *S. typhimurium* exconjugants following digestion with *Eco*RV. The filter was probed with the kanamycin resistance gene of the mini-Tn*5* transposon.

Figure 3 shows a colony blot hybridisation analysis of DNA from 48 *S. typhimurium* exconjugants from a half of a microtitre dish (A1-H6). The filter was hybridised with a probe comprising labelled amplified tags from DNA isolated from a pool of the first 24 colonies (A1-D6).

Figure 4 shows a DNA colony blot hybridisation analysis of 95 *S. typhimurium* exconjugants of a microtitre dish (A1-H11), which were injected into a mouse. Replicate filters were hybridised with labelled amplified tags from the pool (inoculum pattern), or with labelled amplified tags from DNA isolated from over 10,000 pooled colonies that were recovered from the spleen of the infected animal (spleen pattern). Colonies B6, A11 and C8 gave rise to weak hybridisation signals on both sets of filters. Hybridisation signals from colonies A3, C5, G3 (*aroA*), and F10 are present on the inoculum pattern but not on the spleen pattern.

Figure 5 shows the sequence of a *Salmonella* gene isolated using the method of the invention and a comparison to the *Escherichia coli* clp

protease genome.

Figure 6 shows partial sequences of further *Salmonella* gene isolated using the method of the invention (SEQ ID Nos. 8 to 36).

Figure 7 shows the mapping of VGC2 on the *S. typhimurium* chromosome. (A) DNA probes from three regions of VGC2 were used in Southern hybridisation analysis of lysates from a set of *S. typhimurium* strains harbouring locked in Mu*d*-P22 prophages. Lysates which hybridised to a 7.5 kb *Pst*I fragment (probe A in Figure 8) are shown. The other two probes used hybridised to the same lysates. (B) The insertion points and packaging directions of the phage are shown along with the map position in minutes (edition VIII, ref 22 in Example 4). The phage designations correspond to the following strains: 18P, TT15242; 18Q, 15241; 19P, TT15244; 19Q, TT15243; 20P, TT15246 and 20Q, TT15245 (Ref in Example 4). The locations of mapped genes are shown by horizontal bars and the approximate locations of other genes are indicated.

Figure 8 shows a physical and genetic map of VGC2. (A) The positions of 16 transposon insertions are shown above the line. The extent of VGC2 is indicated by the thicker line. The position and direction of transcription of ORFs described in the text of Example 4 are shown by arrows below the line, together with the names of similar genes, with the exception of ORFs 12 and 13 whose products are similar to the sensor and regulatory components respectively, of a variety of two component regulatory systems. (B) The location of overlapping clones and an *Eco*RI/*Xba*I restriction fragment from Mu*d*-P22 prophage strain TT15244 are shown as filled bars. Only the portions of the λ clones which have been mapped are shown and the clones may extend beyond these limits.

(C) The positions of restriction sites are marked: B, *Bam*HI; E, *Eco*RI; V, *Eco*RV; H, *Hin*dIII; P, *Pst*I and X, *Xba*I. The positions of the 7.5 kb *Pst*I fragment (probe A) used as a probe in Figure 7 and that of the 2.2 kb *Pst*I/*Hin*dIII fragment (probe B) used as a probe in Figure 10 are shown below the restriction map. The positions of Sequence 1 (described in Figure 11) and Sequence 2 (described in Figure 12) are shown by the thin arrows (labelled Sequence 1 and Sequence 2).

Figure 9 describes mapping the boundaries of VGC2. (A) The positions of mapped genes at minutes 37 to 38 on the *E. coli* K12 chromosome are aligned with the corresponding region of the *S. typhimurium* LT2 chromosome (minutes 30 to 31). An expanded map of the VGC2 region is shown with 11 *S. typhimurium* (*S. t.*) DNA fragments used as probes (thick bars) and the restriction sites used to generate them: B, *Bam*HI; C, *Cla*I; H, *Hin*dII; K, *Kpn*I; P, *Pst*I; N, *Nsi*I and S, *Sal*I. Probes that hybridised to *E. coli* K12 (*E. c.*) genomic DNA are indicated by +; those which failed to hybridise are indicated by -.

Figure 10 shows that VGC2 is conserved among and specific to the *Salmonellae*. Genomic DNA from *Salmonella* serovars and other pathogenic bacteria was restricted with *Pst*I (A), *Hin*dIII or *Eco*RV (B) and subjected to Southern hybridisation analysis, using a 2.2 kb *Pst*I/*Hin*dIII fragment from λ clone 7 as a probe (probe B Figure 2). The filters were hybridised and washed under stringent (A) or non-stringent (B) conditions.

Figure 11 shows the DNA sequence of "Sequence 1" of VGC2 from the centre to the left-hand end (see the arrow labelled Sequence 1 in Figure 2). The DNA is translated in all six reading frames and the start and stop positions of putative genes. and the transposon insertion positions for

various mutants identified by STM are indicated (SEQ ID No 37).

As is conventional a * indicates a stop codon and standard nucleotide ambiguity codes are used where necessary.

Figure 12 shows the DNA sequence of "Sequence 2" of VGC2 (cluster C) (see the arrow labelled Sequence 2 in Figure 2). The DNA is translated in all six reading frames and the start and stop positions of putative genes, and the transposon insertion positions for various mutants identified by STM are indicated (SEQ ID No 38).

As is conventional a * indicates a stop codon and standard nucleotide ambiguity codes are used where necessary.

Figures 7 to 12 are most relevant to Example 4.

Example 1: Identification of virulence genes in *Salmonella typhimurium*

Materials and Methods

Bacterial Strains and Plasmids

Salmonella typhimurium strain 12023 (equivalent to American Type Culture Collection (ATCC) strain 14028) was obtained from the National Collection of Type Cultures (NCTC), Public Health Laboratory Service, Colindale, London, UK. A spontaneous nalidixic acid resistant mutant of this strain (12023 Nal^r) was selected in our laboratory. Another derivative of strain 12023, CL1509 (*aroA*::Tn*10*) was a gift from Fred Heffron. *Escherichia coli* strains CC118 λ*pir* (Δ[*ara-leu*], *araD*, Δ*lacX74*, *galE*,

galK, phoA20, thi-1, rpsE, rpoB, argE(Am), *recA1*, λ*pir* phage lysogen) and S17-1 λ*pir* (Tp^r, Sm^r, *recA, thi, pro, hsdR$^-$M$^+$*, RP4:2-Tc:Mu:KmTn*7*, λ*pir*) were gifts from Kenneth Timmis. *E. coli* DH5α was used for propagating pUC18 (Gibco-BRL) and Bluescript (Stratagene) plasmids containing *S. typhimurium* DNA. Plasmid pUTmini-Tn*5*Km2 (de Lorenzo *et al*, 1990) was a gift from Kenneth Timmis.

Construction of semi-random sequence tags and ligations

The oligonucleotide pool RT1(5′-CTAGGTACCTACAACCTCAAGCTT-$[NK]_{20}$-AAGCTTGGTTAGAATGGGTACCATG-3′) (SEQ ID No 1), and primers P2 (5′-TACCTACAACCTCAAGCT-3′) (SEQ ID No 2), P3 (5′-CATGGTACCCATTCTAAC-3′) (SEQ ID No 3), P4 (5′-TACCCATTCTAACCAAGC-3′) (SEQ ID No 4) and P5 (5′-CTAGGTACCTACAACCTC-3′) (SEQ ID No 5) were synthesized on a oligonucleotide synthesizer (Applied Biosystems, model 380B). Double stranded DNA tags were prepared from RT1 in a 100 μl volume PCR containing 1.5 mM $MgCl_2$, 50 mM KCl, and 10 mM Tris-Cl (pH 8.0) with 200 pg of RT1 as target; 250 μM each dATP, dCTP, dGTP, dTTP; 100 pM of primers P3 and P5; and 2.5 U of Amplitaq (Perkin-Elmer Cetus). Thermal cycling conditions were 30 cycles of 95°C for 30 s, 50°C for 45 s, and 72°C for 10 s. The PCR product was gel purified (Sambrook *et al*, 1989), passed through an elutipD column (available from Schleicher and Schull) and digested with *Kpn*I prior to ligation into pUC18 or pUTmini-Tn*5*Km2. For ligations, plasmids were digested with *Kpn*I and dephosphorylated with calf intestinal alkaline phosphatase (Gibco-BRL). Linearized plasmid molecules were gel-purified (Sambrook *et al*, 1989) prior to ligation to remove any residual uncut plasmid DNA from the digestion. Ligation reactions contained approximately 50 ng each of plasmid and double stranded tag DNA in a 25 μl volume with 1 unit T4 DNA ligase (Gibco-BRL) in a buffer supplied with the enzyme.

Ligations were carried out for 2 h at 24°C. To determine the proportion of bacterial colonies arising from either self ligation of the plasmid DNA or uncut plasmid DNA, a control reaction was carried out in which the double stranded tag DNA was omitted from the ligation reaction. This yielded no ampicillin resistant bacterial colonies following transformation of *E. coli* CC118 (Sambrook *et al*, 1989), compared with 185 colonies arising from a ligation reaction containing the double stranded tag DNA.

Bacterial Transformation and Matings

The products of several ligations between pUT mini-Tn*5*Km2 and the double stranded tag DNA were used to transform *E. coli* CC118 (Sambrook *et al*, 1989). A total of approximately 10,300 transformants were pooled and plasmid DNA extracted from the pool was used to transform *E. coli* S-17 λ*pir* (de Lorenzo & Timmis, 1994). For mating experiments, a pool of approximately 40,000 ampicillin resistant *E. coli* S-17 λ*pir* transformants, and *S. typhimurium* 12023 Nal^r were cultured separately to an optical density $(OD)_{580}$ of 1.0. Aliquots of each culture (0.4 ml) were mixed in 5 ml 10 mM $MgSO_4$, and filtered through a Millipore membrane (0.45 μm diameter). The filters were placed on the surface of agar containing M9 salts (de Lorenzo & Timmis, 1994) and incubated at 37°C for 16 h. The bacteria were recovered by shaking the filters in liquid LB medium for 40 min at 37°C and exconjugants were selected by plating the suspension onto LB medium containing 100 μg ml^{-1} nalidixic acid (to select against the donor strain) and 50 μg ml^{-1} kanamycin (to select for the recipient strain). Each exconjugant was checked by transferring nalidixic acid resistant (nal^r), kanamycin resistant (kan^r) colonies to MacConkey Lactose indicator medium (to distinguish between *E. coli* and *S. typhimurium*), and to LB medium containing ampicillin. Approximately 90% of the nal^r, kan^r colonies were sensitive to ampicillin, indicating that these resulted from authentic

transposition events (de Lorenzo & Timmis, 1994). Individual ampicillin-sensitive exconjugants were stored in 96 well microtitre dishes containing LB medium. For long term storage at -80°C, either 7% DMSO or 15% glycerol was included in the medium.

Phenotypic characterisation of mutants

Mutants were replica plated from microtitre dishes onto solid medium containing M9 salts and 0.4% glucose (Sambrook *et al*, 1989) to identify auxotrophs. Mutants with rough colony morphology were detected by low magnification microscopy of colonies on agar plates.

Colony Blots, DNA extractions, PCRs, DNA labelings and hybridisations

For colony blot hybridizations, a 48-well metal replicator (Sigma) was used to transfer exconjugants from microtitre dishes to Hybond N nylon filters (Amersham, UK) that had been placed on the surface of LB agar containing 50 μg ml^{-1} kanamycin. After overnight incubation at 37°C, the filters supporting the bacterial colonies were removed and dried at room temperature for 10 min. The bacteria were lysed with 0.4 N NaOH and the filters washed with 0.5 N Tris-Cl pH 7.0 according to the filter manufacturer's instructions. The bacterial DNA was fixed to the filters by exposure to UV light from a Stratalinker (Stratagene). Hybridisations to ^{32}P-labelled probes were carried out under stringent conditions as previously described (Holden *et al*, 1989). For DNA extractions, *S. typhimurium* transposon mutant strains were grown in liquid LB medium in microtitre dishes or resuspended in LB medium following growth on solid media. Total DNA was prepared by the hexadecyltrimethylammoniumbromide (CTAB) method according to Ausubel *et al* (1987). Briefly, cells from 150 to 1000 μl volumes were precipitated by centrifugation and resuspended in

576 μl TE. To this was added 15 μl of 20% SDS and 3 μl of 20 mg ml^{-1} proteinase K. After incubating at 37°C for 1 hour, 166 μl of 3 M NaCl was added and mixed thoroughly, followed by 80 μl of 10% (w/v) CTAB and 0.7 M NaCl. After thorough mixing, the solution was incubated at 65°C for 10 min. Following extraction with phenol and phenol-chloroform, the DNA was precipitated by addition of isopropanol, washed with 70% ethanol and resuspended in TE at a concentration of approximately 1 μg μl^{-1}.

The DNA samples were subjected to two rounds of PCR to generate labelled probes. The first PCR was performed in 100 μl reactions containing 20 mM Tris-Cl pH 8.3; 50 mM KCl; 2 mM $MgCl_2$; 0.01% Tween 80; 200 μM each dATP, dCTP, dGTP, dTTP; 2.5 units of Amplitaq polymerase (Perkin-Elmer Cetus); 770 ng each primer P2 and P4; and 5 μg target DNA. After an initial denaturation of 4 min at 95°C, thermal cycling consisted of 20 cycles of 45 s at 50°C, 10 s at 72°C, and 30 s at 95°C. PCR products were extracted with chloroform/isoamyl alcohol (24/1) and precipitated with ethanol. DNA was resuspended in 10 μl TE and the PCR products were purified by electrophoresis through a 1.6% Seaplaque (FMC Bioproducts) gel in TAE buffer. Gel slices containing fragments of about 80 bp were excised and used for the second PCR. This reaction was carried out in a 20 μl total volume, and contained 20 mM Tris-Cl pH 8.3; 50 mM KCl; 2 mM $MgCl_2$; 0.01% Tween 80; 50 μM each dATP, dTTP, dGTP; 10 μl ^{32}P-dCTP (3000 Ci/mmol, Amersham); 150 ng each primer P2 and P4; approximately 10 ng of target DNA (1-2 μl of 1.6% Seaplaque agarose containing the first round PCR product); 0.5 units of Amplitaq polymerase. The reaction was overlayed with 20 μl mineral oil and thermal cycling was performed as described above. Incorporation of the radioactive label was quantitated by absorbance to Whatman DE81 paper (Sambrook *et al*, 1989).

Infection Studies

Individual *Salmonella* exconjugants containing tagged transposons were grown in 2% tryptone, 1% yeast extract, 0.92% v/v glycerol, 0.5% Na_2PO_4, 1% KNO_3 (TYGPN medium) (Ausubel *et al*, 1987) in microtitre plates overnight at 37°C. A metal replicator was used to transfer a small volume of the overnight cultures to a fresh microtitre plate and the cultures were incubated at 37°C until the OD_{580} (measured using a Titertek Multiscan microtitre plate reader) was approximately 0.2 in each well. Cultures from individual wells were then pooled and the OD_{550} determined using a spectrophotometer. The culture was diluted in sterile saline to approximately $5x10^5$ cfu ml^{-1}. Further dilutions were plated out onto TYGPN containing nalidixic acid (100 mg ml^{-1}) and kanamycin (50 mg ml^{-1}) to confirm the cfu present in the inoculum.

Groups of three female BALB/c mice (20-25g) were injected intraperitoneally with 0.2 ml of bacterial suspension containing approximately $1x10^5$ cfu ml^{-1}. Mice were sacrificed three days post-inoculation and their spleens were removed to recover bacteria. Half of each spleen was homogenized in 1 ml of sterile saline in a microfuge tube. Cellular debris was allowed to settle and 1 ml of saline containing cells still in suspension was removed to a fresh tube and centrifuged for two minutes in a microfuge. The supernatant was aspirated and the pellet resuspended in 1 ml of sterile distilled water. A dilution series was made in sterile distilled water and 100 ml of each dilution was plated onto TYGPN agar containing nalidixic acid (100 ug ml^{-1}) and kanamycin (50 ug ml^{-1}). Bacteria were recovered from plates containing between 1000 and 4000 colonies, and a total of over 10,000 colonies recovered from each spleen were pooled and used to prepare DNA for PCR generation of probes to screen colony blots.

Virulence gene cloning and DNA sequencing

Total DNA was isolated from *S. typhimurium* exconjugants and digested separately with *Sst*I, *Sal*I, *Pst*I and *Sph*I. Digests were fractionated through agarose gels, transferred to Hybond N^+ membranes (Amersham) and subjected to Southern hybridisation analysis using the kanamycin resistance gene of pUT mini-Tn*5*Km2 as a probe. The probe was labelled with digoxygenin (Boehringer-Mannheim) and chemiluminescence detection was carried out according to the manufacturer's instructions. The hybridisation and washing conditions were as described above. Restriction enzymes which gave rise to hybridising fragments in the 3-5 kb range were used to digest DNA for a preparative agarose gel, and DNA fragments corresponding to the sizes of the hybridisation signals were excised from this, purified and ligated into pUC18. Ligation reactions were used to transform *E. coli* DH5a to kanamycin resistance. Plasmids from kanamycin-resistant transformants were purified by passage through an elutipD column and checked by restriction enzyme digestion. Plasmid inserts were partially sequenced by the di-deoxy method (Sanger *et al*, 1977) using the -40 primer and reverse sequencing primer (United States Biochemical Corporation) and the primers P6 (5'-CCTAGGCGGCCAGATCTGAT-3') (SEQ ID No 6) and P7 (5'GCACTTGTGTATAAGAGTCAG-3') (SEQ ID No 7) which anneal to the I and O termini of Tn*5*, respectively. Nucleotide sequences and deduced amino acid sequences were assembled using the Macvector 3.5 software package run on a Macintosh SE/30 computer. Sequences were compared with the EMBL and Genbank DNA databases using the UNIX/SUN computer system at the Human Genome Mapping Project Resource Centre, Harrow, UK.

Results

Tag Design

The structure of the DNA tags is shown in Figure 1a. Each tag consists of a variable central region flanked by "arms", of invariant sequence. The central region sequence ($[NK]_{20}$) was designed to prevent the occurrence of sites for the commonly used 6 bp-recognition restriction enzymes, but is sufficiently variable to ensure that statistically, the same sequence should only occur once in 2×10^{11} molecules (DNA sequencing of 12 randomly selected tags showed that none shared more than 50% identity over the variable region). (N means any base (A, G, C or T) and K means G or T.) The arms contain *Kpn*I sites close to the ends to facilitate the initial cloning step, and the *Hin*dIII sites bordering the variable region were used to release radiolabelled variable regions from the arms prior to hybridisation analysis. The arms were also designed such that primers P2 and P4 each contain only one guanine residue. Therefore during a PCR using these primers, only one cytosine will be incorporated into each newly synthesised arm, compared to an average of ten in the unique sequence. When radiolabelled dCTP is included in the PCR, an average of ten-fold more label will be present in the unique sequence compared with each arm. This is intended to minimise background hybridisation signals from the arms, after they have been released from the unique sequences by digestion with *Hin*dIII. Double stranded tags were ligated into the *Kpn*I site of the mini-Tn*5* transposon Km2, carried on plasmid pUT (de Lorenzo & Timmis, 1994). Replication of this plasmid is dependent on the R6K-specified π product of the *pir* gene. It carries the *oriT* sequence of the RP4 plasmid, permitting transfer to a variety of bacterial species (Miller & Mekalanos, 1988), and the *tnp*$^{+}$ gene needed for transposition of the mini-Tn*5* element. The tagged mini-Tn*5* transposons were transferred to *S. typhimurium* by conjugation, and 288

exconjugants resulting from transposition events were stored in the wells of microtitre dishes. Total DNA isolated from 12 of these was digested with *Eco*RV, and subjected to Southern hybridisation analysis using the kanamycin resistance gene of the mini-Tn*5* transposon as a probe. In each case, the exconjugant had arisen as a result a single integration of the transposon into a different site of the bacterial genome (Figure 2).

Specificity and sensitivity studies

We next determined the efficiency and uniformity of amplification of the DNA tags in PCRs involving pools of exconjugant DNAs as targets for the reactions. In an attempt to minimise unequal amplification of tags in the PCR, we determined the maximum quantity of DNA target that could be used in a 100 μl reaction, and the minimum number of PCR cycles, that resulted in products which could be visualised by ethidium bromide staining of an agarose gel (5 μg DNA and 20 cycles, respectively).

S. typhimurium exconjugants which had reached stationary growth phase in microtitre dishes were combined, and used to extract DNA. This was subjected to a PCR using primers P2 and P4. PCR products of 80 bp were gel-purified and used as targets for a second PCR, using the same primers but with ^{32}P-labelled CTP. This resulted in over 60% of the radiolabelled dCTP being incorporated into the PCR products. The radiolabelled products were digested with *Hin*dIII and used to probe colony blotted DNA from their corresponding microtitre dishes. Of the 1510 mutants tested in this way, 358 failed to yield a clear signal on an autoradiogram following an overnight exposure of the colony blot. There are three potential explanations for this. Firstly, it is possible that a proportion of the transposons did not carry tags. However, by comparing the transformation frequencies resulting from ligation reactions involving the transposon in the

presence or absence of tags, it seems unlikely that untagged transposons could account for more than approximately 0.5% of the total (see Materials and Methods). More probable causes are that the variable sequence was truncated in some of the tags, and/or that some of the sequences formed secondary structures, both of which might have prevented amplification. Mutants which failed to give clear signals were not included in further studies. The specificity of the efficiently amplifiable tags was demonstrated by generating a probe from 24 colonies of a microtitre dish, and using it to probe a colony blot of 48 colonies, which included the 24 used to generate the probe. The lack of any hybridisation signal from the 24 colonies not used to generate the probe (Figure 3) shows that the hybridisation conditions employed were sufficiently stringent to prevent cross-hybridisation among labelled tags, and suggests that each exconjugant is not reiterated within a microtitre dish.

There are further considerations in determining the maximum pool size that can be used as an inoculum in animal experiments. As the quantity of labelled tag for each transposon is inversely proportional to the complexity of the tag pool, there is a limit to the pool size above which hybridisation signals become too weak to be detected after overnight exposure of an autoradiogram. More importantly, as the complexity of the pool increases, so must the likelihood of failure of a virulent representative of the pool to be present in sufficient numbers, in the spleen of an infected animal, to produce enough labelled probe. We have not determined the upper limit for pool size in the murine model of salmonellosis that we have employed, but it must be in excess of 96.

Virulence tests of the transposon mutants

A total of 1152 uniquely tagged insertion mutants (from two microtitre

dishes) were tested for virulence in BALB/c mice in twelve pools, each representing a 96-well microtitre dish. Animals received an intraperitoneal injection of approximately 10^3 cells of each of 96 transposon mutants of a microtitre dish (10^5 organisms in total). Three days after injection mice were sacrificed, and bacteria were recovered by plating spleen homogenates onto laboratory medium. Approximately 10,000 colonies recovered from each mouse were pooled and DNA was extracted. The tags present in this DNA sample were amplified and labelled by the PCR, and colony blots probed and compared with the hybridisation pattern obtained using tags amplified from the inoculum (Figure 3). As a control, an *aroA* mutant of *S. typhimurium* was tagged and employed as one of the 96 mutants in the inoculum. This strain would not be expected to be recovered in the spleen because its virulence is severely attenuated (Buchmeier *et al*, 1993). Forty-one mutants were identified whose DNA hybridized to labelled tags from the inoculum but not from labelled tags from bacteria recovered from the spleen. The experiment was repeated and the same forty-one mutants were again identified. Two of these were the *aroA* mutant (one per pool), as expected. Another was an auxotrophic mutant (it failed to grow on minimal medium). All of the mutants had normal colony morphology.

Example 2: Cloning and partial characterisation of sequences flanking the transposon

DNA was extracted from one of the mutants described in Example 1 (Pool 1, F10), digested with *SstI*, and subcloned on the basis of kanamycin resistance. The sequence of 450 bp flanking one end of the transposon was determined using primer P7. This sequence shows 80% identity to the *E. coli* clp (lon) gene, which encodes a heat-regulated protease (Figure 5). To our knowledge, this gene has not previously been implicated as a virulence determinant.

Partial sequences of thirteen further *Salmonella typhimurium* virulence genes are shown in Figure 6 (sequences A2 to A9 and B1 to B5). Deduced amino acid sequences of P2D6, S4C3, P3F4, P7G2 and P9B7 bear similarities to a family of secretion-associated proteins that have been conserved throughout bacterial pathogens of animals and plants, and which are known in *Salmonella* as the *inv* family. In *S. typhimurium* the *inv* genes are required for bacterial invasion into intestinal tissue. The virulence of *inv* mutants is attenuated when they are inoculated by the oral route, but not when they are administered intraperitoneally. The discovery of *inv*-related genes that are required for virulence following intraperitoneal inoculation suggests a new secretion apparatus which might be required for invasion of non-phagocytic cells of the spleen and other organs. The products of these new genes might represent better drug targets than the inv proteins in the treatment of established infections.

Further characterisation of the genes identified in this example is described in Example 4.

Example 3: LD_{50} determinations and mouse vaccination study

Mutations identified by the method of the invention attenuate virulence.

Five of the mutations in genes not previously implicated in virulence were transferred by P22-mediated transduction to the nalidixic acid-sensitive parent strain of *S. typhimurium* 12028. Transductants were checked by restriction mapping then injected by the intraperitoneal route into groups of BALB/c mice to determine their 50% lethal dose (LD_{50}). The LD_{50} values for mutants S4C3, P7G2, P3F4 and P9B7 were all several orders of magnitude higher than that of the wild-type strain. No difference in the LD_{50} was detected for mutant P1F10; however, there was a statistically

significant decrease in the proportion of P1F10 cells recovered from the spleens of mice injected with an inoculum consisting of an equal proportion of this strain and the wild-type strain. This implies that this mutation does attenuate virulence, but to a degree that is not detectable by LD_{50}.

Mutants P3F4 and P9B7 were also administered by the oral route at an inoculum level of 10^7 cells/mouse. None of the mice became ill, indicating that the oral LD_{50} levels of these mutants are at least an order of magnitude higher than that of the wild-type strain.

In the mouse vaccination study groups of five female BALB/c mice of 20-25 g in mass were initially inoculated orally (p.o.) or intraperitoneally (i.p.) with serial ten fold dilutions of *Salmonella typhimurium* mutant strains P3F4 and P9B7. After four weeks the mice were then inoculated with 500 c.f.u. of the parental wild type strain. Deaths were then recorded over four weeks.

A group of two mice of the same age and batch as the mice inoculated with the mutant strains were also inoculated i.p. with 500 c.f.u. of the wild type strain as a positive control. Both non-immunised mice died as expected within four weeks.

Results are tabulated below:

1) p.o. initial inoculation with mutant strain P3F4

initial inoculum in c.f.u.	no. mice surviving first challenge	no. mice surviving wild type challenge
5×10^9	5	2 (40%)
5×10^8	5	2 (40%)

5×10^7	5	0 (0%)

2) i.p. initial inoculum with mutant strain P3F4

initial inoculum in c.f.u.	no. mice surviving first challenge	no. mice surviving wild type challenge
5×10^6	3	3 (100%)
5×10^5	5	4 (80%)
5×10^4	6	5 (83%)
$5 X 10^3$	5	4 (80%)

3) p.o. initial inoculum with mutant strain P9B7

initial inoculum in c.f.u.	no. mice surviving first challenge	no. mice surviving wild type challenge
5×10^9	5	0 (0%)

4) i.p. initial inoculum with mutant P9B7

initial inoculum in c.f.u.	no. mice surviving first challenge	no. mice surviving wild type challenge
5×10^6	4	2 (50%)

From these experiments I conclude that mutant P3P4 appears to give some protection against subsequent wild type challenge. This protection appears greater in mice that were immunised i.p.

Example 4: Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium*

Abbreviations used in this Example are VGC1, virulence gene cluster 1; VGC2, virulence gene cluster 2.

Background to the experiments described

Salmonella typhimurium is a principal agent of gastroenteritis in humans and produces a systemic illness in mice which serves as a model for human typhoid fever (1). Following oral inoculation of mice with *S. typhimurium*, the bacteria pass from the lumen of the small intestine through the intestinal mucosa, *via* enterocytes or M cells of the Peyer's patch follicles (2). The bacteria then invade macrophages and neutrophils, enter the reticuloendothelial system and disseminate to other organs, including the spleen and liver, where further reproduction results in an overwhelming and fatal bacteremia (3). To invade host cells, to survive and replicate in a variety of physiologically stressful intracellular and extracellular environments and to circumvent the specific antibacterial activities of the immune system, *S. typhimurium* employs a sophisticated repertoire of virulence factors (4).

To gain a more comprehensive understanding of virulence mechanisms of *S. typhimurium* and other pathogens the transposon mutagenesis system described in Example 1, which is conveniently called 'signature-tagged mutagenesis' (STM), which combines the strength of mutational analysis with the ability to follow simultaneously the fate of a large number of different mutants within a single animal (5 and Example 1; Reference 5 was published after the priority date for this invention). Using this approach we identified 43 mutants with attenuated virulence from a total of 1152 mutants that were screened. The nucleotide sequences of DNA flanking the insertion points of transposons in 5 of these mutants showed that they were related to genes encoding type III secretion systems of a variety of bacterial pathogens (6, 7). The products of the *inv/spa* gene cluster of *S. typhimurium* (8, 9) are proteins that form a type III secretion system required for the assembly of surface appendages mediating entry into

epithelial cells (10). Hence the virulence of strains carrying mutations in the *inv/spa* cluster is attenuated only if the inoculum is administered orally and not when given intraperitoneally (8). In contrast the 5 mutants identified by STM are avirulent following intraperitoneal inoculation (5).

In this example we show that the transposon insertion points of these 5 mutants and an additional 11 mutants identified by STM all map to the same region of the *S. typhimurium* chromosome. Further analysis of this region reveals additional genes whose deduced products have sequence similarity to other components of type III secretion systems. This chromosomal region which we refer to as virulence gene cluster 2 (VGC2) is not present in a number of other enteric bacteria, and represents an important locus for *S. typhimurium* virulence.

Materials and Methods

Bacterial Strains, Transduction and Growth Media. *Salmonella enterica* serotypes 5791 (*aberdeen*), 423180 (*gallinarum*), 7101 (*cubana*) and 12416 (*typhimurium* LT2) were obtained from the National Collections of Type Cultures, Public Health Laboratory Service, UK. *Salmonella typhi* BRD123 genomic DNA was a gift from G. Dougan, enteropathogenic *Escherichia coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), *Vibrio cholera* biotype *El Tor, Shigella flexneri* serotype 2 and *Staphylococcus aureus* were clinical isolates obtained from the Department of Infectious Diseases and Bacteriology, Royal Postgraduate Medical School, UK. Genomic DNA from *Yersinia pestis* was a gift from J. Heesemann. However, genomic DNA can be isolated using standard methods. The bacterial strains and the methods used to generate signature-tagged mini-Tn*5* transposon mutants of *S. typhimurium* NCTC strain 12023 have been described previously (5, 11). Routine propagation of plasmids was in *E. coli* DH5α. Bacteria were

grown in LB broth (12) supplemented with the appropriate antibiotics. Before virulence levels of individual mutant strains were assessed, the mutations were first transferred by phage P22 mediated transduction (12) to the nalidixic acid sensitive parental strain of *S. typhimurium* 12023. Transductants were analysed by restriction digestion and Southern hybridisation before use as inoculum.

Lambda Library Screening. Lambda (λ) clones with overlapping insert DNAs covering VGC2 were obtained by standard methods (13) from a λ1059 library (14) containing inserts from a partial *Sau*3A digest of *S. typhimurium* LT2 genomic DNA. The library was obtained *via* K. Sanderson, from the Salmonella Genetic Stock Centre (SGSC), Calgary, Canada.

Mu*d*-P22 Lysogens. Radiolabelled DNA probes were hybridised to Hybond N (Amersham) filters bearing DNA prepared from lysates of a set of *S. typhimurium* strains harbouring Mu*d*-P22 prophages at known positions in the *S. typhimurium* genome. Preparation of mitomycin-induced Mu*d*-P22 lysates was as described (12, 15). The set of Mu*d*-P22 prophages was originally assembled by Benson and Goldman (16) and was obtained from the SGSC.

Gel Electrophoresis and Southern Hybridisation. Gel electrophoresis was performed in 1% or 0.6% agarose gels run in 0.5 x TBE. Gel fractionated DNA was transferred to Hybond N or N+ membranes (Amersham) and stringent hybridisation and washing procedures (permitting hybridisation between nucleotide sequences with 10% or less mismatches) were as described by Holden *et al*, (17). For non-stringent conditions (permitting hybridisation between sequences with 50% mismatches) filters were hybridised overnight at 42°C in 10% formamide/0.25 M Na_2HPO_4/7% SDS

and the most stringent step was with 20 mM Na_2HPO_4/1% SDS at 42°C. DNA fragments used as probes were labelled with [^{32}P]dCTP using the 'Radprime' system (Gibco-BRL) or with [digoxigenin-11]dUTP and detected using the Digoxigenin system (Boehringer Mannheim) according to the manufacturers' instructions, except that hybridisation was performed in the same solution as that used for radioactively labelled probes. Genomic DNA was prepared for Southern hybridisation as described previously (13).

Molecular Cloning and Nucleotide Sequencing. Restriction endonucleases and T4 DNA ligase were obtained from Gibco-BRL. General molecular biology techniques were as described in Sambrook *et al*, (18). Nucleotide sequencing was performed by the dideoxy chain termination method (19) using a T7 sequencing kit (Pharmacia). Sequences were assembled with the MacVector 3.5 software or AssemblyLIGN packages. Nucleotide and derived amino acid sequences were compared with those in the European Molecular Biology Laboratory (EMBL) and SwissProt databases using the BLAST and FASTA programs of the GCG package from the University of Wisconsin (version 8) (20) on the network service at the Human Genome Mapping Project Resource Centre, Hinxton, UK.

Virulence Tests. Groups of five female BALB/c mice (20-25g) were inoculated orally (p.o.) or intraperitoneally (i.p.) with 10-fold dilutions of bacteria suspended in physiological saline. For preparation of the inoculum, bacteria were grown overnight at 37°C in LB broth with shaking (50 rpm) and then used to inoculate fresh medium for various lengths of time until an optical density (OD) at 560 nm of 0.4 to 0.6 had been reached. For cell densities of 5 x 10^8 colony forming units (cfu) per ml and above, cultures were concentrated by centrifugation and resuspended in saline. The concentration of cfu/ml was checked by plating a dilution series of the inoculum onto LB agar plates. Mice were inoculated i.p. with 0.2 ml

volumes and p.o. by gavage with the same volume of inoculum. The LD_{50} values were calculated after 28 days by the method of Reed and Meunch (21).

Results

Localisation of Transposon Insertions. The generation of a bank of *Salmonella typhimurium* mini Tn*5* transposon mutants and the screen used to identify 43 mutants with attenuated virulence have been described previously (5). Transposons and flanking DNA regions were cloned from exconjugants by selection for kanamycin resistance or by inverse PCR. Nucleotide sequences of 300-600 bp of DNA flanking the transposons were obtained for 33 mutants. Comparison of these sequences with those in the DNA and protein databases indicated that 14 mutants resulted from transposon insertions into previously known virulence genes, 7 arose from insertions into new genes with similarity to known genes of the enterobacteria and 12 resulted from insertions into sequences without similarity to entries in the DNA and protein databases (ref. 5, Example 1 and this Example).

Three lines of evidence suggested that 16 of 19 transposon insertions into new sequences were clustered in three regions of the genome, initially designated A, B and C. First, comparing nucleotide sequences from regions flanking transposon insertion points with each other and with those in the databases showed that some sequences overlapped with one another or had strong similarity to different regions of the same gene. Second, Southern analysis of genomic DNA digested with several restriction enzymes and probed with restriction fragments flanking transposon insertion points indicated that some transposon insertions were located on the same restriction fragments. Third, when the same DNA probes were hybridised

to plaques from a *S. typhimurium* λ DNA library, the probes from mutants which the previous two steps had suggested might be linked were found to hybridise to the same λ DNA clones. Thus two mutants (P9B7 and P12F5) were assigned to cluster A, five mutants (P2D6, P9B6, P11C3, P11D10 and P11H10) to cluster B and nine mutants (P3F4, P4F8, P7A3, P7B8, P7G2, P8G12, P9G4, P10E11 and P11B9) to cluster C (Figure 8).

Hybridisation of DNA probes from these three clusters to lysates from a set of *S. typhimurium* strains harbouring locked-in Mu*d*-P22 prophages (15, 16) showed that the three loci were all located in the minute 30 to 31 region (edition VIII, ref. 22) (Figure 7), indicating that the three loci were closely linked or constituted one large virulence locus. To determine if any of the λ clones covering clusters A, B and C contained overlapping DNA inserts, DNA fragments from the terminal regions of each clone were used as probes in Southern hybridisation analysis of the other λ clones. Hybridising DNA fragments showed that several λ clones overlap and that clusters A, B and C comprise one contiguous region (Figure 8). DNA fragments from the ends of this region were then used to probe the λ library to identify further clones containing inserts representing the adjacent regions. No λ clones were identified that covered the extreme right hand terminus of the locus so this region was obtained by cloning a 6.5 kb *Eco*RI/*Xba*I fragment from a lysate of the Mu*d*-P22 prophage strain TT15244 (16).

Restriction mapping and Southern hybridisation analysis were then used to construct a physical map of this locus (Figure 8). To distinguish this locus from the well characterised *inv/spa* gene cluster at minute 63 (edition VIII, ref. 22) (8, 9, 23, 24, 25, 26), we refer to the latter as virulence gene cluster 1 (VGC1) and have termed the new virulence locus VGC2. Figure 2 shows the position of two portions of DNA whose nucleotide sequence has been determined ("Sequence 1" and "Sequence 2"). The nucleotide

sequence is shown in Figures 11 and 12.

Mapping the boundaries of VGC2 on the *S. typhimurium* chromosome. Nucleotide sequencing of λ clone 7 at the left hand side of VGC2 revealed the presence of an open reading frame (ORF) whose deduced amino acid sequence is over 90% identical to the derived product of a segment of the *ydhE*[+] gene of *E. coli* and sequencing of the 6.5 kb *Eco*RI/*Xba*I cloned fragment on the right hand side of VGC2 revealed the presence of an ORF whose predicted amino acid sequence is over 90% identical to pyruvate kinase I of *E. coli* encoded by the *pykF* gene (27). On the *E. coli* chromosome *ydhE* and *pykF* are located close to one another, at minute 37 to 38 (28). Eleven non-overlapping DNA fragments distributed along the length of VGC2 were used as probes in non-stringent Southern hybridisation analysis of *E. coli* and *S. typhimurium* genomic DNA. Hybridising DNA fragments showed that a region of approximately 40 kb comprising VGC2 was absent from the *E. coli* genome and localised the boundaries of VGC2 to within 1 kb (Figure 9). Comparison of the location of the *Xba*I site close to the right hand end of VGC2 (Figure 8) with a map of known *Xba*I sites (29) at the minute 30 region of the chromosome (22) enables a map position of 30.7 minutes to be deduced for VGC2.

Structure of VGC2. Nucleotide sequencing of portions of VGC2 has revealed the presence of 19 ORFs (Figure 8). The G+C content of approximately 26 kb of nucleotide sequence within VGC2 is 44.6%, compared to 47% for VGC1 (9) and 51-53% estimated for the entire *Salmonella* genome (30).

The complete deduced amino acid sequences of ORFs 1-11 are similar to those of proteins of type III secretion systems (6, 7), which are known to be required for the export of virulence determinants in a variety of bacterial

pathogens of plants and animals (7). The predicted proteins of ORFs 1 - 8 (Figure 8) are similar in organisation and sequence to the products of the *yscN-U* genes of *Yersinia pseudotuberculosis* (31), to *invC/spaS* of the *inv/spa* cluster in VGC1 of *Salmonella typhimurium* (8, 9) and to *spa47/spa40* of the *spa/mxi* cluster of *Shigella flexneri* (32, 33, 34, 35,). For example the predicted amino acid sequence of ORF 3 (Figure 8) is 50% identical to YscS of *Y. pseudotuberculosis* (31), 34% identical to Spa9 from *S. flexneri* (35) and 37% identical to SpaQ of VGC1 of *S. typhimurium* (9). The predicted protein product of ORF9 is closely related to the LcrD family of proteins with 43% identity to LcrD of *Y. enterocolitica* (36), 39% identity to MxiA of *S. flexneri* (32) and 40% identity to InvA of VGC1 (23). Partial nucleotide sequences for the remaining ORFs shown in Figure 8 indicate that the predicted protein from ORF10 is most similar to *Y. enterocolitica* YscJ (37) a lipoprotein located in the bacterial outer membrane, with ORF11 similar to *S. typhimurium* InvG, a member of the PulD family of translocases (38). ORF12 and ORF13 show significant similarity to the sensor and regulatory subunits respectively, from a variety of proteins comprising two component regulatory systems (39). There is ample coding capacity for further genes between ORFs 9 and 10, ORFs 10 and 11, and between ORF 19 and the right hand end of VGC2.

VGC2 is conserved among and is specific to the *Salmonellae*. A 2.2 kb *Pst*I/*Hin*dIII fragment located at the centre of VGC2 (probe B, Figure 8) lacking sequence similarity to entries in the DNA and protein databases was used as a probe in Southern hybridisation analysis of genomic DNA from *Salmonella* serovars and other pathogenic bacteria (Figure 10A). DNA fragments hybridising under non-stringent conditions showed that VGC2 is present in *S. aberdeen, S. gallinarum, S. cubana, S. typhi* and is absent from EPEC, EHEC, *Y. pestis, S. flexneri, V. cholera* and *S. aureus*. Thus VGC2 is conserved among and is likely to be specific to the *Salmonellae*.

To determine if the organisation of the locus is conserved among the *Salmonella* serovars tested, stringent Southern hybridisations with genomic DNA digested with two further restriction enzymes were carried out. Hybridising DNA fragments showed that there is some heterogeneity in the arrangement of restriction sites between *S. typhimurium* LT2 and *S. gallinarum, S. cubana* and *S. typhi* (Figure 10B). Furthermore, *S. gallinarum* and *S. typhi* contain additional hybridising fragments to those present in the other *Salmonellae* examined, suggesting that regions of VGC2 have been duplicated in these species.

VGC2 is required for virulence in mice. Previous experiments showed that the LD_{50} values for i.p. inoculation of transposon mutants P3F4, P7G2, P9B7 and P11C3 were at least 100-fold greater than the wild type strain (5). In order to clarify the importance of VGC2 in the process of infection, the p.o. and i.p. LD_{50} values for mutants P3F4 and P9B7 were determined (Table 1). Both mutants showed a reduction in virulence of at least five orders of magnitude by either route of inoculation in comparison with the parental strain. This profound attenuation of virulence by both routes of inoculation demonstrates that VGC2 is required for events in the infective process after epithelial cell penetration in BALB/c mice.

Table 1. LD_{50} values of *S. typhimurium* strains.

	LD_{50} (cfu)	
Strain	**i.p.**	**p.o.**
12023 wild type	4.2	6.2×10^4
P3F4	1.5×10^6	$>5 \times 10^9$
P9B7	$>1.5 \times 10^6$	$>5 \times 10^9$

cfu. colony forming units

Discussion

A hitherto unknown virulence locus in *S. typhimurium* of approximately 40 kb located at minute 30.7 on the chromosome by mapping the insertion points of a group of signature-tagged transposon mutants with attenuated virulence has been identified (5). This locus is referred to as virulence gene cluster 2 (VGC2) to distinguish it from the *inv/spa* virulence genes at 63 minutes (edition VIII, ref. 22) which we suggest be renamed VGC1. VGC1 and VGC2 both encode components of type III secretion systems. However, these secretion systems are functionally distinct.

Of 19 mutants that arose from insertions into new genes (ref. 5 and this example) 16 mapped to the same region of the chromosome. It is possible that mini-Tn*5* insertion occurs preferentially in VGC2. Alternatively, as the negative selection used to identify mutants with attenuated virulence (5) was very stringent (reflected by the high LD_{50} values for VGC2 mutants) it is possible that, among the previously unknown genes, only mutations in those of VGC2 result in a degree of attenuation sufficient to be recovered in the screen. The failure of previous searches for *S. typhimurium* virulence determinants to identify VGC2 might stem from reliance on cell culture assays rather than a live animal model of infection. A previous study which identified regions of the *S. typhimurium* LT2 chromosome unique to *Salmonellae* (40) located one such region (RF333) to minutes 30.5 - 32. Therefore, RF333 may correspond to VGC2, although it was not known that RF333 was involved in virulence determination.

Comparisons with the type III secretion systems encoded by the virulence plasmids of *Yersinia* and *Shigella* as well as with VGC1 of *Salmonella* indicates that VGC2 encodes the basic structural components of the secretory apparatus. Furthermore, the order of ORFs 1-8 in VGC2 is the

same as the gene order in homologues in *Yersinia, Shigella* and VGC1 of *S. typhimurium*. The fact that the organisation and structure of the VGC2 secretion system is no more closely related to VGC1 than to the corresponding genes of *Yersinia*, together with the low G+C content of VGC2 suggests that VGC2, like VGC1 (40, 41, 42) was acquired independently by *S. typhimurium via* horizontal transmission. The proteins encoded by ORFs 12 and 13 show strong similarity to bacterial two component regulators (39) and could regulate either ORFs 1-11 and/or the secreted proteins of this system.

Many genes in VGC1 have been shown to be important for entry of *S. typhimurium* into epithelial cells. This process requires bacterial contact (2) and results in cytoskeletal rearrangements leading to localised membrane ruffling (43, 44). The role of VGC1 and its restriction to this stage of the infection is reflected in the approximately 50-fold attenuation of virulence in BALB/c mice inoculated p.o. with VGC1 mutants and by the fact that VGC1 mutants show no loss of virulence when administered i.p. (8). The second observation also explains why no VGC1 mutants were obtained in our screen (5). In contrast, mutants in VGC2 are profoundly attenuated following both p.o. and i.p. inoculation. This shows that, unlike VGC1, VGC2 is required for virulence in mice after epithelial cell penetration, but these findings do not exclude a role for VGC1 in this early stage of infection.

Thus in summary mapping the insertion points of 16 signature-tagged transposon mutants on the *Salmonella typhimurium* chromosome led to the identification of a 40 kb virulence gene cluster at minute 30.7. This locus is conserved among all other *Salmonella* species examined, but not present in a variety of other pathogenic bacteria or in *Escherichia coli* K12. Nucleotide sequencing of a portion of this locus revealed 11 open reading frames whose predicted proteins encode components of a type III secretion

system. To distinguish between this and the type III secretion system encoded by the *inv/spa* invasion locus we refer to the *inv/spa* locus as virulence gene cluster 1 (VGC1) and the new locus as VGC2. VGC2 has a lower G+C content than that of the *Salmonella* genome and is flanked by genes whose products share greater than 90% identity with those of the *E. coli ydhE* and *pykF* genes. Thus VGC2 was probably acquired horizontally by insertion into a region corresponding to that between the *ydhE* and *pykF* genes of *E. coli*. Virulence studies of VGC2 mutants have shown them to be attenuated by at least five orders of magnitude compared with the wild type strain following oral or intraperitoneal inoculation.

References for this Example

1. Carter, P.B. & Collins, F.M. (1974) *J. Exp. Med.* **139**, 1189-1203.
2. Takeuchi, A. (1967) *Am. J. Pathol.* **50**, 109-136.
3. Finlay, B.B. (1994) *Curr. Top. Microbiol. Immunol.* **192**, 163-185.
4. Groisman, E.A. & Ochman, H. (1994) *Trends Microbiol.* **2**, 289-293.
5. Hensel, M., Shea, J.E., Gleeson, C., Jones, M.D., Dalton, E. & Holden, D.W. (1995) *Science* **269**, 400-403.
6. Salmond, G.P.C. & Reeves, P.J. (1993) *Trends Biochem. Sci.* **18**, 7-12.
7. Van Gijsegem, F., Genin., S. & Boucher, C. (1993) *Trends Microbiol.* **1**, 175-180.
8. Galan, J.E. & Curtiss, R. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6383-6387.
9. Groisman, E.A. & Ochman, H. (1993) *EMBO J.* **12**, 3779-3787.
10. Ginocchio, C.C., Olmsted, S.B., Wells, C.L. & Galan, J.E. (1994) *Cell* **76**, 717-724.
11. de Lorenzo, V. & Timmis, K.N. (1994) *Methods Enzymol.* **264**,

386-405.

12. Davis, R.H., Botstein, D. & Roth, J.R. (1980) *Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

13. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) *Current Protocols in Molecular Biology* Vol 4 John Wiley and Sons, Inc, New York.

14. Maurer, R., Osmond, B.C., Shekhtman, E., Wong, A. & Botstein, D. (1984) *Genetics* **108**, 1-23.

15. Youderain, P., Sugiono, P., Brewer, K.L., Higgins, N.P. & Elliott, T. (1988) *Genetics* **118**, 581-592.

16. Benson, N.R. & Goldman, B.S. (1992) *J. Bacteriol.* **174**, 1673-1681.

17. Holden, D.W., Kronstad, J.W. & Leong, S. (1989) *EMBO J.* **8**, 1927-1934.

18. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular cloning: a laboratory manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

19. Sanger, F., Nicklen, S. & Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467.

20. Devereux, J., Hearberli, P. & Smithies, O. (1984) *Nucl. Acids Res.* **12**, 387-399.

21. Reed, L.J. & Muench, H. (1938) *Am. J. Hyg.* **27**, 493-497.

22. Sanderson, K.E., Hessel, A. & Rudd, K.E. (1995) *Microbiol. Rev.* **59**, 241-303.

23. Galan, J.E., Ginocchio, C. & Costeas, P. (1992) *J. Bacteriol.* **174**, 4338-4349.

24. Ginocchio, C., Pace, J. & Galan, J.E. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5976-5980.

25. Eichelberg, K., Ginocchio, C.C. & Galan, J.E. (1994) *J. Bacteriol.* **176**, 4501-4510.

26. Collazo, C.M., Zierler, M.K. & Galan, J.E. (1995) *Mol. Microbiol.* **15**, 25-38.

27. Ohara, O., Dorit, R.L. & Gilbert, W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6883-6887.

28. Bachman, B. (1990) *Micro. Rev.* **54**, 130-197.

29. Liu, S.L., Hessel, A. & Sanderson, K.E. (1993) *J. Bacteriol.* **175**, 4104-4120.

30. Fasman, G.D. (1976) *CRC Handbook of Biochemistry and Molecular Biology*, CRC Press, Cleveland.

31. Bergman, T., Erickson, K., Galyov, E., Persson, C. & Wolf Watz, H. (1994) *J. Bacteriol.* **176**, 2619-2626.

32. Andrews, G.P. & Maurelli, A.T. (1992) *Infect. Immun.* **60**, 3287-3295.

33. Allaoui, A., Sansonetti, P.J. & Parsot, C. (1993) *Mol. Microbiol.* **7**, 59-68.

34. Venkatesan, M., Buysse, J.M. & Oaks, E.V. (1992) *J. Bacteriol.* **174**, 1990-2001.

35. Sasakawa, C., Komatsu, K., Tobe, T., Suzuki, T. & Yoshikawa, M. (1993) *J. Bacteriol.* **175**, 2334-2346.

36. Plano, G.V., Barve, S.S. & Straley, S.C. (1991) *J. Bacteriol.* **173**, 7293-7303.

37. Michiels, T., Vanooteghem, J.C., Lambert de Rouvroit, C., China, B., Gustin, A., Boudry, P. & Cornelis, G.R. (1991) *J. Bacteriol.* **173**, 4994-5009.

38. Kaniga, K., Bossio, J.C. & Galan, J.E. (1994) *Mol. Microbiol.* **13**, 555-568.

39. Ronson, C.W., Nixon, B.T. & Ausubel, F.M. (1987) *Cell* **49**, 579-581.

40. Groisman, E.A., Sturmoski, M.A., Solomon, F.R., Lin, R. & Ochman, H. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1033-1037.

41. Altmeyer, R.M., McNern, J.K., Bossio, J.C., Rosenshine, I., Finlay, B.B. & Galan, J.E. (1993) *Mol. Microbiol.* **7**, 89-98.

42. Li, J., Ochman, H., Groisman, E.A., Boyd, E.F., Soloman, F., Nelson, K. & Selander, R.K. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7252-7256.

43. Finlay, B.B. & Rauschkowski, S. (1991) *J. Cell Sci.* **99**, 283-296.

44. Francis, C.L., Starnbach, M.N. & Falkow, S. (1992) *Mol. Microbiol.* **6**, 3077-3087.

Example 5: Identification of virulence genes in *Streptococcus pneumoniae*

(a) Mutagenesis

In the absence of a convenient transposon system, the most efficient way of creating tagged mutants of *Streptococcus pneumoniae* is to use insertion-duplication mutagenesis (Morrison *et al* (1984) *J. Bacteriol.* **159**, 870). Random *S. pneumoniae* DNA fragments of 200-400 bp will be generated by genomic DNA digestion with a restriction enzyme or by physical shearing by sonication followed by gel fractionation and DNA end-repair using T4 DNA polymerase. The fragments are ligated into plasmid pJDC9 (Pearce *et al* (1993) *Mol. Microbiol.* **9**, 1037 which carries the *erm* gene for erythromycin selection in *E. coli* and *S. pneumoniae*), previously modified by incorporation of DNA sequence tags into one of the polylinker cloning sites. The size of cloned *S. pneumoniae* DNA is sufficient to ensure homologous recombination, and reduces the possibility of generating an unrepresentative library in *E. coli* (expression of *S. pneumoniae* proteins can be toxic to *E. coli*). Alternative vectors carrying different selectable markers are available and can be used in place of pJDC9. Tagged plasmids carrying DNA fragments are introduced to an

appropriate *S. pneumoniae* strain selected on the basis of serotype and virulence in a murine model of pneumococcal pneumonia. Regulation of competence for genetic transformation in *S. pneumoniae* is governed by competence factor, a peptide of 17 amino acids which has been characterized recently by Don Morrison's group at the University of Illinois at Chicago and which is described Havarstein, Coomaraswamy and Morrison (1995) *Proc. Natl. Acad. Sci. USA* **92**, 11140-11144. Incorporation of minute quantities of this peptide in transformation experiments leads to very efficient transformation frequencies in some encapsulated clinical isolates of *S. pneumoniae*. This overcomes a major hurdle in pneumococcal molecular genetics and the availability of the peptide greatly facilitates the construction of *S. pneumoniae* mutant banks and allows flexibility in choosing the strain(s) to be mutated. A proportion of transformants are analysed to verify homologous integration of the plasmid sequences, and checked for stability. The very low level of reversion associated with mutants generated by insertion-duplication is minimized by the fact that the duplicated regions will be short (200-400 bp); however if the level of reversion is unacceptably high, antibiotic selection is maintained during growth of the transformants in culture and during growth in the animal.

(b) Animal model

The *S. pneumoniae* mutant bank is organized into pools for inoculation into Swiss and/or C57B1/6 mice. Preliminary experiments are conducted to determine the optimum complexity of the pools and the optimum inoculum level. One attractive model utilises inocula of 10^5 cfu, delivered by mouth to the trachea (Veber *et al* (1993) *J. Antimicrobial Chemotherapy* **32**, 473). Swiss mice develop acute pneumonia within 3-4 days, and C57B1/6 mice develop subacute pneumonia within 8-10 days. These pulmonary models of infection yield 10^8 cfu/lung (Veber *et al* (1993) *J. Antimicrobial*

Chemotherapy 32, 473) at the time of death. If required, mice are also injected intraperitoneally for the identification of genes required for bloodstream infection (Sullivan *et al* (1993) *Antimicrobial Agents and Chemotherapy* 37, 234).

(c) Virulence gene identification

Once the parameters of the infection model are optimized, a mutant bank consisting of several thousand strains is subjected to virulence tests. Mutants with attenuated virulence are identified by hybridisation analysis, using labelled tags from the 'input' and 'recovered' pools as probes. If *S. pneumoniae* DNA cannot be colony blotted easily, chromosomal DNA is liberated chemically or enzymatically in the wells of microtitre dishes prior to transfer onto nylon membranes using a dot-blot apparatus. DNA flanking the integrated plasmid is cloned by plasmid rescue in *E. coli* (Morrison *et al* (1984) *J. Bacteriol.* **159**, 870), and sequenced. Genomic DNA libraries are constructed in appropriate vectors maintained in either *E. coli* or a Gram-positive host strain, and are probed with restriction fragments flanking the integrated plasmid to isolate cloned virulence genes which is then fully sequenced and subjected to detailed functional analysis.

Example 6: Identification of virulence genes in *Enterococcus faecalis*

(a) Mutagenesis

Mutagenesis of *E. faecalis* is accomplished using plasmid pAT112 or a derivative, developed for this purpose. pAT112 carries genes for selection in both Gram-negative and Gram-positive bacteria, and the *att* site of Tn*1545*. It therefore requires the presence in the host strain of the integrase for transposition, and stable, single copy insertions are obtained if the host does not contain an excisionase gene (Trieu-Cuot *et al* (1991) *Gene* **106**, 21). Recovery of DNA flanking the integrated plasmid is accomplished by

restriction digestion of genomic DNA, intramolecular ligation and transformation of *E. coli*. The presence of single sites for restriction enzymes in pAT112 and its derivatives will (Trieu-Cuot *et al* (1991) *Gene* **106**, 21) allows the incorporation of DNA sequence tags prior to transfer to a virulent strain of *E. faecalis* carrying plasmid pAT145 (to provide the integrase function) by either conjugation, electroporation or transformation (Trieu-Cuot *et al* (1991) *Gene* **106**, 21; Wirth *et al* (1986) *J. Bacteriol.* **165**, 831).

(b) Animal model

A large number of insertion mutants are analysed for random integration of the plasmid by isolating DNA from transcipients, restriction enzyme digestion and Southern hybridisation. Individual mutants are stored in the wells of microtitre dishes, and complexity and size of pooled inocula are optimised prior to screening of the mutant bank. Two different models of infection caused by *E. faecalis* are employed. The first is a well established rat model of endocarditis, involving tail vein injection of up to 10^8 cfu of *E. faecalis* into animals that have a catheter inserted across the aortic valve (Whitman *et al* (1993) *Antimicrobial Agents and Chemotherapy* **37**, 1069). Animals are sacrificed at various times after inoculation, and bacterial vegetations on the aortic valve are excised, homogenized and plated to culture medium to recover bacterial colonies. Virulent bacteria are also recovered from the blood at various times after inoculation. The second model is of peritonitis in mice, following intraperitoneal injection of up to 10^9 cfu of *E. faecalis* (Chenoweth *et al* (1990) *Antimicrobial Agents and Chemotherapy* **34**, 1800). As with the *S. pneumoniae* model, preliminary experiments are done to establish the optimum complexity of the pools and the optimum inoculum level, prior to screening the mutant bank.

(c) Virulence gene identification

Isolation of DNA flanking the site of integration of pAT112 using its *E. coli* origin of replication is simplified by the lack of sites for most of the commonly used 6 bp recognition restriction enzymes in the vector. Therefore DNA from the strains of interest are digested with one of these enzymes, self-ligated, transformed into *E. coli* and sequenced using primers based on the sequences adjacent to the *att* sites on the plasmid. A genomic DNA library of *E. faecalis* are probed with sequences of interest to identify intact copies of virulence genes which are then sequenced.

Example 7: Identification of virulence genes in Pseudomonas aeruginosa

(a) Mutagenesis

Since transposon Tn*5* has been used by others to mutagenise *Pseudomonas aeruginosa*, and the mini-Tn*5* derivative that was used for the identification of *Salmonella typhimurium* virulence genes (Example 1) is reported to have broad utilisation among Gram-negative bacteria, including several pseudomonads (DeLorenzo and Timaris (1994) *Methods Enzymol.* **264**, 386), a *P. aeruginosa* mutant bank is constructed using our existing pool of signature tagged mini-Tn*5* transposons by conjugal transfer of the suicide vector to one or more virulent (and possibly mucoid) recipient strains. This approach represents a significant time saving. Other derivatives of Tn*5* designed specifically for *P. aeruginosa* mutagenesis (Rella *et al* (1985) *Gene* **33**, 293), may alternatively be employed with the mini Tn*5* transposon.

(b) Animal model and virulence gene identification

The bank of *P. aeruginosa* insertion mutants is screened for attenuated virulence in a chronic pulmonary infection model in rats. Suspensions of *P. aeruginosa* cells are introduced into a bronchus following tracheotomy.

and disease develops over a 30 day period (Woods *et al* (1982) *Infect. Immun.* **36**, 1223). Bacteria are recovered by plating lung homogenates to laboratory medium and sequence tags from these are used to probe DNA colony blots of bacteria used as the inoculum. It is also possible to subject the mutant bank to virulence tests in a model of endogenous bacteremia (Hirakata *et al* (1992) *Antimicrobial Agents and Chemotherapy* **36**, 1198), and cystic fibrosis (Davidson *et al* (1995) *Nature Genetics* **9**, 351) in mice. Cloning and sequencing of DNA flanking the transposons is done as described in Example 1. Genomic DNA libraries for the isolation and sequencing of intact copies of the genes are constructed in the laboratory by standard methods.

Example 8: Identification of virulence genes in *Aspergillus fumigatus*

(a) Mutagenesis

The functional eqiuvalent of transposon mutagenesis in fungi is restriction enzyme mediated integration (REMI) of transforming DNA (Schiestl and Petes (1991) *Proc. Natl. Acad. Sci.* **88**, 7585). In this process, fungal cells are transformed with DNA fragments carrying a selectable marker in the presence of a restriction enzyme, and single copy integrations occur at different genomic sites, defined by the target sequence of the restriction enzyme. REMI has already been used successfully to isolate virulence genes of Cochliobolus (Lu *et al* (1994) *Proc. Natl. Acad. Sci. USA* **91**, 12649) and Ustilago (Bolker *et al* (1995) *Mol. Gen. Genet.* **248**, 547), and have shown that incorporation of active restriction enzyme with a plasmid encoding hygromycin resistance leads to single and apparently random integration of the linear plasmid into the *A. fumigatus* genome. Sequence tags are introduced into a convenient site in one of two vectors for hygromycin resistance, and used to transform a clinical isolate of *A. fumigatus*.

(b) Animal model and virulence gene identification

The low-dose model of aspergillosis in neutropenic mice in particular closely matches the course of pulmonary disease in humans (Smith *et al* (1994) *Infect. Immun.* **62**, 5247). Mice are inoculated intranasally with up to 1,000,000 conidiospores/mouse, and virulent fungal mutants are recovered 7-10 days later by using lung homogenates to inoculate liquid medium. Hyphae are collected after a few hours, from which DNA is extracted for amplification and labelling of tags to probe colony blots of DNA from the pool of transformants comprising the inoculum. DNA from the regions flanking the REMI insertion points are cloned by digesting the transformant DNA with a restriction enzyme that cuts outside the REMI vector, self ligation and transformation of *E. coli*. Primers based on the known sequence of the plasmid are used to determine the adjacent *A. fumigatus* DNA sequences. To prove that the insertion of the vector was the cause of the avirulent phenotype, the recovered plasmid is recut with the same restriction enzyme used for cloning, and transformed back into the wild-type *A. fumigatus* parent strain. Transformants that have arisen by homologous recombination are then subjected to virulence tests.

REFERENCES (other than for Example 4)

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) *Current Protocols in Molecular Biology*, New York: John Wiley and Sons.

Buchmeier, N.A., Lipps, C.J., So, M.Y. and Heffron, F. (1993) Recombination-deficient mutants of *Salmonella typhimurium* are avirulent and sensitive to the oxidative burst of macrophages. *Mol. Microbiol.* **7**, 933-936.

Carter, P.B. and Collins, F.M. (1974) The route of enteric infection in normal mice. *J. Exp. Med.* **139**, 1189-1203.

de Lorenzo, V. and Timmis, K.N. (1994) Analysis and construction of stable phenotypes in Gram-negative bacteria with Tn*5*-and Tn*10*-derived minitransposons. *Methods Enzymol.* **264**, 386-405.

de Lorenzo, V., Herrero, M., Jakubzik, U. and Timmis, K.N. (1990) Mini-Tn*5* transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria. *J. Bacteriol.* **172**, 6568-6572.

Fields, P.I., Groisman, E.A. and Heffron, F. (1989) A *Salmonella* locus that controls resistance to microbicidal proteins from phagocytic cells. *Science* **243**, 1059-1062.

Finlay, B.B., Starnbach, M.N., Francis, C.L., Stocker, B.A., Chatfield, S., Dougan, G. and Falkow, S. (1988) Identification and characterization of Tn*phoA* mutants of *Salmonella* that are unable to pass through a

polarized MDCK epithelial cell monolayer. *Mol. Microbiol.* **2**, 757-766.

Groisman, E.A., Chiao, E., Lipps, C.J., Heffron, F. (1989) *Salmonella typhimurium phoP* virulence gene is a transcriptional regulator. *Proc. Natl. Acad. Sci. USA.* **86**, 7077-7081.

Groisman, E.A. and Ochman, H. (1994) How to become a pathogen. *Trends Microbiol.* **2**, 289-293.

Groisman, E.A. and Saier, M.H., Jr. (1990) *Salmonella* virulence: new clues to intramacrophage survival. *Trends Biochem. Sci.* **15**, 30-33.

Herrero, M., de Lorenzo, V. and Timmis, K.N. (1990) Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in Gram-negative bacteria. *J. Bacteriol.* **172**, 6557-6567.

Holden D.W., Kronstad J.W., Leong S.A. (1989) Mutation in a heat-regulated hsp70 gene of *Ustilago maydis*. *EMBO J.* **8**, 1927-1934.

Holland J., Towner K.J., Williams P. (1992) Tn*916* insertion mutagenesis in *Escherichia coli* and *Haemophilus influenzae* type b following conjugative transfer. *J. Gen. Microbiol.* **138**, 509-515.

Mahan, M.J., Slauch, J.M., Mekalanos, J.J. (1993) Selection of bacterial virulence genes that are specifically induced in host tissues. *Science* **259**, 686-688.

Miller, S.I., Kukral, A.M. and Mekalanos. J.J. (1989a) A two-component regulatory system (*phoP phoQ*) controls *Salmonella typhimurium* virulence. *Proc. Natl. Acad. Sci. USA.* **86**, 5054-5058.

Miller, I., Maskell, D., Hormaeche, C., Johnson, K., Pickard, D. and Dougan, G. (1989b) Isolation of orally attenuated *Salmonella typhimurium* following Tn*phoA* mutagenesis. *Infect. Immun.* **57**, 2758-2763.

Miller, V.L. and Mekalanos, J.J. (1988) A novel suicide vector and its use in construction of invertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. *J. Bacteriol.* **170**, 2575-2583.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**, 5463-5467.

Schiestl R.H., and Petes T.D. (1991) Integration of DNA fragments by illegitimate recombination in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci USA.* **88**, 7585-7589.

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced



CLAIMS

1. A method for identifying a microorganism having a reduced adaptation to a particular environment comprising the steps of:

(1) providing a plurality of microorganisms each of which is independently mutated by the insertional inactivation of a gene with a nucleic acid comprising a unique marker sequence so that each mutant contains a different marker sequence, or clones of the said microorganism;

(2) providing individually a stored sample of each mutant produced by step (1) and providing individually stored nucleic acid comprising the unique marker sequence from each individual mutant;

(3) introducing a plurality of mutants produced by step (1) into the said particular environment and allowing those microorganisms which are able to do so to grow in the said environment;

(4) retrieving microorganisms from the said environment or a selected part thereof and isolating the nucleic acid from the retrieved microorganisms;

(5) comparing any marker sequences in the nucleic acid isolated in step (4) to the unique marker sequence of each individual mutant stored as in step (2); and

(6) selecting an individual mutant which does not contain any of the marker sequences as isolated in step (4).

2. A method according to Claim 1 wherein the plurality of microorganisms as defined in step (1) is produced from a plurality of microorganisms, each of which comprises a nucleic acid comprising a unique marker sequence, by changing their condition from a first given condition to a second given condition wherein (a) in the first given condition the said nucleic acid comprising a unique marker is maintained episomally and (b) in the second given condition the said nucleic acid comprising a



THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

unique marker sequence insertionally inactivates a gene.

3. A method according to Claims 1 or 2 further comprising the steps:

(1A) removing auxotrophs from the plurality of mutants produced in step (1); or

(6A) determining whether the mutant selected in step (6) is an auxotroph; or

both (1A) and (6A).

4. A method of identifying a gene which allows a microorganism to adapt to a particular environment, the method comprising the method of any one of Claims 1 to 3 followed by the step:

(7) isolating the insertionally-inactivated gene from the individual mutant selected in step (6).

5. A method according to Claim 4 further comprising the step:

(8) isolating from a wild-type microorganism the corresponding wild-type gene using the insertionally-inactivated gene isolated in step (7) as a probe.

6. A method according to any one of Claims 1 to 5 wherein the particular environment is a differentiated multicellular organism.

7. A method according to Claim 6 wherein the multicellular organism is a plant.

8. A method according to Claim 6 wherein the multicellular organism is a non-human animal.

9. A method according to Claim 8 wherein the animal is a mouse, rat,

* C L M S *

This Sheet Has Been Replaced

rabbit, dog or monkey.

10. A method according to Claim 9 wherein the animal is a mouse.

11. A method according to any one of Claims 6 to 10 wherein in step (4) the microorganisms are retrieved from the said environment at a site remote from the site of introduction in step (3).

12. A method according to any one of Claims 8 to 10 wherein in step (3) the microorganism is introduced orally or intraperitoneally.

13. A method according to Claim 12 when dependent on Claims 8 or 9 wherein in step (4) the microorganisms are retrieved from the spleen.

14. A method according to any one of the preceding claims wherein the microorganism is a bacterium.

15. A method according to any one of Claims 1 to 13 wherein the microorganism is a fungus.

16. A method according to Claim 7 wherein the microorganism is a bacterium pathogenic to plants.

17. A method according to Claim 7 wherein the microorganism is a fungus pathogenic to plants.

18. A method according to any one of Claims 8 to 10 wherein the microorganism is a bacterium pathogenic to animals.

19. A method according to any one of Claims 8 to 10 wherein the

* C L M S *

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced



microorganism is a fungus pathogenic to animals.

20. A method according to Claim 18 wherein the bacterium is any one of *Bordetella pertussis*, *Campylobacter jejuni*, *Clostridium botulinum*, *Escherichia coli*, *Haemophilus ducreyi*, *Haemophilus influenzae*, *Helicobacter pylori*, *Klebsiella pneumoniae*, *Legionella pneumophila*, *Listeria* spp., *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Pseudomonas* spp., *Salmonella* spp., *Shigella* spp., *Staphylococcus aureus*, *Streptococcus pyogenes*, *Streptococcus pneumoniae*, *Vibrio* spp., and *Yersinia pestis*.

21. A method according to Claim 19 wherein the fungus is any one of *Aspergillus* spp., *Cryptococcus neoformans* and *Histoplasma capsulatum*.

22. A method according to any one of the preceding claims wherein in step (1) the gene is insertionally inactivated using a transposon or transposon like element or other DNA sequence carrying a unique marker sequence.

23. A method according to any one of the preceding claims wherein in step (1) each different marker sequence is flanked on either side by sequences common to each said nucleic acid.

24. A method according to Claim 23 wherein in step (2) the nucleic acid comprising the unique marker is isolated using DNA amplification techniques and oligonucleotide primers which hybridise to the said common sequences.

25. A method according to Claim 23 or 24 wherein in step (4) the nucleic acid comprising a plurality of said marker sequences is isolated using DNA amplification techniques and oligonucleotide primers which



THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

hybridise to the said common sequences.

26. A microorganism obtained using the method of any one of the preceding claims.

27. A microorganism comprising a mutation in a gene identified using the method of Claim 5.

28. A microorganism obtained according to Claim 26, when dependent on Claim 8, or Claim 27 for use in a vaccine.

29. A vaccine comprising a microorganism according to Claim 26, when dependent on Claim 8, or Claim 27 and a pharmaceutically-acceptable carrier.

30. A gene obtained using the method of Claims 4 or 5.

31. A gene according to Claim 30 which is isolated from the *Salmonella typhimurium* genome and hybridises to the sequence shown in Figure 5 under stringent conditions.

32. A gene according to Claim 30 which is isolated from the *Salmonella typhimurium* genome and hybridises to a sequence shown in Figure 6 under stringent conditions.

33. A polypeptide encoded by a gene according to any one of Claims 30 to 32.

34. A method of identifying a compound which reduces the ability of a microorganism to adapt to a particular environment comprising the step of

* C L M S *

selecting a compound which interferes with the function of a gene according to any one of Claims 30 to 32 or a polypeptide according to Claim 33.

35. A compound identifiable by the method of Claim 34.

36. A compound according to Claim 35 wherein the particular environment is a host organism.

37. A compound according to Claim 36 wherein the host organism is a plant.

38. A compound according to Claim 36 wherein the host organism is an animal.

39. Use of a compound according to any one of Claim 36 to Claim 38 for treating infection of said host organism with said microorganism.

40. A molecule which selectively interacts with, and substantially inhibits the function of, a gene according to any one of Claims 30 to 32 or a nucleic acid product thereof.

41. A molecule according to Claim 40 which is an antisense nucleic acid or nucleic acid derivative.

42. A molecule according to Claim 40 or 41 which is an antisense oligonucleotide.

43. A molecule according to any one of Claims 40 to 42 for use in medicine.

44. A method of treating a host which has, or is susceptible to, an infection with a microorganism, the method comprising administering an effective amount of a molecule or compound according to Claim 36 or 40 wherein said gene is present in said microorganism, or a close relative of said microorganism.

45. A pharmaceutical composition comprising a molecule or compound according to Claim 38 or 40 and a pharmaceutically acceptable carrier.

46. The VGC2 DNA of *Salmonella typhimurium* or a part thereof, or a variant of said DNA or a variant of a part thereof.

47. A mutant bacterium wherein if the bacterium normally contains a gene that is the same as or equivalent to a gene in VGC2, said gene is mutated or absent in said mutant bacterium.

48. A method of making a bacterium according to Claim 47.

49. Use of a mutant bacterium according to Claim 47 in a vaccine.

50. A pharmaceutical composition comprising a bacterium according to Claim 47 and a pharmaceutically acceptable carrier.

51. A polypeptide encoded by VGC2 DNA of *Salmonella typhimurium* or a part thereof, or a variant of said polypeptide or a variant of a part thereof.

52. A method of identifying a compound which reduces the ability of a bacterium to infect or cause disease in a host comprising the step of selecting a compound which interferes with the function of a gene in VGC2

THIS SHEET HAS BEEN REPLACED

according to Claim 46 or a polypeptide according to Claim 51.

53. A compound identifiable by the method of Claim 52.

54. A molecule which selectively interacts with, and substantially inhibits the function of, a gene in VGC2 of *Salmonella typhimurium* or a nucleic product thereof.

55. A molecule or compound according to Claim 53 or 54 for use in medicine.

56. Any novel feature or combination of features disclosed herein.

This Sheet Has Been Replaced





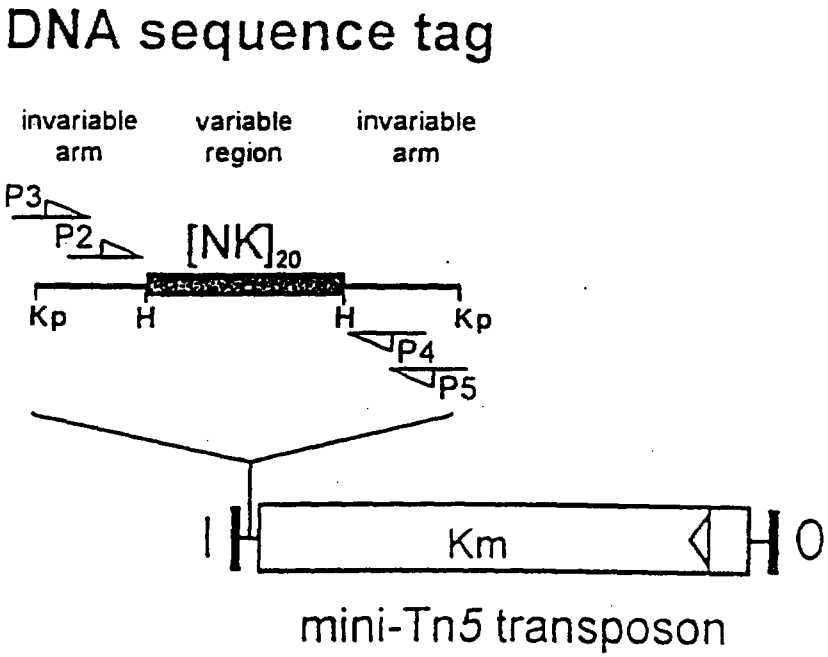

DNA sequence tag
invariable arm
variable region
invariable arm
P3
P2
[NK]20
Kp
H
H
Kp
P4
P5
I
Km
O
mini-Tn5 transposon

Figure 1a

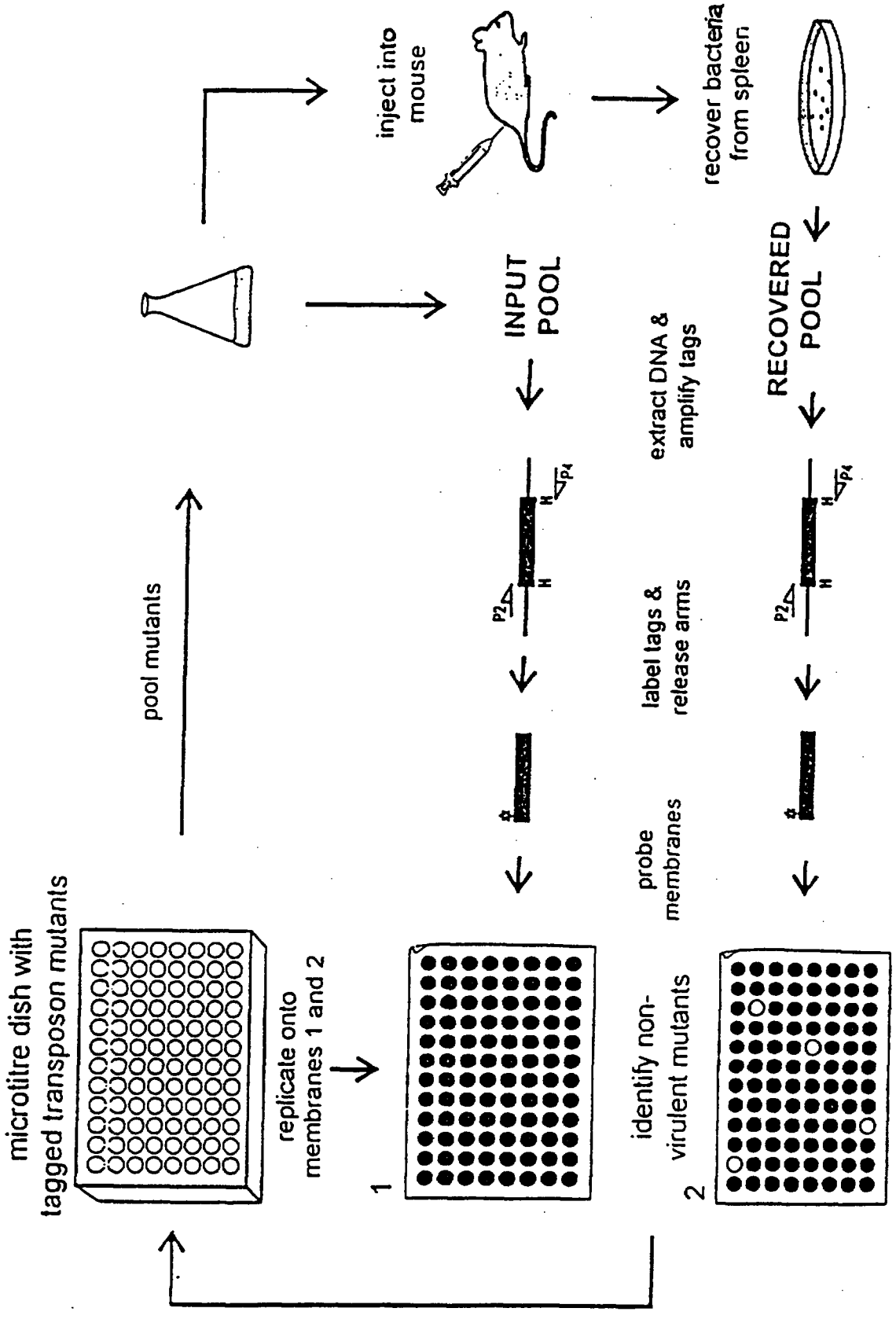

microtitre dish with tagged transposon mutants
replicate onto membranes 1 and 2
1
identify non-virulent mutants
2
probe membranes
label tags & release arms
extract DNA & amplify tags
pool mutants
INPUT POOL
RECOVERED POOL
inject into mouse
recover bacteria from spleen
P1
P2
H

Figure 1b

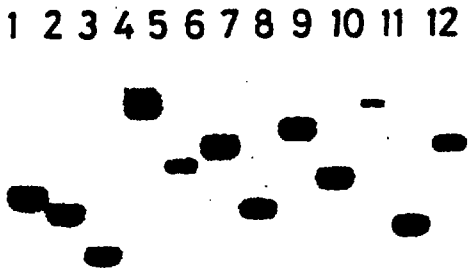


Fig. 2

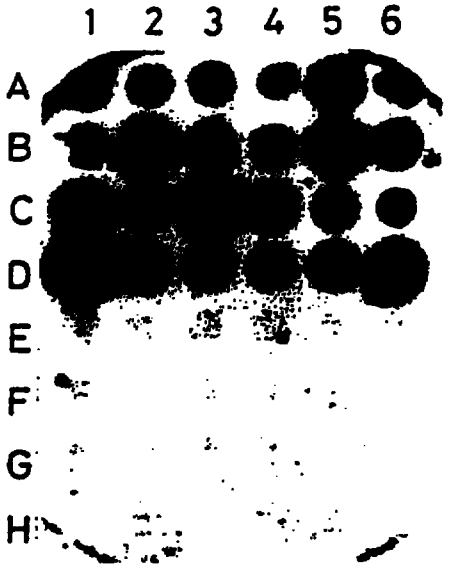
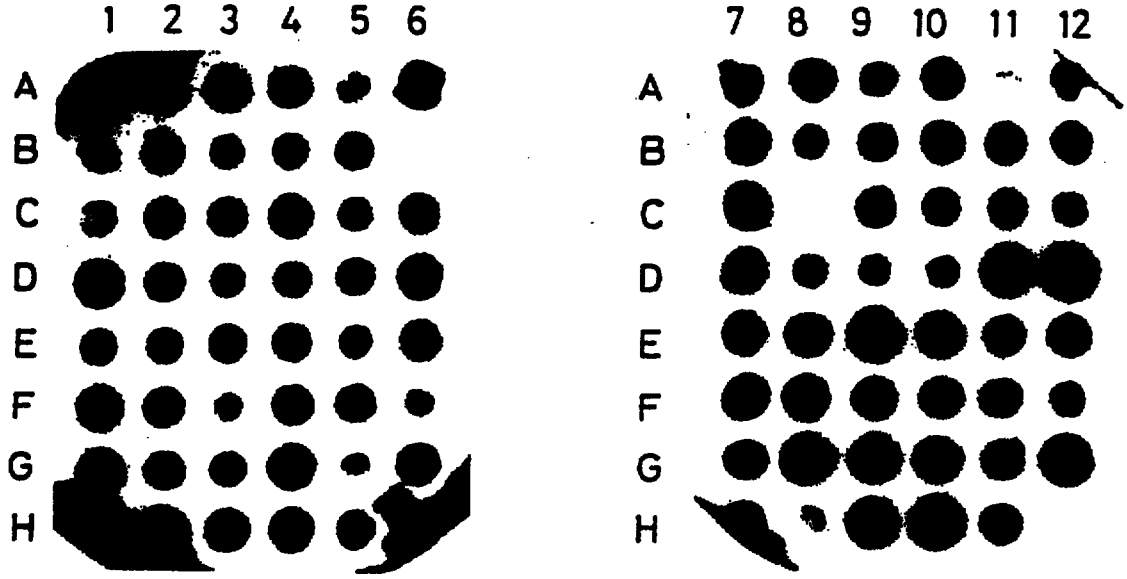


Fig. 3

Inoculum pattern



Spleen pattern

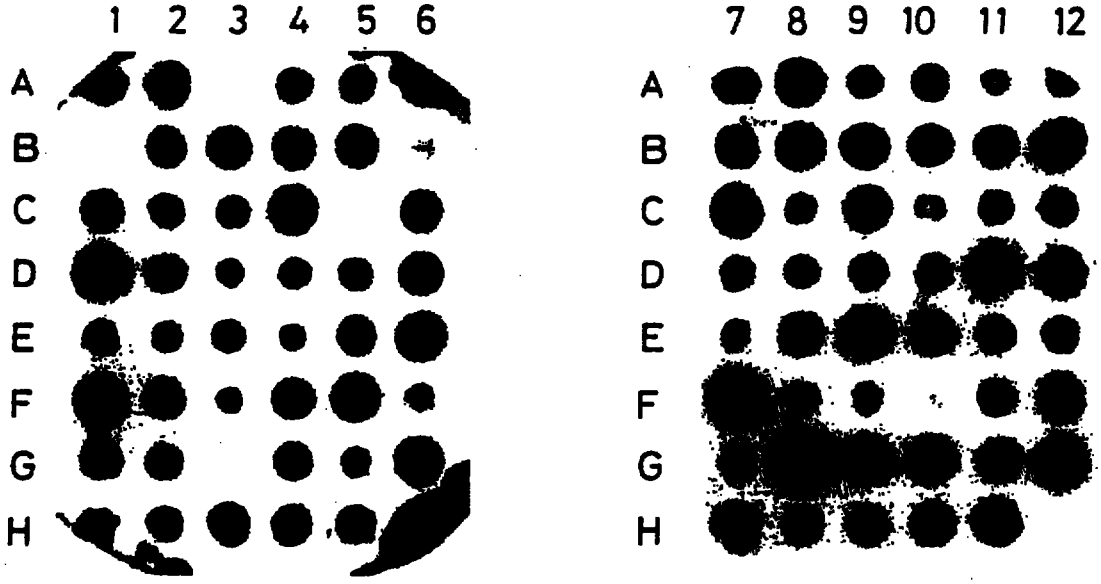


Fig. 4

```
Name: mpcc2 1

J05534 Escherichia coli ATP-dependent clp protease proteolytic
  component (clpP) gene, complete cds.
 Length = 1236

  Minus Strand HSPs:

 Score = 453 (125.2 bits), Expect = 4.3e-28, P = 4.3e-28
 Identities = 113/141 (80%), Positives = 113/141 (80%), Strand = Minus
                        Query is our Salmonella sequence
Query:    359 CCACCAGCCGCTGGGGTACCAGGGCCAGGCGACGGATATTGAAATTCACGCCCGCGAAAT 300
              ||| | |  |    || ||||||||||||||||| ||||| |||||||| ||||| |||||
Sbjct:    785 CCAACCGTTGGGCGGCTACCAGGGCCAGGCGACCGATATCGAAATTCATGCCCGTGAAAT 844
                                                        c/pP gene
Query:    299 TTTGAAAGTAAAAGGGCGCATGAATGAACTTATGRMKYKMMATACGGGTCANTCTCTTGA 240
              | ||||||| |||||||||||||||||||||||||        |||||||||| ||  | ||
Sbjct:    845 TCTGAAAGTTAAAGGGCGCATGAATGAACTTATGGCGCTTCATACGGGTCAATCATTAGA 904

Query:    239 GCAGATTGAASGTGATACTGA 219
              ||||||||| ||||||| ||
Sbjct:    905 ACAGATTGAACGTGATACCGA 925

 Score = 231 (63.8 bits), Expect = 4.0e-24, Poisson P(2) = 4.0e-24
 I  tities = 55/66 (83%), Positives = 55/66 (83%), Strand = Minus

Query:    194 TGAAGCGGTAGAGTACGGTTTGGTTGACTCAATTTTGACCCATCGTAATTGATGCCCTGG 135
              ||||||||| || |||||| |||| || || ||| ||||||||||||||||||||    |
Sbjct:    950 TGAAGCGGTGGAATACGGTCTGGTCGATTCGATTCTGACCCATCGTAATTGATGCCAGAG 1009

Query:    134 ACGCAA 129
              ||||||
Sbjct:   1010 GCGCAA 1015

>ECCLPXGNA Z23278 E.coli ClpX gene, complete CDS
           Length = 1945

  Minus Strand HSPs:

 Score = 364 (100.6 bits), Expect = 1.6e-20, P = 1.6e-20
 Identities = 88/107 (82%), Positives = 88/107 (82%), Strand = Minus

Query:    325 GATATTGAAATTCACGCCCGCGAAATTTTGAAAGTAAAAGGGCGCATGAATGAACTTATG 266
              ||||| |||||||| ||||| |||||| ||||||| ||||||||||||||||||||||||
Sbjct:      1 GATATCGAAATTCATGCCCGTGAAATTCTGAAAGTTAAAGGGCGCATGAATGAACTTATG 60

Query:    265 RMKYKMMATACGGGTCANTCTCTTGAGCAGATTGAASGTGATACTGA 219
                        |||||||||| ||  | || ||||||||| ||||||| ||
Sbjct:     61 GCGCTTCATACGGGTCAATCATTAGAACAGATTGAACGTGATACCGA 107

 Score = 231 (63.8 bits), Expect = 6.8e-24, Poisson P(2) = 6.8e-24
 Identities = 55/66 (83%), Positives = 55/66 (83%), Strand = Minus

Query:    194 TGAAGCGGTAGAGTACGGTTTGGTTGACTCAATTTTGACCCATCGTAATTGATGCCCTGG 135
              ||||||||| || |||||| |||| || || ||| ||||||||||||||||||||    |
Sbjct:    132 TGAAGCGGTGGAATACGGTCTGGTCGATTCGATTCTGACCCATCGTAATTGATGCCAGAG 191

Query:    134 ACGCAA 129
              ||||||
Sbjct:    192 GCGCAA 197
```

Fetch ⟶ Gb_ba:Ecoclppa
- OK then type J Biol Chem 265, 12536 (1990)

Figure 5

A) new virulence factors with similarity to sequenced genes:

1. p1F10

similarity to *clpP (E.coli)*

(Figure 5 of application)

2. p2D6

similarity to *lcrD (Yersinia spp.)*

sequence p2D6_1_I

GGTCTTAATGTACGGGCATGGTCTGCATCGATAACTCCGGCACGCAAATCGCCATCGATACTCATTTGT
TTGGCTGGCATCCCATCAAGCGAGAAACGTGCGCTAACTTCCGCCACCCTCTCGATACCTTTTGTAATG
ACAATAAATTGCACGATAGTAATGATGGTAAATACGACCAACCCAACGGTGAGATTTCCTCCTACGACA
AACTTACCGAAAGCATCCACAAATATTACCGGCATTATGTTGTAACAGTACCCAGCCGTGATGTGCTGA
TTGGGGAGTTAACAACCGATTTAT

3. s4C3

probably same gene as p2D6, but different region

similarity to *S. typhimurium invA* and *Yersinia spp.lcrD*

sequence s4C3_1_U

GCGCGGACGCTAGTGTGGTGGGTGACAGCCAGACGTTACCGAACGGGATGGGGCAGATCTGTTGGCTTA
CAAAAGACATGGCCCATAAGGCGCAAGGTTTTGGGACTGGACGTTTTCGCGGGCAGACAACGTATCTCT
GTCTTATTAAAATGTGTCCTGCTTCGGCATATGTATCGAACCCTCGGAGCAAAGTCGTTTGGGCGCAGA
ATTAGTACGTTTGGGTCGGTTGCTGTTATTCCTTGGGCTCGGAAAAAGAGTGCCAGCGTGAAGGAGTGG
GATTTGGCAGACTGGCCGCCTAAT

sequence s4C3_1_R

CACTATAGGGAAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCTACTAGTCATATGGATTGCACTT
GTGTATAAGAGTCAGGATTAGAGGACATGCGCCGGGAACCATACTATCTTTTTCCGGTGCTTCGACGCC
ATTTGCGGAAACCACAGACTTTTTGCGGCGAATGAGGATAATTGGCAATGCTAACAACGCTGAAAAGAA
AGCGAGAGTGATAAAAGGAAAGCCAGGAATTAAAGCGAGGAGCATTAAAACCACAGCGGCTAATATGAG
CGACTGAGGTTGTCTGGCAATTTG

4. p3F4

similarity to *invG (S.typhimurium)*

sequence p3F4_1_U

TGCAGGCCGACTCTAGAGGATCCCCGGGTACCGGTAATTTCTTTAACCTCGCATCCCGGTGGATGAAAG
GATATTCTGGCTGCGTAAGTAATGAATGAACCGCCCAGTAGATAAAATATTGAAAGTGATAACCTGATG
TTTTAATAACGATGCAGGATATACATATAACATGCTGGCATCAAACCAGGTAAGCAAATCATATTGTGC
TGCCAGGTTATTCAAACTATCGACCGGTGGTCCAGGCGGGAATTTTTCCACTAAATGTAGGTGGGATCA
ATGGGCTAATTGGTATAGGCGGAT

Figure 6 Sheet 1 of 5

5. p7G2 7/39

similarity to *yscC (Yersinia spp.)*

sequence p7G2_1_U

CCTGTGATTCCGGATGAAATAGCTTTTACGAAAGCTGTCAGACNTGCTGAAGAATACGCTGCAAATGGT
AAGCTTGTAACTTTTGGGTATTGTTCCAACGCATGCTGAAACGGGTTATGGATATATTCGTCGCGGTGA
GTTGATAGGAAATGACGCTTATGCAGTGGCTGAATTTGTGGAGAAACCGGATATCGATACCGCCCGTGA
CTATTTCAAATCAGGGGAAATATTACTGGCCTAGCGGCGATGTTTTTATTTCGCGCAAAGCCCTTATTT
AAACGAATTAAACGTATCTATCACCCCCAAATTCATACAGCTTGTGAA

sequence p7G2_3_O

TTACTAAACAGGGCCCCGGACCaTGTAAACACCACGCtTGCCAAcACTAAAAAACGATGCtTGCcGTAA
AAAAATTGAAcGTTATTTACTTAATAcGCCTATTTTATTTACATTATGCACGGACAGAGGGTGAGGATT
AAATGGATAATATTGATAATAAGTATAcTCCACAGCTATGTAAAATTTTgGGGGcTATATCgGATtTGg
TTGtTTtTAATTTAGCCtTATGGcTTtCACTAGGATGTGTCTATTTTTTTtGTGGtCAAGCACAGAGAT
TTATTCCCCaACCACC

sequence p7G2_1_I

TTTCCTTGCCGTGACAGTCCGGGATGCGAGGTTAACGAAATTACCGGCACCAAAGCTGTGGAGGTGAGC
GGTGTCCCCAGCTGCCTGACTCGTATTAGTCAATTAGCTTCAGTGCTGGATAATGCGTTAATCAAACGA
AAAGACAGTGCGGTGAGTGTAAGTATATACACGCTTAAGTATGCCACTGCGATGGATACCCAGTACCAT
TATCGCGATCAGTCCGTCGTGGTTCCAGGGGTCGCCTAGTGTATTGCGTGAGATGAGTAACACCAGCGT
CCCGACGTCATCGACGAACAATGG

6. p9B7

similarity to *fliQ, invX (E.coli)*

sequence p9B7_1_I

CATGAGTAACCTACCCAACTGTAATCTTTACCAATATGCATCATAATCTTCTGCTGGTAAATGATTGGT
AATATCGGAAAGGTAAGTGACATAAGCACGCCATTACGTAAAAGTGCGGCCCCTAAACTGCCACTTTTT
AATAAGGGAAGTAATAAAGAAAGGCTCAATGGTCGAATAAAAGCCACAGCCAATGCAATAAGCCACTCA
TTTACCTGTTGTGCCATTCAACCATGCTCTCCAATTCGTAACATTATCTGCCGGGTATAATTCAACAGG
ATACCGCTAAGCCATGGGTAG

sequence p9B7_3_O

ATTCCAGCCCCCGGGCCATCTAACCACTATGAACAATCATCTTCTGGGTGGACAATCATTGGTACCATC
GGCCAGGCTTGTGCAATATGTATGTCATCACGTAAAAGCGCGGCCCCTTAATCTCCCCATTCTTCCTTA
AGGGCAGTTATCACGGCTGGCTCAATGGCCGGCTTAACAGCCACAG

7. s6F5

similarity to *yscU (Y. enterocolitica)*

sequence s6F5_1_O

GAGGCGCGTCTTCGGTTGAGGGTCGCCCTCCAGATCTTTATGCTCCTGTTTTACGTCATCTTTACTCAT
TTTAAGATCTTTTCTAATCTTATAATATTGAAAAGAATAGTCCAGTATGCCAACGACGAAATAAAGAAA
CATCACCCCAACCCATAACCATTTTTTCAATGATGAAAGCACAAGCACGCCACAGGCTACCACCACAGCC
CGGAGGGGGCCGGAAAGTGCTGGGATCTTGATTAATGAAAAAGGCAAAGGGAAGAGATAGGATGATGCA
TGCTGGTTGGAGGCAGATTATTCATCTTCG

Figure 6 Sheet 2 of 5

B) new sequences without similarity to entries in DNA or protein databases:

1. s4D10

sequence s4D10_1_U

AGTTGCCGTATTTATTAAATATTCACCTCAGGTCAATATGGAGGTCTTCCCGGCTAAAAATCATTGCTT
TACTAGAGATATCACTCCCTGGGTTGCAATACAGTACGATTAGTTATCTTGATGCAGCCTGCTGATTTC
AGAATGGCAGCTGACGTACCCGCGAGACAAACATTCTGGATTATGGACGTTATCAACGCCAATATAGGG
AAGGTGGTGAAGTGGTTGATGAAATACCCCTATCCCTTGCATGTTATCGCTGACAGGACTGTTATCAGG
AGCGGGCATCCTCGATCGGCT

sequence s4D10_1_R

CAAGAGACAGATCCAACTCGGGCCGATCGCCATAACGCCAGCAGTTTGAAAGATGAAAGCCCAGCTTAT
CCAGCCATTCCGGTACAGCGTAACGAGCAGGTTGCCAGAAATAACGATAAAGTTGCAACACCTCGGGAT
CAGGTCGGCTCAAAAACGGGGTCTCAGGCAAAAATAGCCGATCAGGATGCCCACTCCTAATAACAGTCC
TGTCAACGATAACATCAACGGATAAGGGTATTTCATCAACCACTTCACCACCTTCCCTTTATTGGCGTT
GGATAACGTCCATAATCCAGA

2. s4H10

sequence s4H10_1_U

AGGGCTTTATTGATTCCATTTTTACACTGATGAATGTTCCGTTGCGCTGCCCGGATTACAGCCGGATCC
TCTAGAGTCGACCTGCAGAACCGAGCCAGGAGCAAATTAATTTTTTTGGGCAATTGCTGAAAGATGAAG
CATCCACCAGTAACGCCAGTGCTTTATTACCGCAGGTTATGTTGACCAGACAAATAGATTATATGCAGT
TAACGGTAGGCGTCGATTATCTTGTCAGAATATCAGGCGCAGCATCGCAAGCGCTTAATAAGCTGGGTA
ACATGGCATGAAGGGGCAACCC

sequence s4H10_1_R

CACTATAGGGAAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCTACTAGTCATATGGATTCCTAGG
CGGCCAGATCTGATCAAGAGACAGATCCAACTCGGGCCGATCGCCATAACGCCAGCAGTTTGAAAGATG
AAAGCCCAGCTTATCCAGCCATTCCGGTACAGCGTAACGAGCAGGTTGCCAGAAATAACGATAAAGTTG
CAACACCTCGGGATCAGGTCGGCTCAAAAACGGGGTCTCAGGCAAAAATAGCCGATCAGGATGCCCACT
CCTAATAACAGTCCTGTCAACG

3. p4G5

sequence p4G5_1_O

CCCCCCCCCTTCTCCTGGCTTACACAGCCCCAGACCGGCGCTGGAAAAGGCCATTCCCGCCATACAGGA
GGCCAGCAACATATTTTCACGCGCCGCCAGATCGTGGCCGTAACCCACGGCTTTCGGCAGCGATTTGCC
AATCATCGCTATCGCGCCAATCGCCAGGCTGTCGGTAAACGGCGTGGCGTTGAGCGCGCTGTAGGCCTC
AATCGCATGCGTCAACGCATCGATACCGGTCATCGCCGTCACGTTTGGCGGAACGCCTTCGGTCACGGA
AGCATCAAGAATCGCCACGTCCGGC

sequence p4G5_1_U

CGCGAACGTGCGCCGCAACTGCTTGTGGACGGTGAATTGCAGTTTGACGCCGCTTTCGTGCCGGAGGTC
GCCGCGCAAAAAGCGCCTGACAGCCCGCTGCAAGGCCGCGCCAACGTGATGATTTTCCCGTCGCTGGAG
GCGGGCAATATTGGCTACAAAATCACTCAGCGTCTGGGAGGCTATCGCGCTGTTGGGCCGCTAATTCAG
GGGCTTGGCGCGCCGCTTCACGACCTCTCCCGAGGCTGTAGCGTGCAGGAAATTATCGAACTGCGGTTG
GTGAGAAAACCAA

Figure 6 Sheet 3 of 5

4. p7A3

sequence p7A3_1_U

CGCCCTAGCATGCCTGGCGTTGTCCGGTTATTGCTCGTCAAGCGAACAGATGCAAAAGGTGAGAGCGAC
TCTCGAATCATGGGGGGTCATGTATCGGGATGGTGTAATCTGTGATGACTTATTGGTACGAGAAGTGCA
GGATGTTTTGGATAAAAATGGGTTACCCGCATGCTGAAGTATCCAGCGAAGGGCCGGGGAGCGTGTTAA
TTCATGATGATATACAAATGGATCAGCAATGGCGCAAGGTTCAACCATTACTTGCAGATATTCCCGGGT
TATTGCACTGGCAGATTAGTCACTCTC

sequence p7A3_1_I

CCCTTCCCAGGCTCGACAGGTACACAGCCAGCCACTGGTGCAGGCAGTTACTTGCTTTCATCATGGGAA
GGAGCAATATCCTGATATATTAAAGAAAGAGCGGGATCCCCTTTCTTTACTGCTGCTAACGTTTCTTGC
AAAATGCGTTGATGAGATTCATCCAGCACACCACTGATAACAAAAGAGCGCCGCATTGGCGTAACATTG
ACAAGCCCCACTAAACCGCTCTCTATTATCGCAGAAATAATATCATCCCCCTGAGACTGATGAGAGTGA
CTATTCTGCCAGCGCAAATAACCC

5. p10E11

sequence p10E11_1

ATACCGAGTATTAAGCGGCTGTGTAACATCGTCATCCAACAACATACGCAGCGAGCCGCCACGCCGGAA
AAACCGCATCGTGTCATGTGCCTGTTGTAGGGTCGGGTCTTTTTTCATGAGTACGTTTTCTGCGCTATC
ATACTGGAAATTTCCCCCCACTTACTGATAAGCCCTGTCAGTTGGGTAAGGACAGAGTTAAGCTCCTGA
GACATTTTTTGGAATGGTTATCTTTCCCCGACTCATAAAATCGGTATTCCCGCTGGGGGCAATATCCAA
AGACGCTTTGGTCGCCCGTAGGGCACC

sequence p10E11_U

GCCGTATGCCTGCAGTTGCCCGGTTATTGCTCGTCAAGCGAACCGATGCCAAAGGTGAGAGCGACTCTC
GAATCATGGGGGGTCATGTATCGGGATGGTGTAATCTGTGATGACTTATTGGTACGAGAAGTGCAGGAT
GTTTTGGTAAAAATGGGTTACCCCCATGCTGAAGTATCCAGCGAAGGGGCGGGGAGCGTGTTAATTCAC
GATGATATTCAAATGGGTCAGCAATGGGGCAAGGTTCAACCCCCACTTGCAGATATTCCCCCCCCTATT
GGACTGGCAGATTAGTCACTCTCA

6. s4B9

sequence s4B9_1_O

GGGCGACCTGCCCGCGGCGCAACTTTCCCCGAAGCGTTTTCCATTTCCTTGTTCTTAAATGACCTGGAA
AGCTTACCTAAGCCTTGTCTTGCCTATGTGACAATACTGCTTGGAGAACACCCGGACGTCCATGATTAT
GCTATACAGATCACAGCGGATGGGGGATGGTGAATCGGTTATTATACCACAAGTCGCAGCTCTGAGCTT
ATTGCTATTGAGATAGAAAAACACCCCGCTTCAACTTGGATTTTGAATAATGTAATACGCAATCACCAT
ACACTATATTCGGGTGGCGTATAA

sequence s4B9_1_R

TTCGAGCTGGGGCACCGCTAATATCTTTAACCTCGCATCCCGGTGATGAAAGGATATTCTGGCTGCGTA
AGTAATGAATGAACCGCCCAGCAGATAAAATATTGACAGTGATAACCCGATGTTTTTTTAACGATGCAG
GCTATACATATAACATAGCTGGCCACCAACACAGCTGAAGTAAATCATATTGTTGCTGCCAGGCTACTT
CACACTATTGTCCGGCGGGCCAGCGGGGATTTTCCCCCTAAATCTCGCTGGTTCTCAAA

7. p4F8

sequence p4F8_1_I

AGTCTACGATTTCGCTATATCTTCTCTTAATCATGGCCGCCATTTGTGGATGCGATTTTAAAATATCCG
GGCGATCTTTCATTAAAAAATAAAGATTCCCCATGACTTCACAGATAAAGGTATCGGTATTTTGAGTGA
TACGTAACAATTCGTTCTCTTCGTGTGGGTCCATGATGCGAAGAATAATGGTGGCATCATTTTCATGAG

GATTATGAACCCGAAATCTTTCTCTTTGCGATGCGCAGGCTAACTCTTTCAACTCAAAAAAAATCTCTG
TAAGCCGCTCTCGTGTGGGGGCGC

8. p7B8

sequence p7B8_1_O

GCGCCCCTTTAATTGGTTGAGGCGGCTGGTATTCTTGTAAGGGTAATACTAGCGAGACCCAGGTTCCAC
CCCCGGGGACACTTTTTAGTGTCAGATTACCGCCCATCATTTTAGCCAGGCTTGACGCAATAGTCAGTC
CAATTCCTGTACCTTGCGAATTTGTGTCTGCTTGATAAAAAGCAGAAAAGATTTGAGACTGCTGCTGTT
TTTCAATCCCCCCCACCGCTATCGCTAACCAGAAATATTAATTGTTCCTCACCAAGATTGAGCGCCAGAC
GTATCCCTCCCCCCTCGGGAAAT

9. p8G12

sequence p8G12_1_I

GGATAAGATCCCGGATAAGTATGTCAGGCTCGTATGCACAACAGGCATTATAAACCTCTAGACCATTTT
TAACATGCTCTACTATTTTAAAATGAGGCCAGGGTAATAAGGCATTCATAATGCCGTTAATGATGATTT
CATGATCGTCTACTAATAAGATCTTATATTCTTTCATTTGGCTGCCCTCGCGAAAATTAAGATAATATT
AAGTAATGGTGTAGGTTGTGGAGATCATACGTATTTTCTGGCGTAAGTCGGTTAGTTCCTCCAGCGCGA
TGATTTTCCCCATTTTTACGCGAT

10. p9G4

sequence p9G4_1_O

TTCCATATTGCTCGTCCGGGGAGCGTGTTAATTCTTGATGATATACCAATGGATCTGCAATGGCGCAAG
GTTCAACCATTACTTGGAGATATTCCCGGGTTATTGTACTGGGAGATTAGTCACTCTCATCAGTCTCAG
GGGGGTGATGTTATTTCTGGGATAATAGAGCAACGGCGTTAGCAGGGGTCGGTCAGTAGTCACGGCCAA
CTTCGGTGCACTTTTGCGTATCACTGGGGTATCATAACTGAATCTCATCCCCCCCACTTTGGTAATCAC
AC

sequence p9G4_1_U

AATTCTTTTACCTCCATAAGCTGCGTGGCATAGCGATACAGAGTATTAAGCGGGTGTGTTACATCGTCA
TCCAACAACATACGCAGCGAGCCGCCACGCCGGAAAAACCGCATCGTGTCATGTGCCTGTTGTAGGGTC
GGGTCTTTTTTTCATGAGTACGTGTTCTGCGCTATCATACTGGAAATTTCCCCCCACTTACTGATAAGC
CCTGTCAGTTGGGTAAGGACAGCGTTAAGCTCCTGAGACATTTTTTGAGTTGTTATCTGCCCCCCGACT
CATAAGATCGGGTATTCCGCGGTGG

11. p9B6

sequence p9B6_1

ATATCCCTAATGCTTTTCCTTAAAATAAATACCACGGAAGGATACTGGCCACCTAGCCAAATTTAGAAA
GCAATGAACATCCGGTTTATTCCTGAAAACGATTACTCCGGCGCACGTTGTTCTGGCGTTACCTGAGCC
AGCAAACGATATAATGGGGTGGTGACCCGCATACCGGTCATTGGCATCCCATCCACACCGGAGGGAGTA
AAACTCATTAGGCCATAGGTAATATCATTAAGACGCTCTAATAAATGAGGGTGGGGGGCCCAAACTACC
ACTCCAGTATGTATTGAGTCA

12. p6G5

sequence p6G5_2_I

CCCATGGGCGCAATTTGTTGCGCAGCGTTTACCCGACCATCGCGTTTATGAGCTGTAATTCATGGGGGG
TAAAAACGGGCGTGACGACCCCAACGGAAGATAAGGCCGGGCTTAAACAGGAGATTATTGCTAATGCGC
AGCGCAAAGTGTTGCTGGCGGACAGCAGTAAGTATGGCGCGCATTCGCTCTTTAATGTGGTGCCGCTTG
AGCGCTTTAATGACGTGATTACCGACGTCAATCTGCCGCCGTCAGCGCAGGTTGAACTGAAAGGGCGCG
CTTTTTGCGCTAACG

Figure 6 Sheet 5 of 5

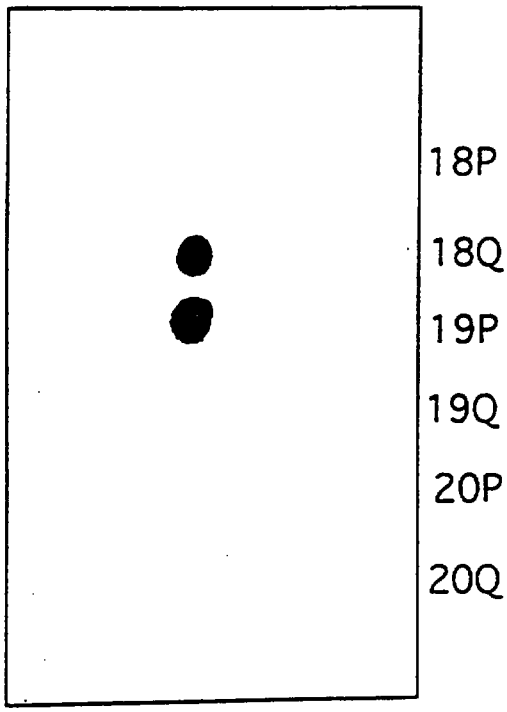


Figure 7a

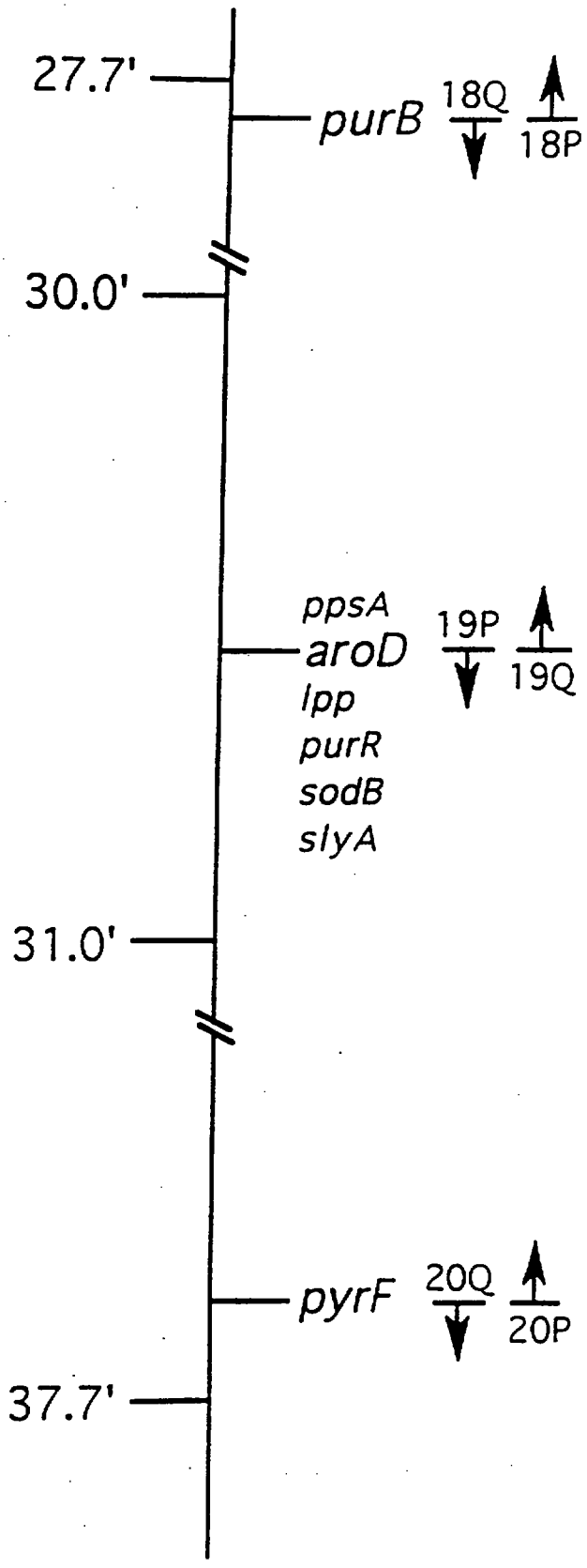

27.7'
purB
18Q
18P
30.0'
ppsA
aroD
lpp
purR
sodB
slyA
19P
19Q
31.0'
pyrF
20Q
20P
37.7'

Figure 7b

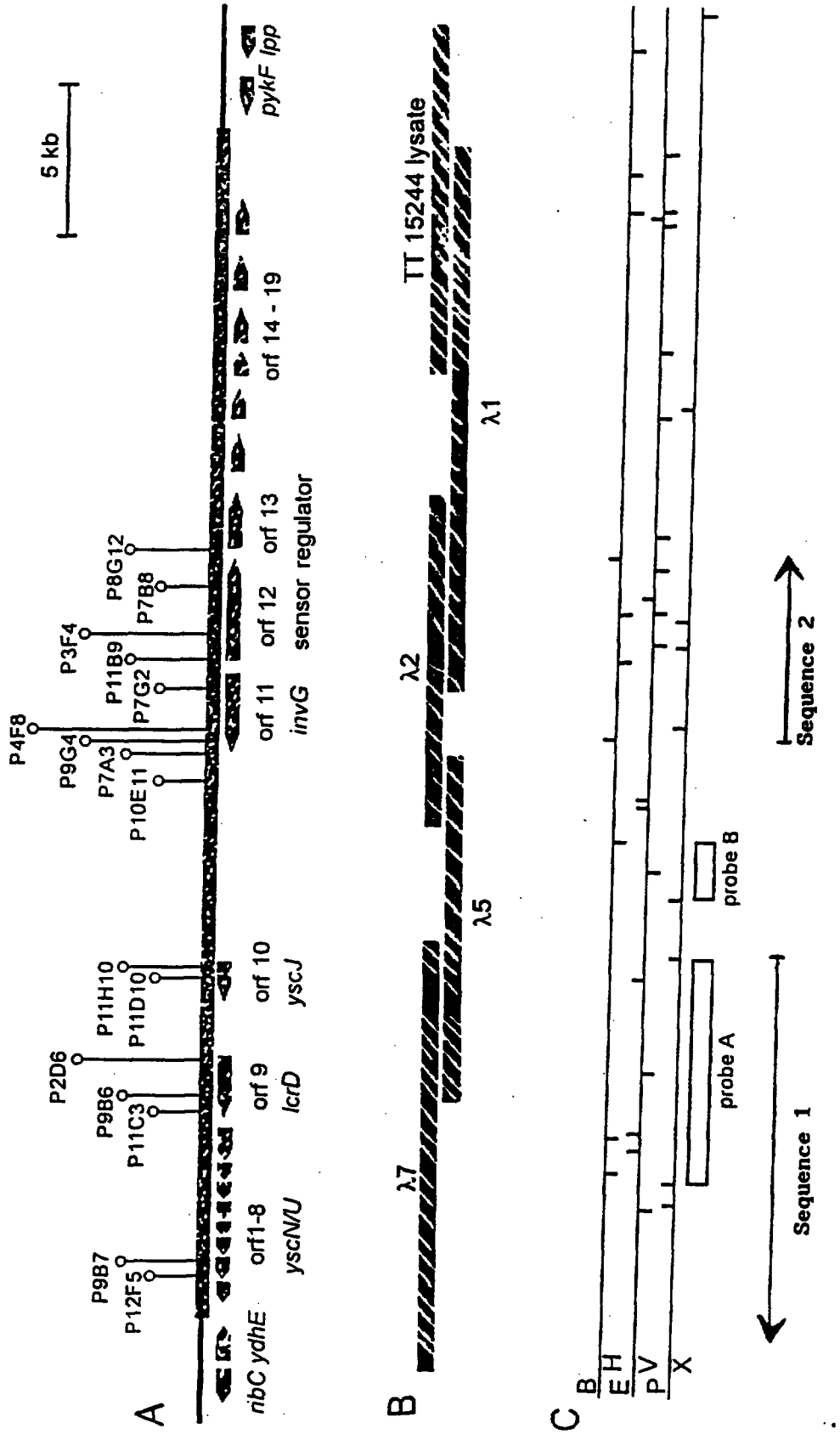

5 kb
pykF lpp
orf 14 - 19
orf 13
orf 12
orf 11
invG
sensor regulator
P8G12
P7B8
P3F4
P11B9
P7G2
P4F8
P9G4
P7A3
P10E11
P11H10
P11D10
P2D6
P9B6
P11C3
P9B7
P12F5
orf 10
yscJ
orf 9
lcrD
orf1-8
yscN/U
nibC ydhE
A
TT 15244 lysate
λ1
λ2
λ5
λ7
B
B
E
H
P
V
X
probe A
probe B
Sequence 1
Sequence 2
C

Figure 8

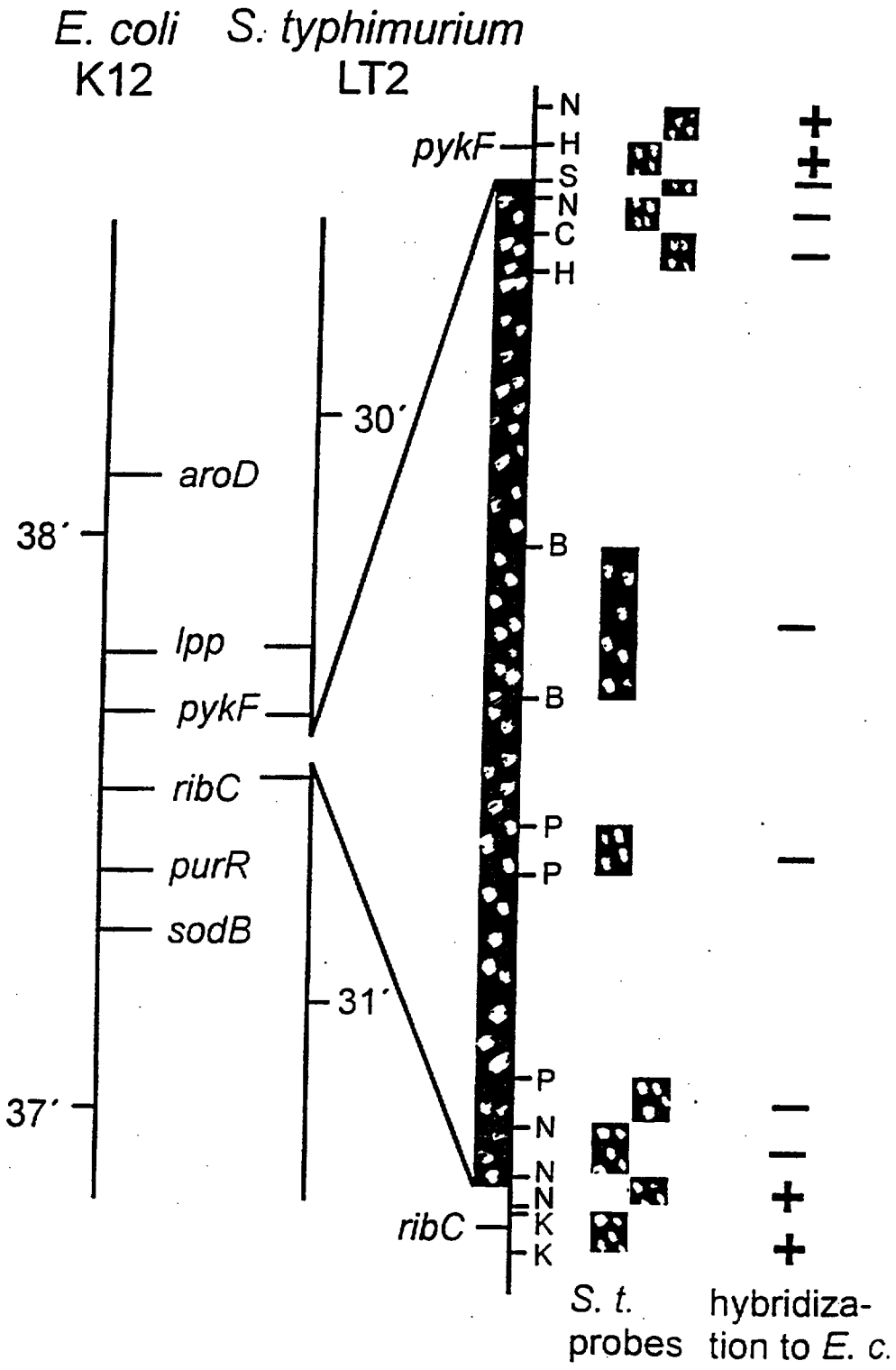

E. coli K12
S. typhimurium LT2
pykF
N
H
S
N
C
H
30′
aroD
38′
B
lpp
B
pykF
ribC
P
purR
P
sodB
31′
P
37′
N
N
N
ribC
K
K
S. t. probes
hybridization to E. c.

Figure 9

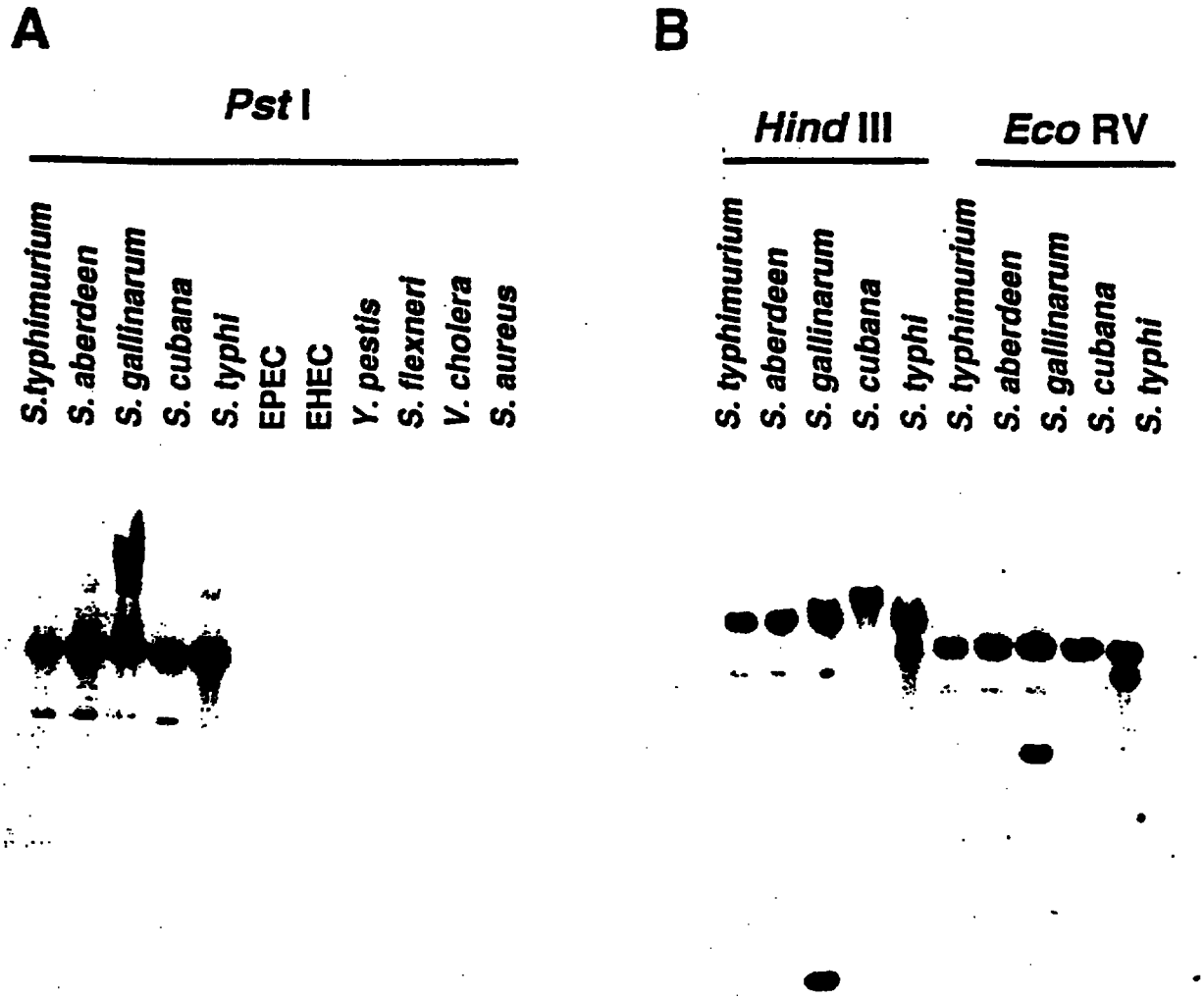
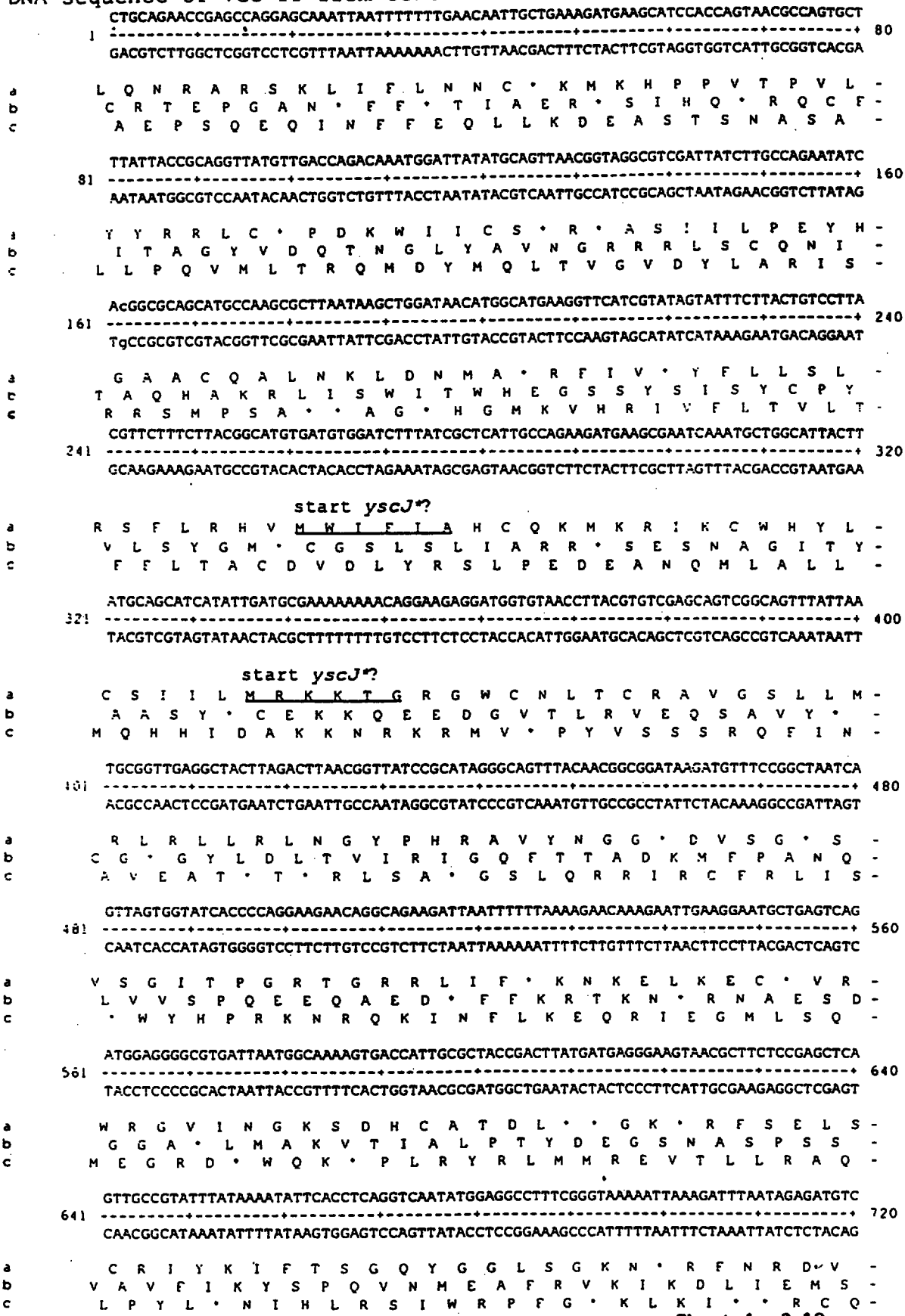

A
Pst I
S.typhimurium
S. aberdeen
S. gallinarum
S. cubana
S. typhi
EPEC
EHEC
Y. pestis
S. flexneri
V. cholera
S. aureus
B
Hind III
Eco RV
S. typhimurium
S. aberdeen
S. gallinarum
S. cubana
S. typhi
S. typhimurium
S. aberdeen
S. gallinarum
S. cubana
S. typhi

Fig. 10

DNA sequence of VGC II from centre to left hand end



Sequence 1

Figure 11

Sheet 1 of 19

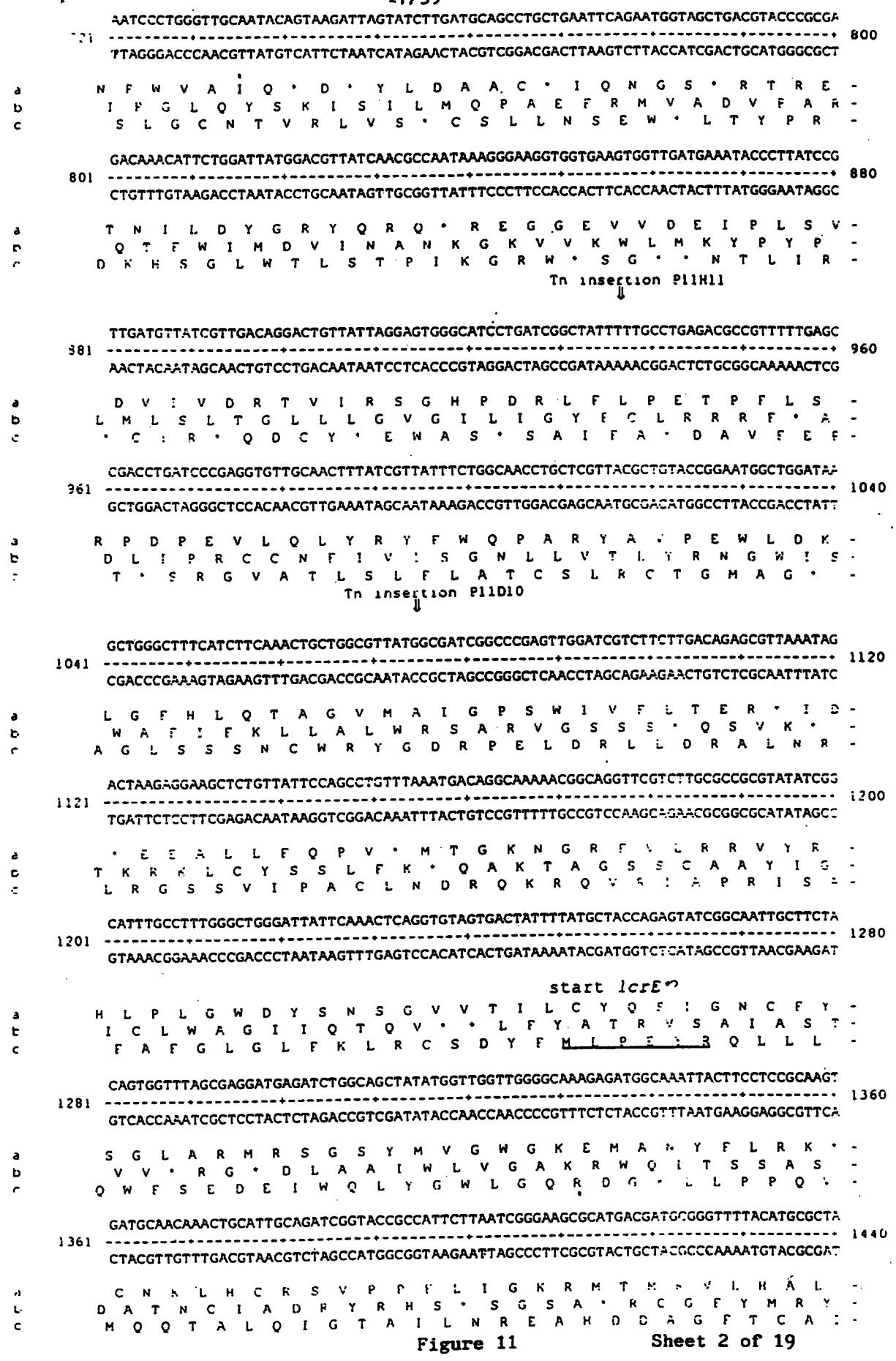


Figure 11 Sheet 2 of 19

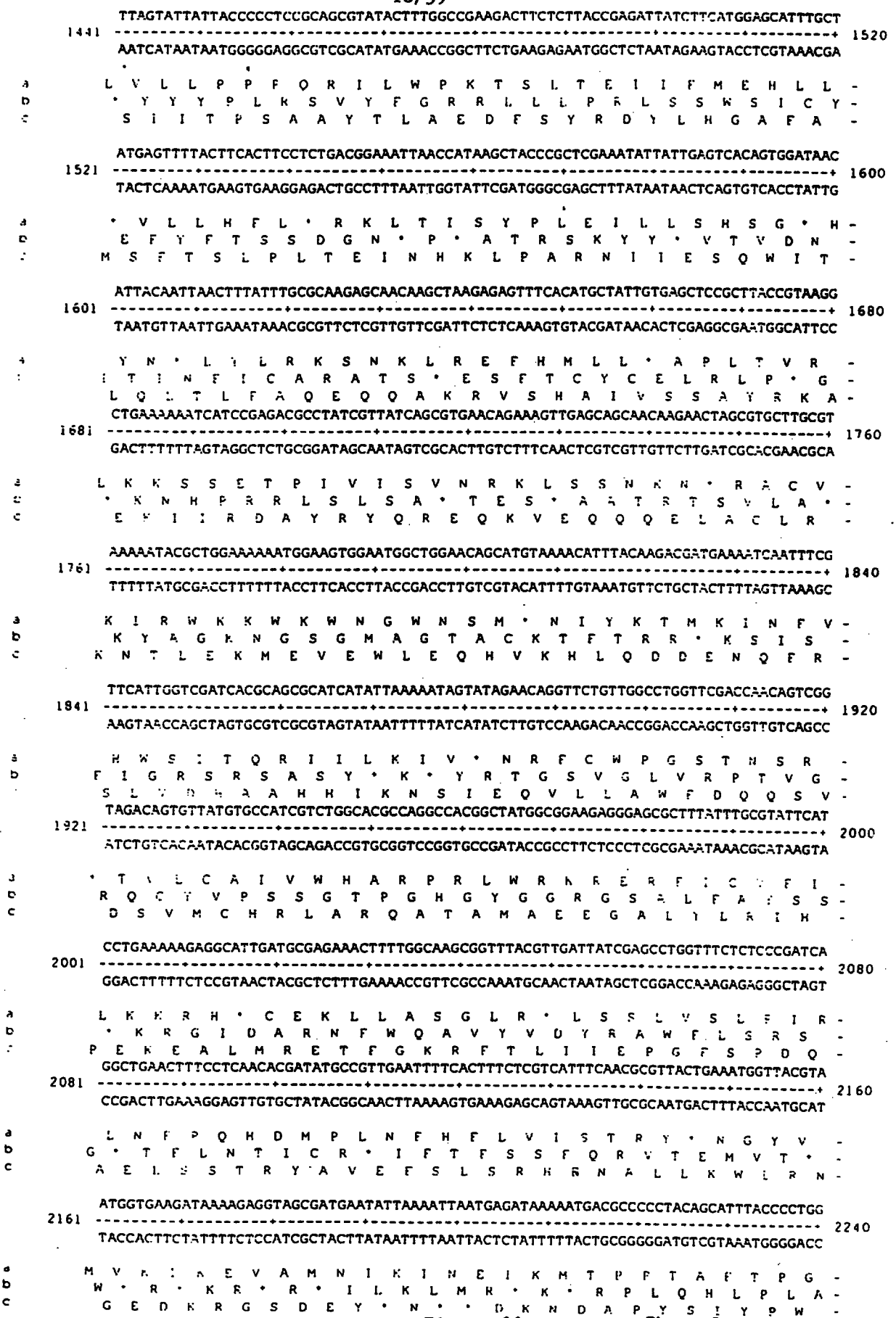


Figure 11 Sheet 3 of 19

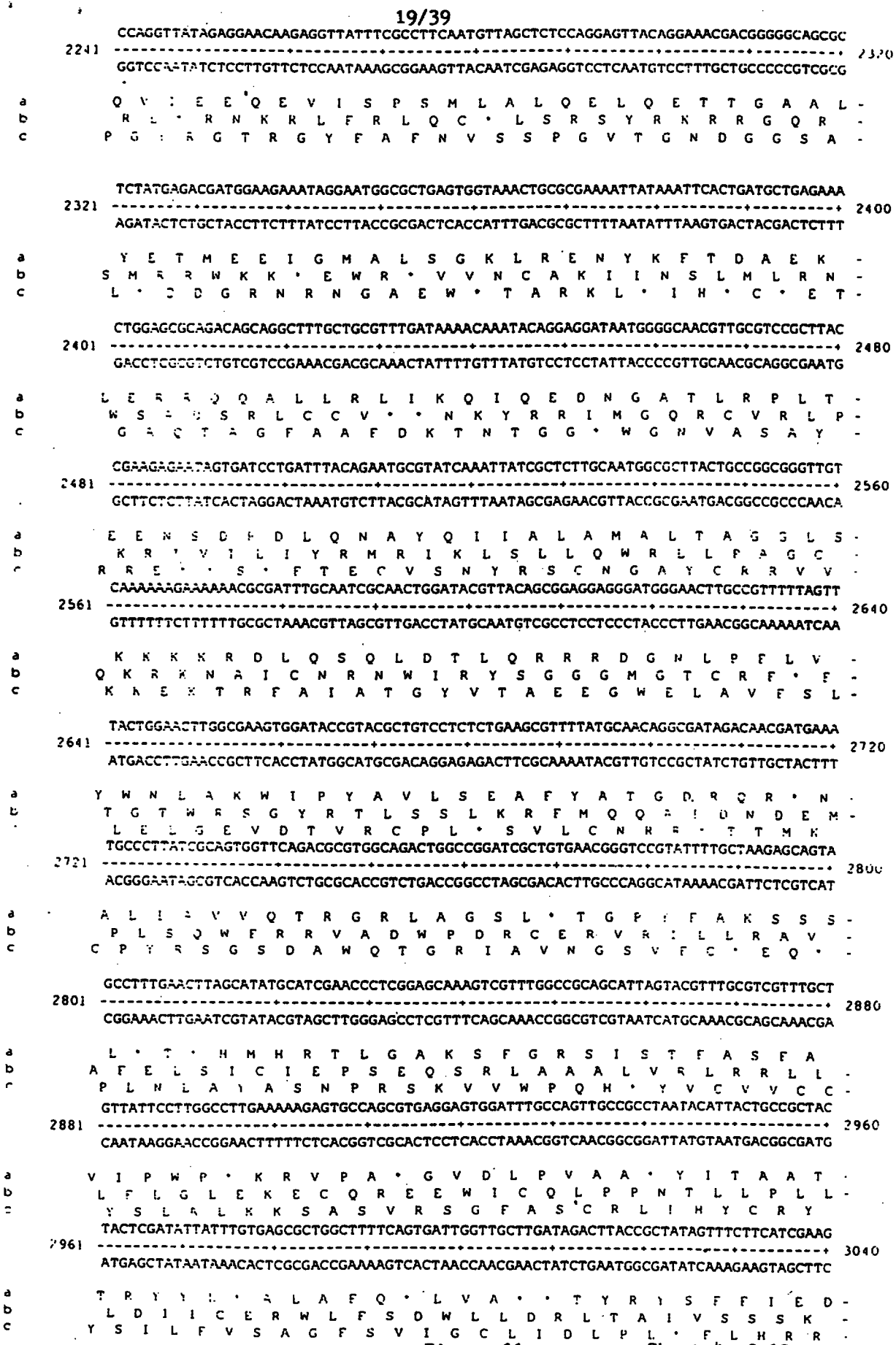


Figure 11 Sheet 4 of 19

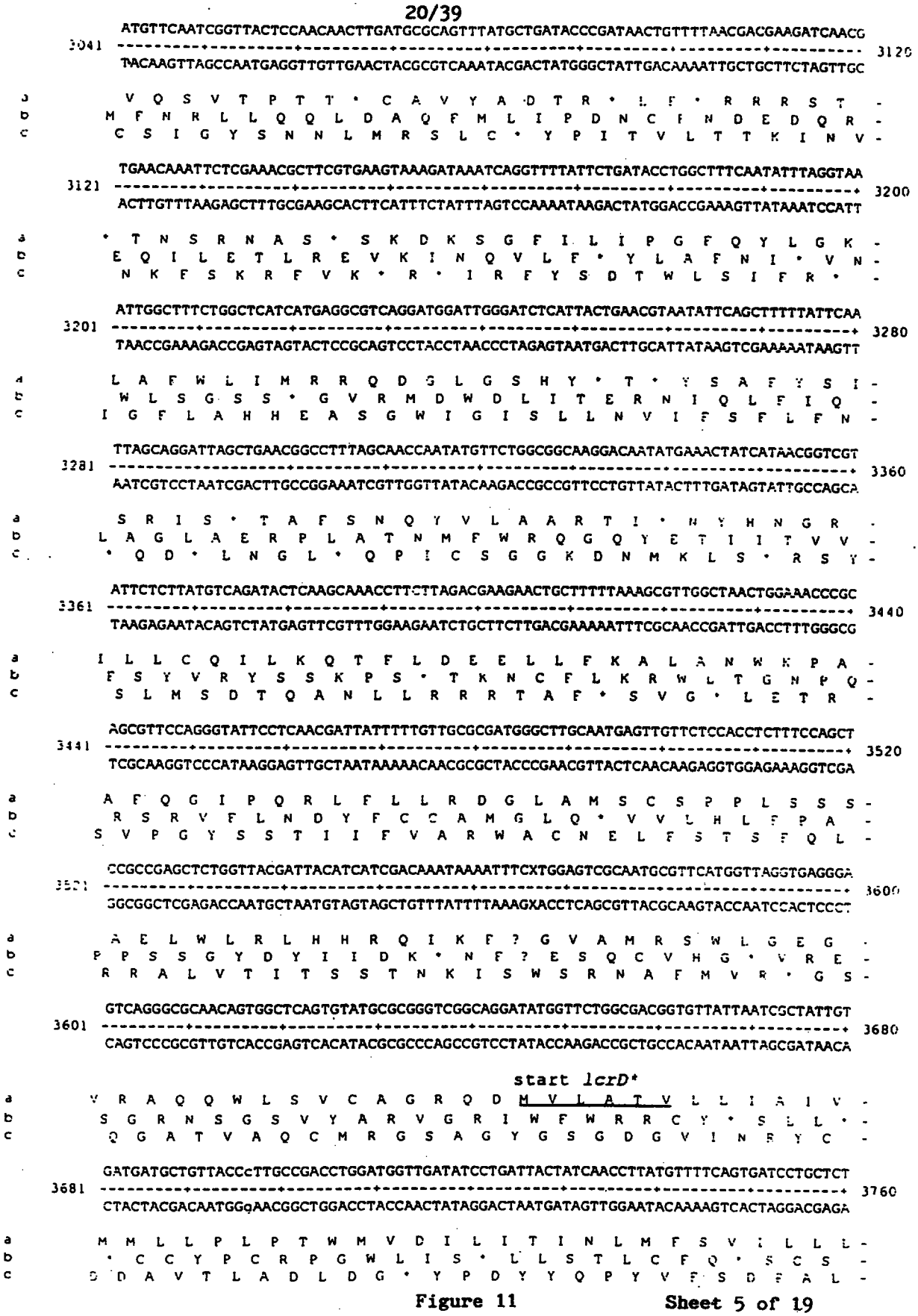


Figure 11 Sheet 5 of 19

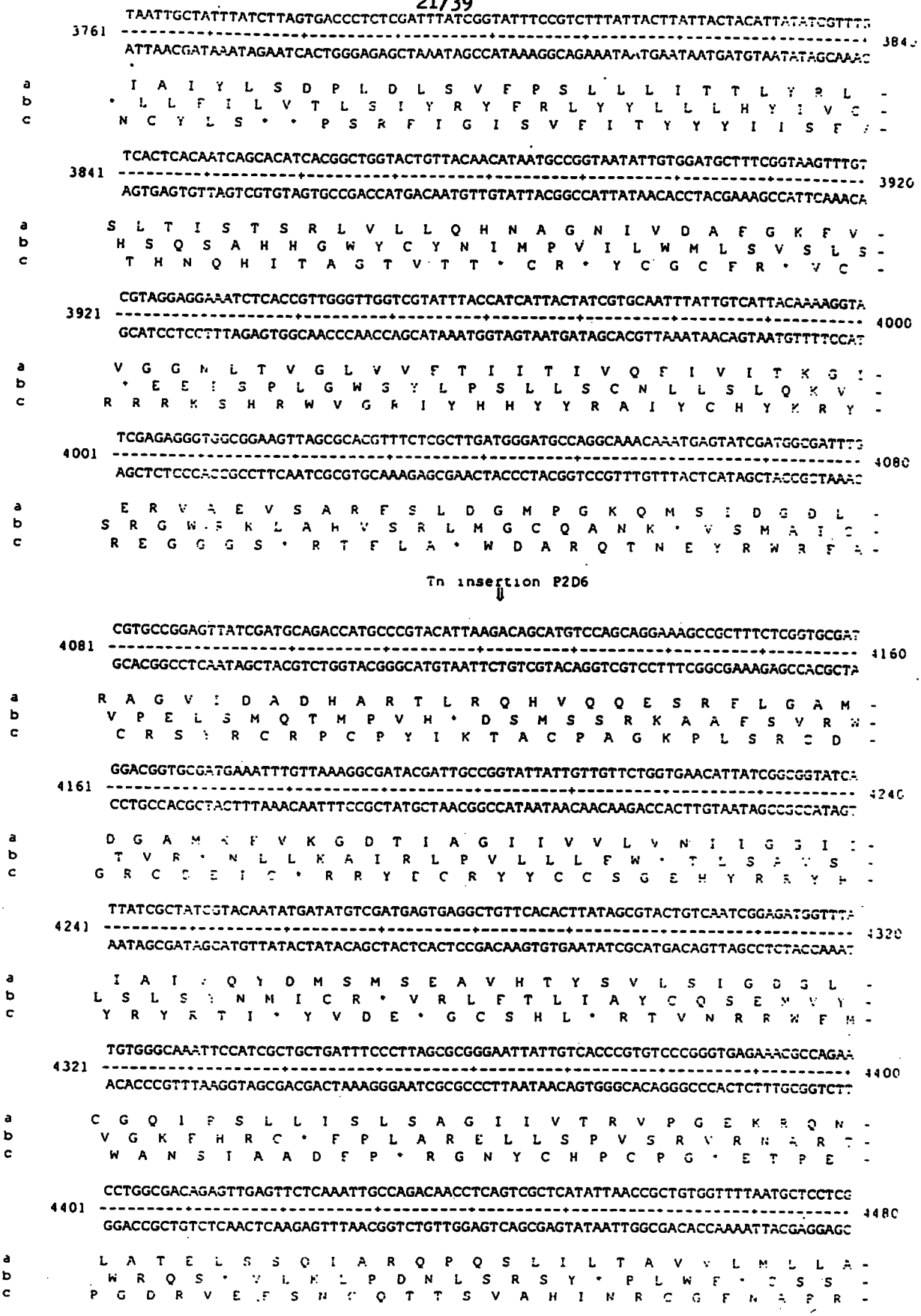


Figure 11 Sheet 6 of 19

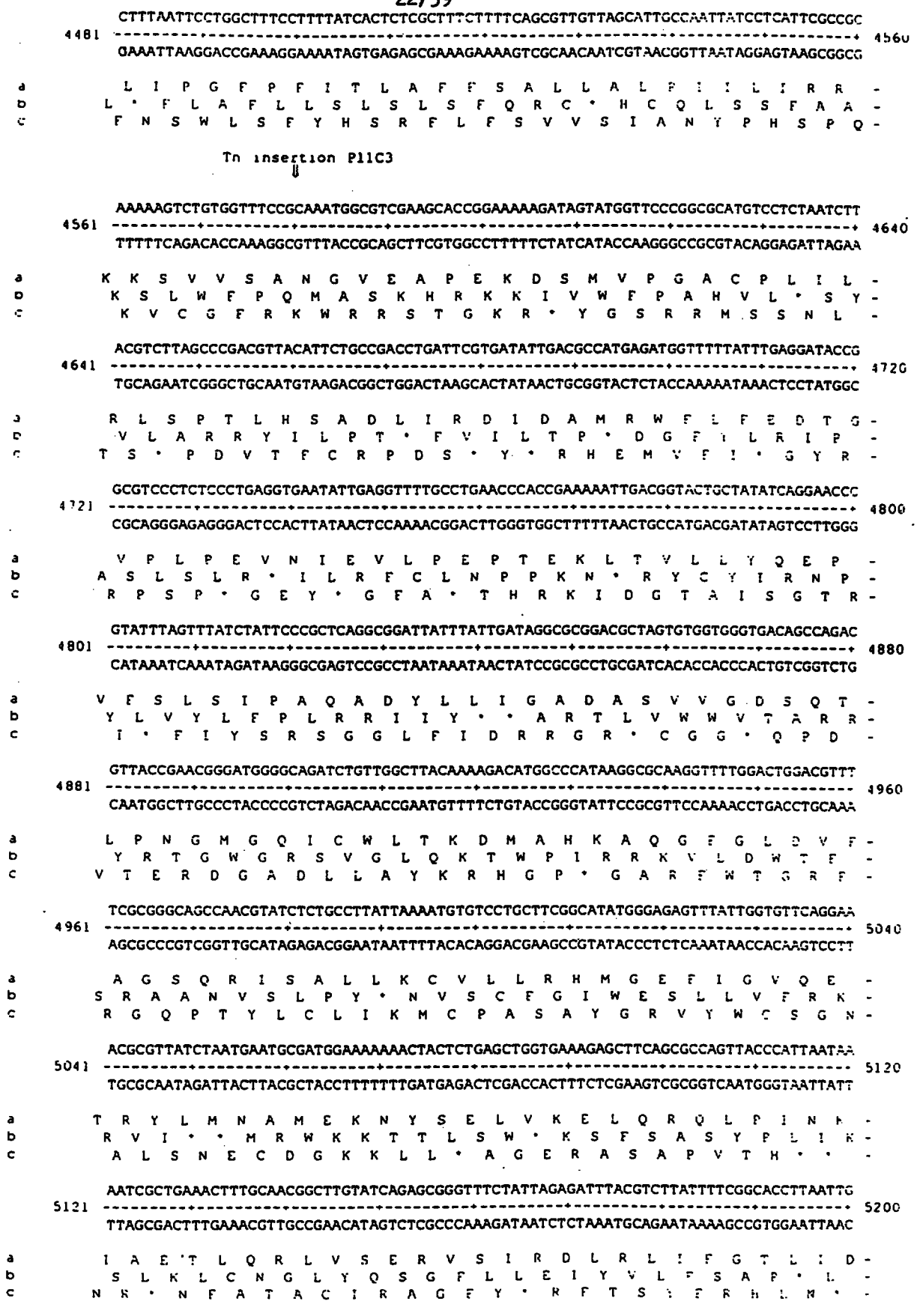


Figure 11 Sheet 7 of 19

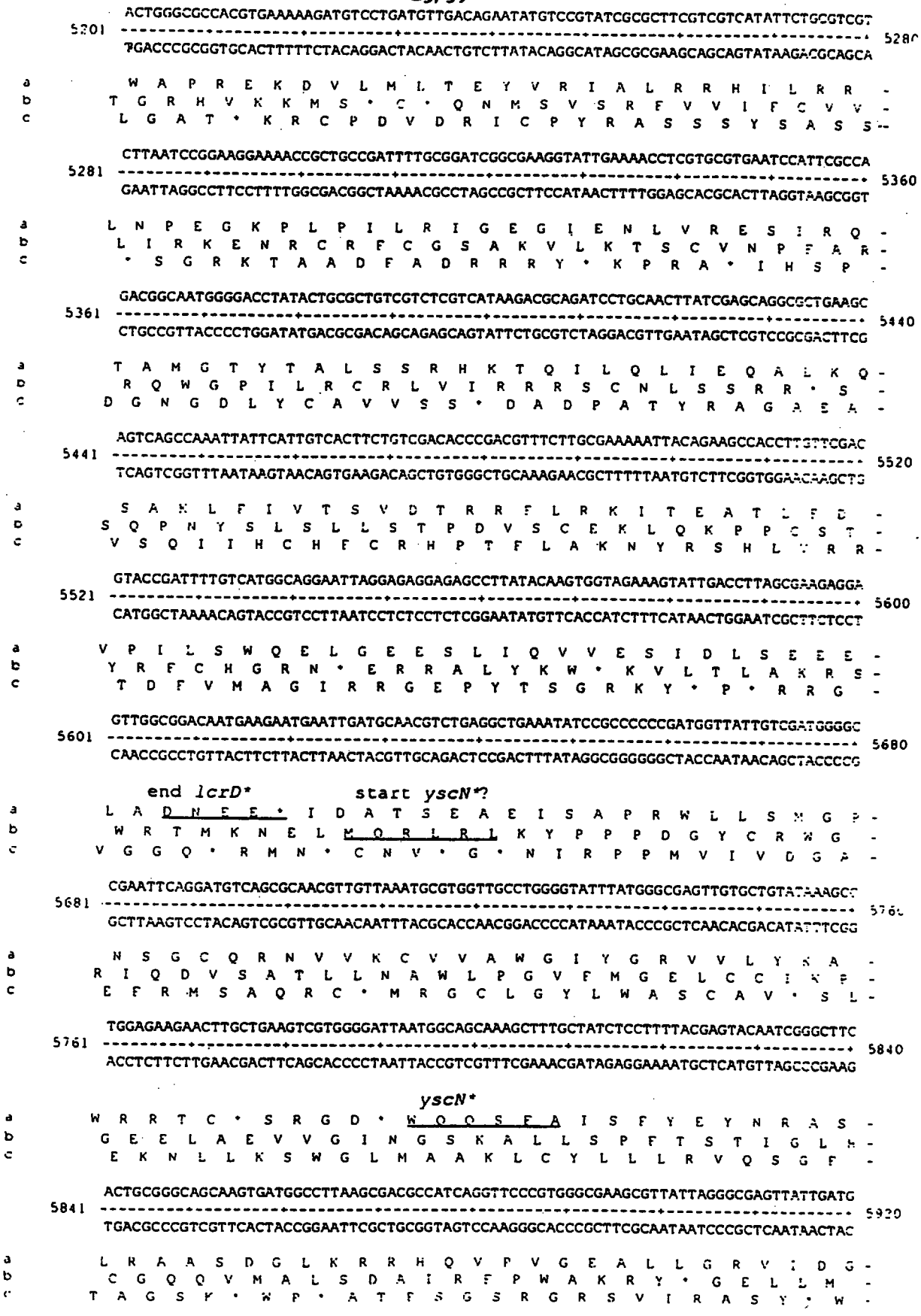


Figure 11 Sheet 8 of 19

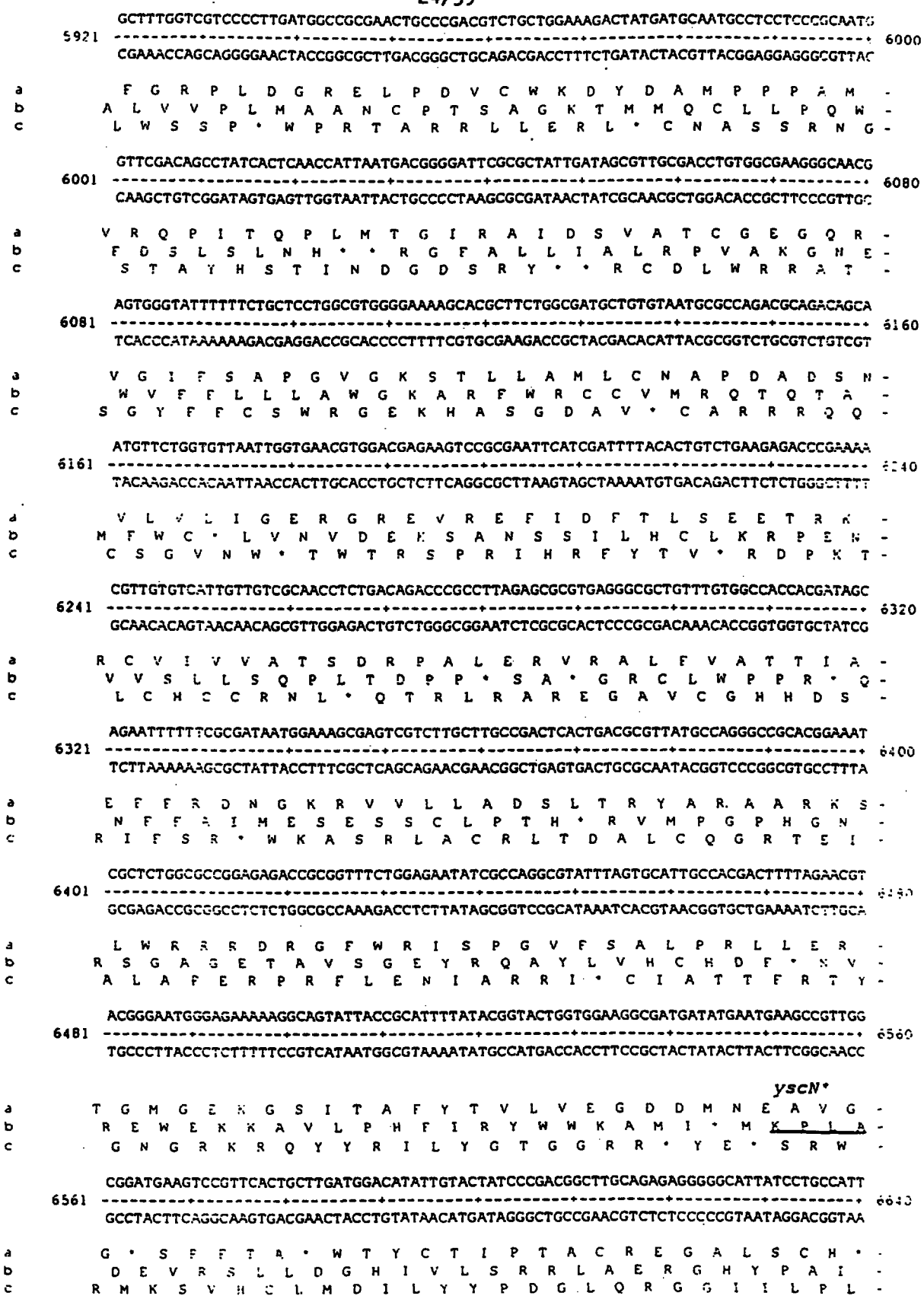


Figure 11 Sheet 9 of 19

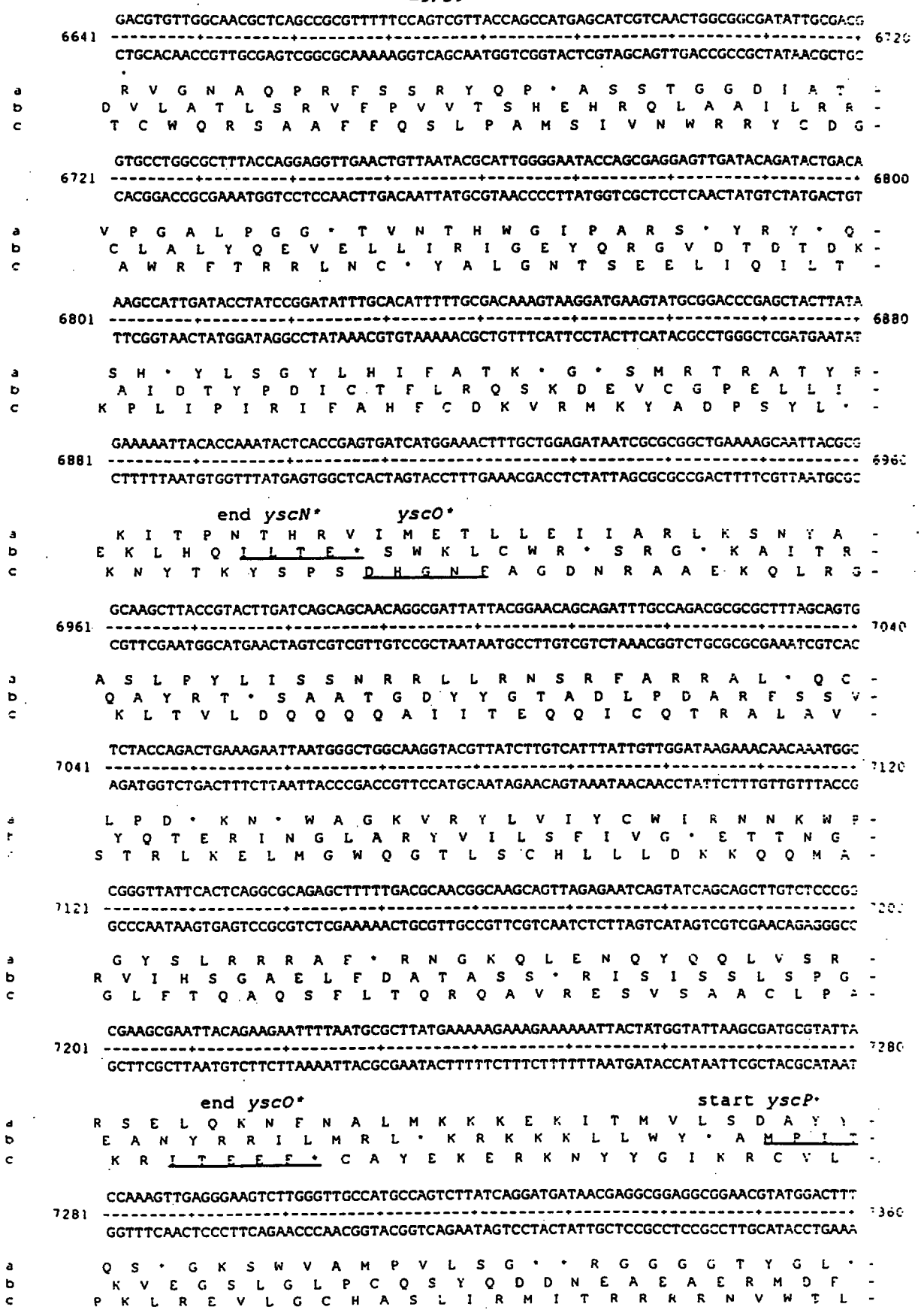


Figure 11 Sheet 10 of 19

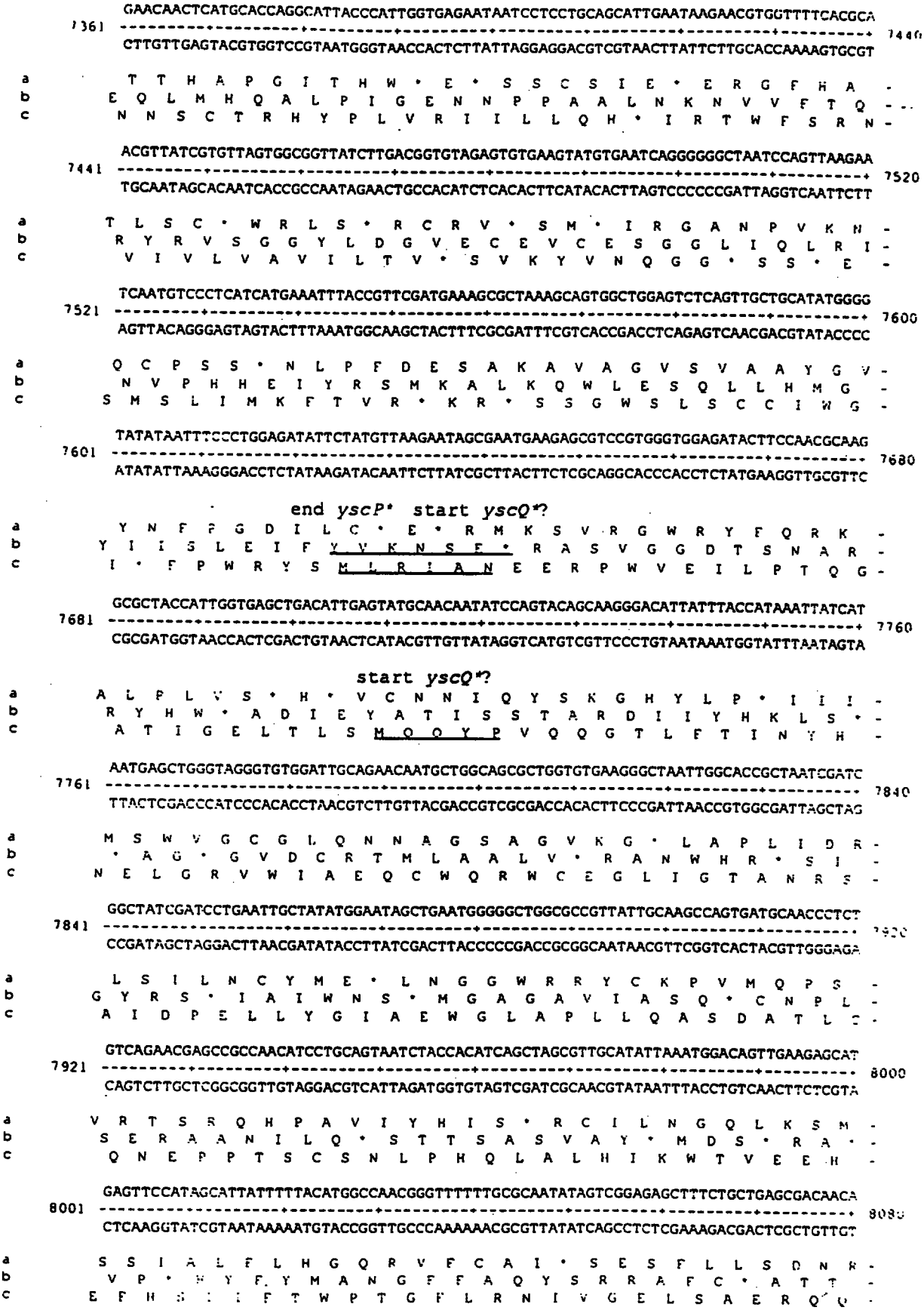


Figure 11 Sheet 11 of 19

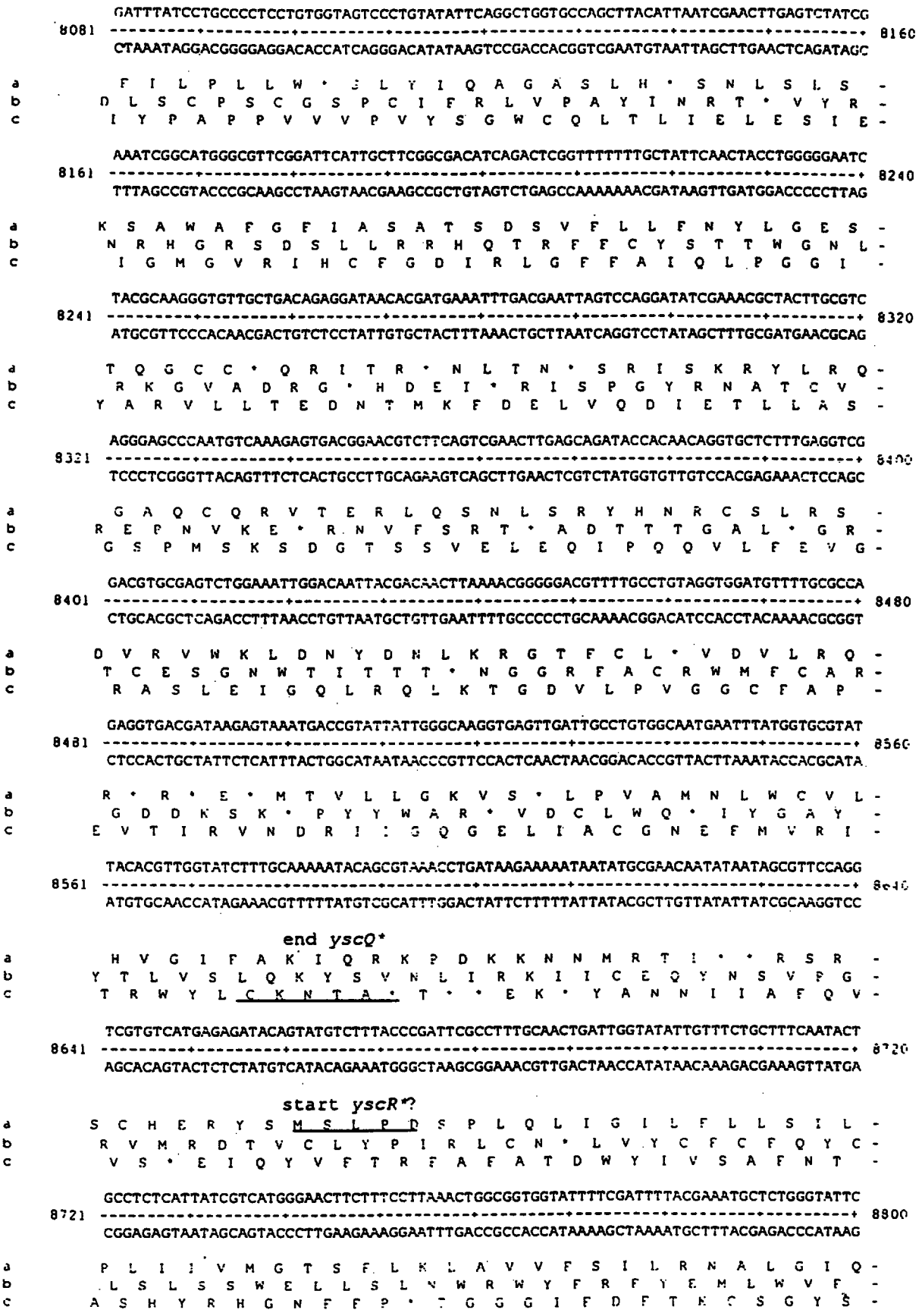


Figure 11 Sheet 12 of 19

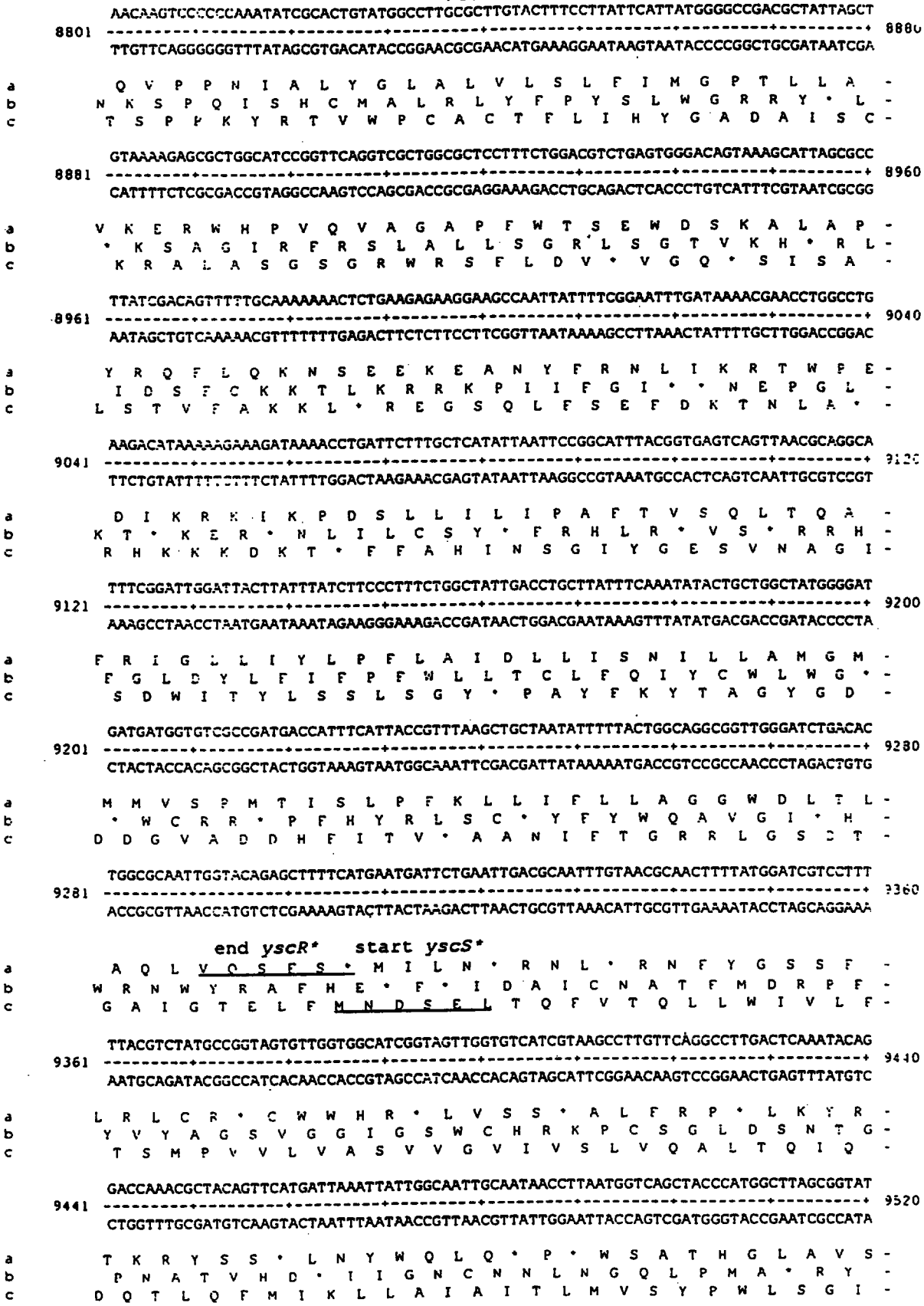


Figure 11 Sheet 13 of 19

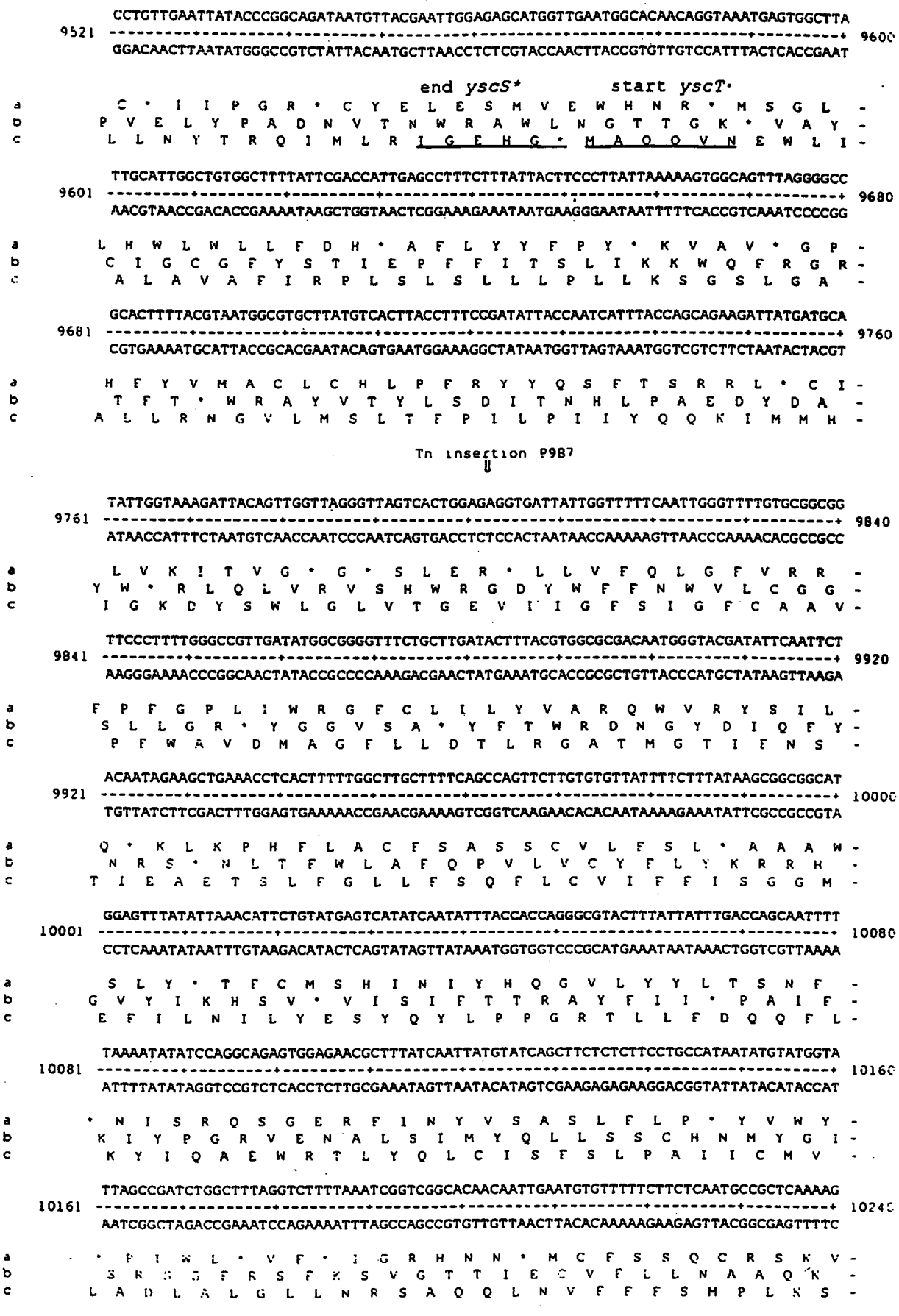


Figure 11 Sheet 14 of 19

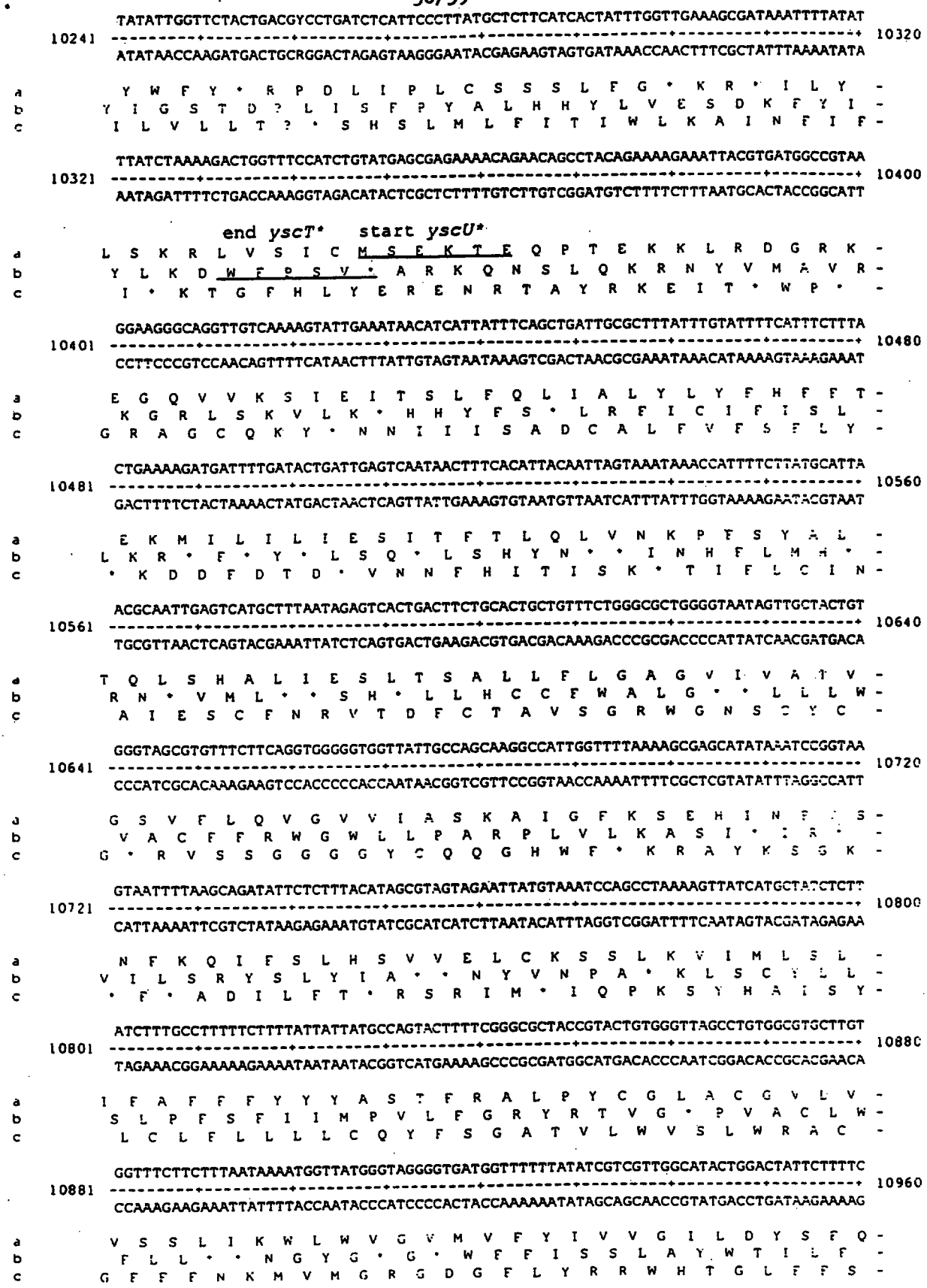


Figure 11 Sheet 15 of 19

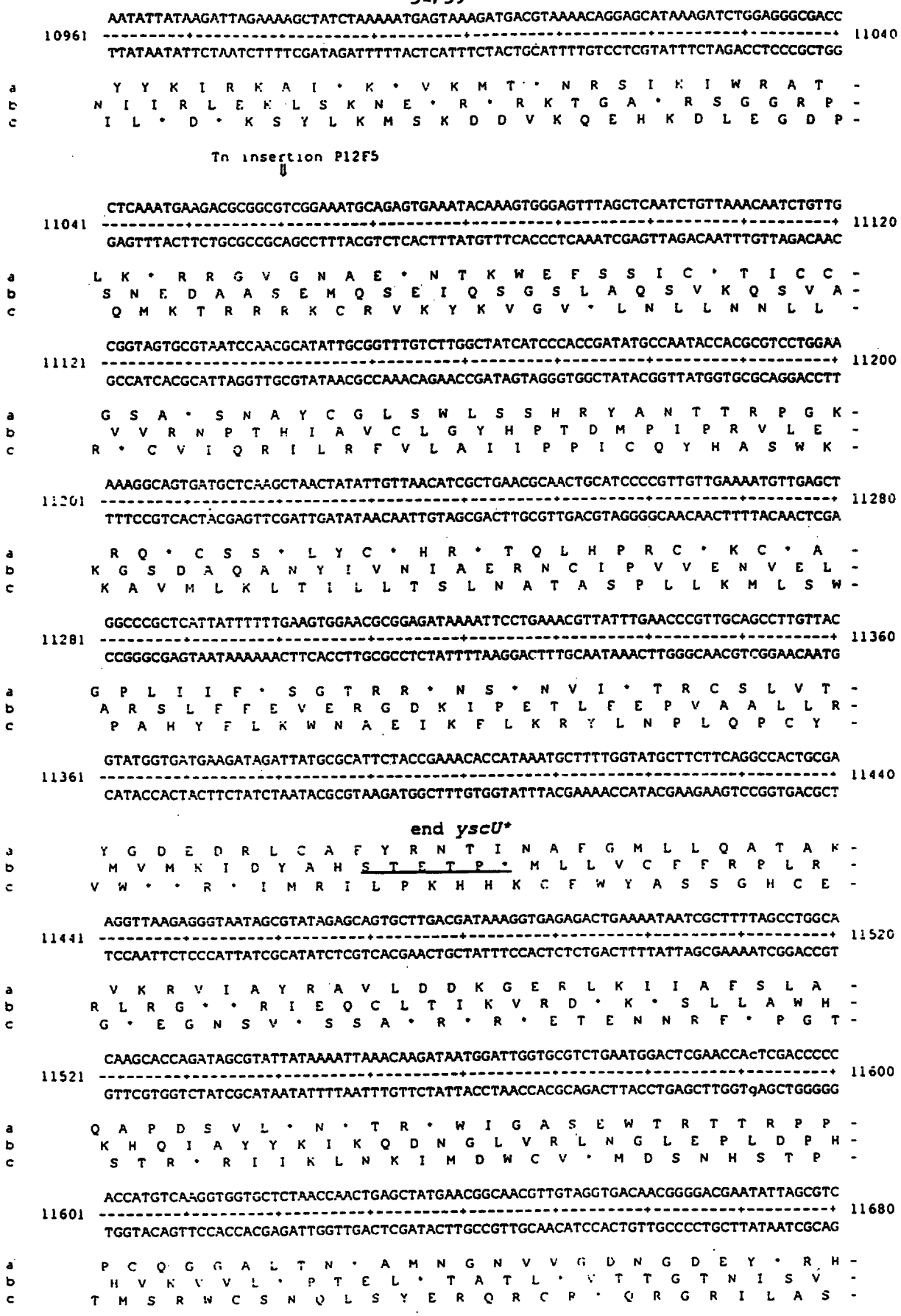


Figure 11 Sheet 16 of 19

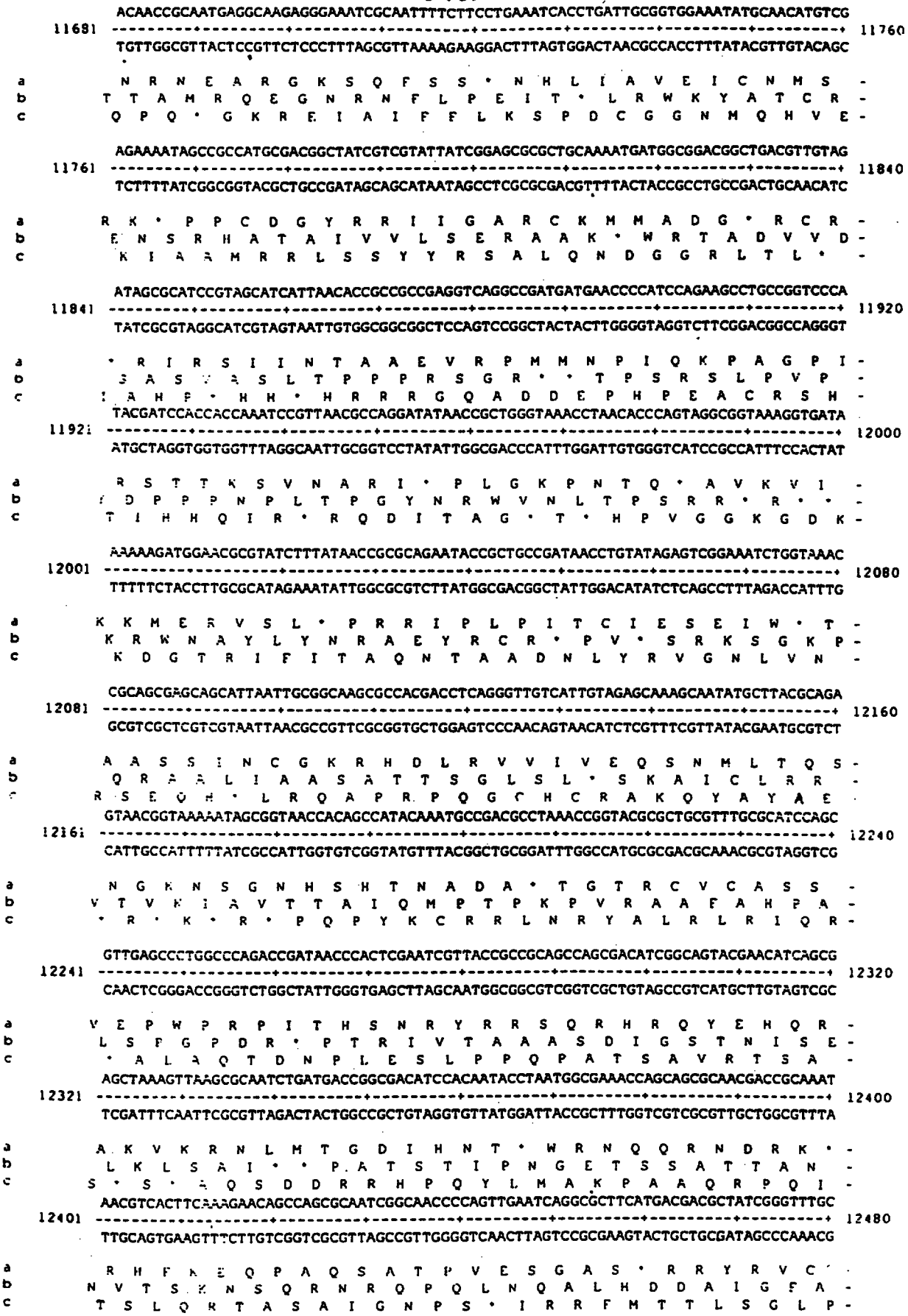


Figure 11 Sheet 17 of 19

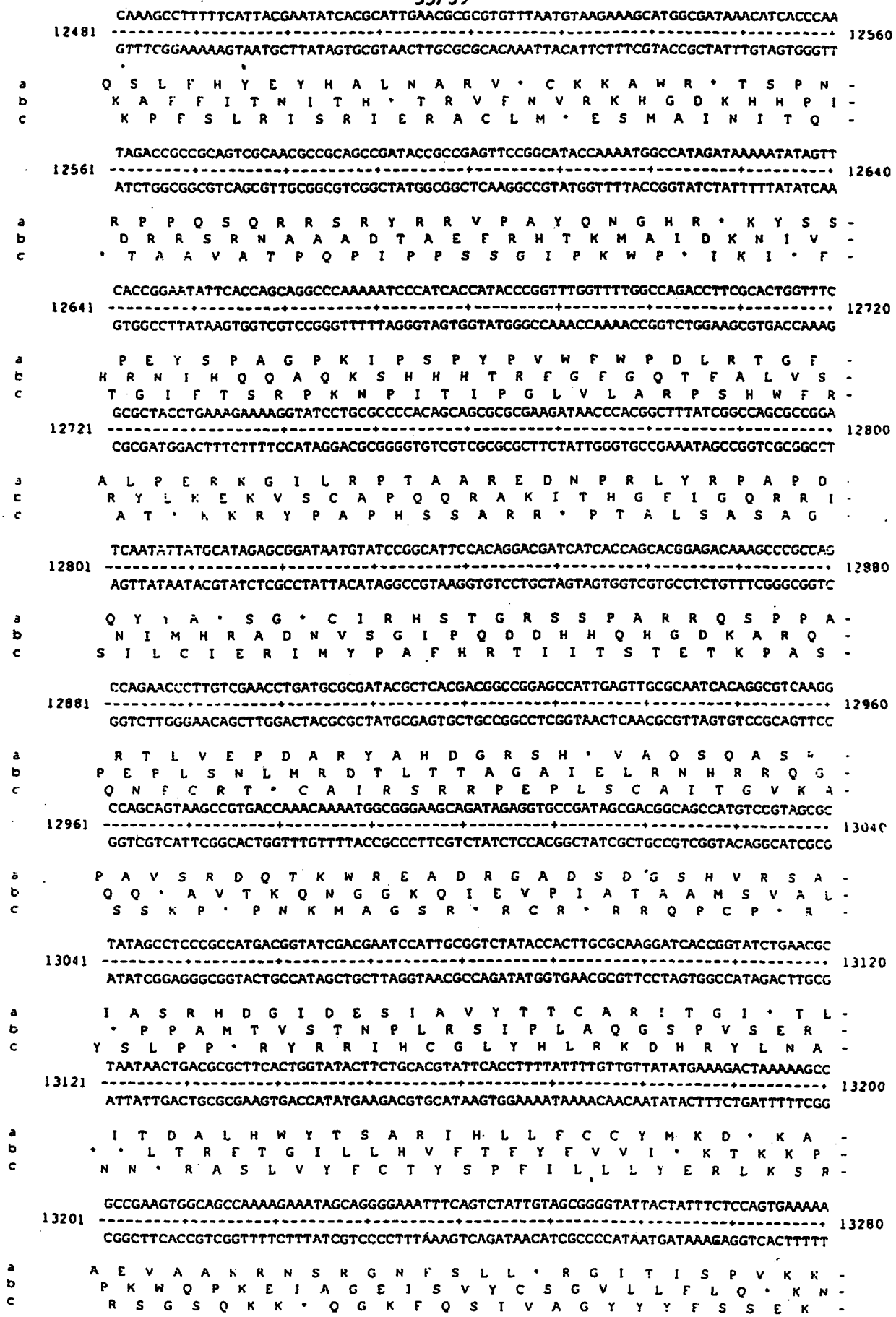


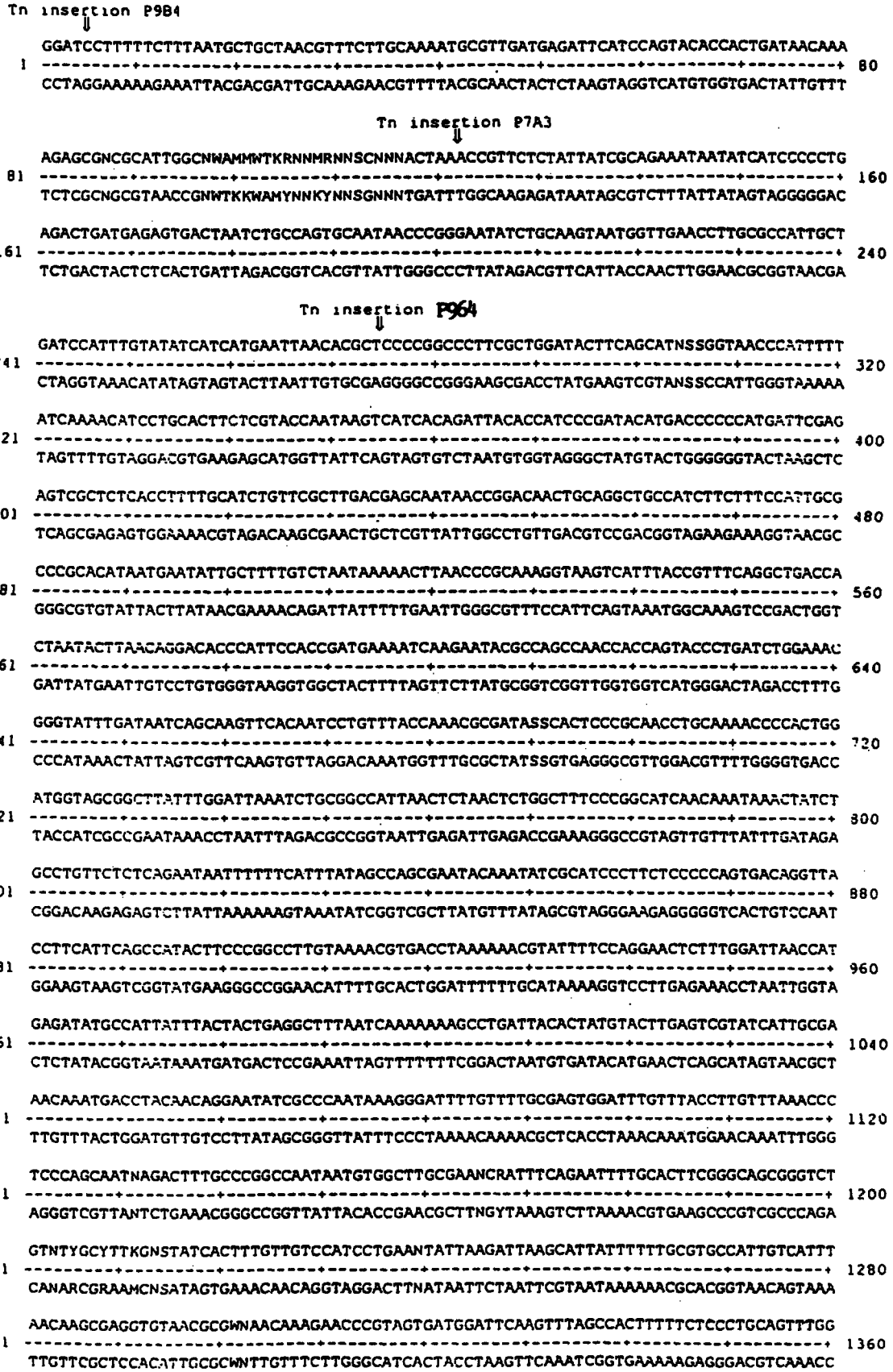
Figure 11 Sheet 18 of 19



Figure 11

SUBSTITUTE SHEET (RULE 26)

DNA sequence of VGC II cluster C



Sequence 2 Figure 12 Sheet 1 of 5

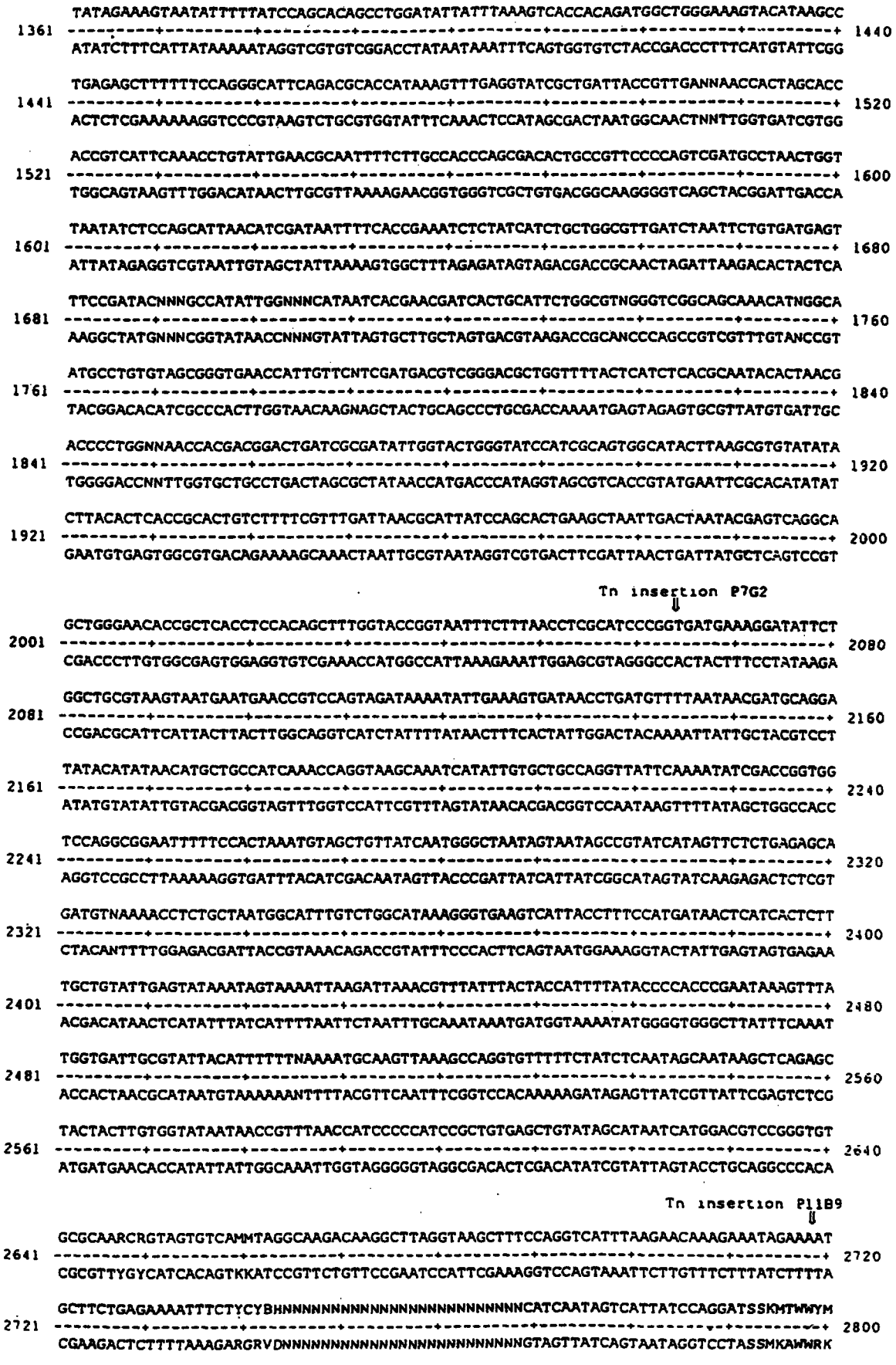


Figure 12 Sheet 2 of 5

```
      NYYKSSSCYSWKATMYYSWRWWTTAATGGAATGCCTTTTAAAACTGCCAGCATGAATCCCTCCTCAGACATAAATGGGAG
2801  ---------+---------+---------+---------+---------+---------+---------+---------+ 2880
      NRRMSSSGRSWMTAKRRSWYWWAATTACCTTACGGAAAATTTTGACGGTCGTACTTAGGGAGGAGTCTGTATTTACCCTC

      TTTCTATCAAATTCGCTCACAACCACATCCGTAAAAAGCCTGATTCACATTTATTTCGACTATACTCTTCTTGTACAATA
2881  ---------+---------+---------+---------+---------+---------+---------+---------+ 2960
      AAAGATAGTTTAAGCGAGTGTTGGTGTAGGCATTTTTCGGACTAAGTGTAAATAAAGCTGATATGAGAAGAACATGTTAT

      TCAGGATGCTGTCTACATATACCTTGTCACAGGCGATTCTATCATTCGGATTTTCCGATAAATTNMMCAATTACATTTTC
2961  ---------+---------+---------+---------+---------+---------+---------+---------+ 3040
      AGTCCTACGACAGATGTATATGGAACAGTGTCCGCTAAGATAGTAAGCCTAAAAGGCTATTTAANKKGTTAATGTAAAAG

      AGCATTGACATAAAAACTTACAATTTGNAAAATTATTTATTAAATAAACTGTTACGATGTTTTTACATCGCCATCTTATT
3041  ---------+---------+---------+---------+---------+---------+---------+---------+ 3120
      TCGTAACTGTATTTTTGAATGTTAAACNTTTTAATAAATAATTTATTTGACAATGCTACAAAAATGTAGCGGTAGAATAA

      AAAAAGTAATTGTAGTCATCGACTNGGTTATATATGAAGAAATTTATCTTCCTAATGATAACACCATCGATTAATCWWCT
3121  ---------+---------+---------+---------+---------+---------+---------+---------+ 3200
      TTTTTCATTAACATCAGTAGCTGANCCAATATATACTTCTTTAAATAGAAGGATTACTATTGTGGTAGCTAATTAGWWGA

      GATGAAACTATATGTACTGCGATAGTGATCAAGTGCCAAAGATTTTGCAACAGGCAACTGGAGGGAAGCATTATGAATTT
3201  ---------+---------+---------+---------+---------+---------+---------+---------+ 3280
      CTACTTTGATATACATGACGCTATCACTAGTTCACGGTTTCTAAAACGTTGTCCGTTGACCTCCCTTCGTAATACTTAAA

      SSTCAATCTCAAGAATACSSYSYRNNNNNNTCTTTAGTAATCAGGCTAACTTTTTTATTTTTATTAACAACAATAATTWT
3281  ---------+---------+---------+---------+---------+---------+---------+---------+ 3360
      SSAGTTAGAGTTCTTATGSSRSRYNNNNNNAGAAATCATTAGTCCGATTGAAAAAATAAAAATAATTGTTGTTATTAAWA

      TTGGCTGCTATCTGTGCTTACCGCAGCTTATATATCAATGGTTCRGAAACGGCAGCATATAATAGAGGATTTATCCGTTC
3361  ---------+---------+---------+---------+---------+---------+---------+---------+ 3440
      AACCGACGATAGACACGAATGGCGTCGAATATATAGTTACCAAGYCTTTGCCGTCGTATATTATCTCCTAAATAGGCAAG

      TATCCGAGATGAATATTGTACTAAGCAATCAACGGTTTGAAGAAGCTGAACGTGACGCTAAAAATTTAATGTATCAATGC
3441  ---------+---------+---------+---------+---------+---------+---------+---------+ 3520
      ATAGGCTCTACTTATAACATGATTCGTTAGTTGCCAAACTTCTTCGACTTGCACTGCGATTTTTAAATTACATAGTTACG

      TCATTAGCGACTGAGATTCATCATAACGATATTTTCCCTGAGGTGAGCCGGCATCTATCTGTCGGTCCTTCAAATTGCAC
3521  ---------+---------+---------+---------+---------+---------+---------+---------+ 3600
      AGTAATCGCTGACTCTAAGTAGTATTGCTATAAAAGGGACTCCACTCGGCCGTAGATAGACAGCCAGGAAGTTTAACGTG

      MGCCGACGCTNAACGGAGAGAAGCACCGTCTCTTTCTGCAGTCCTCTGATATCGATGAAAATAGCTTTCGTCGCGATAGT
3601  ---------+---------+---------+---------+---------+---------+---------+---------+ 3680
      KCGGCTGCGANTTGCCTCTCTTCGTGGCAGAGAAAGACGTCAGGAGACTATAGCTACTTTTATCGAAAGCAGCGCTATCA

         Tn insertion P3F4
                ⇓
      TTTATTCTTAATCATAAAAATGAGATTTCGTTATTATCTACTGATAACCCTTCAGATTATTCAACTCTACAGCCTTTAAC
3681  ---------+---------+---------+---------+---------+---------+---------+---------+ 3760
      AAATAAGAATTAGTATTTTTACTCTAAAGCAATAATAGATGACTATTGGGAAGTCTAATAAGTTGAGATGTCGGAAATTG

      GCGAAAAAGCTTTCCTTTATACCCAACCCATGCCGGGTTTTACTGGAGTGAACCAGAATACATAAACGGCAAAGGATGGC
3761  ---------+---------+---------+---------+---------+---------+---------+---------+ 3840
      CGCTTTTTCGAAAGGAAATATGGGTTGGGTACGGCCCAAAATGACCTCACTTGGTCTTATGTATTTGCCGTTTCCTACCG

      AACGCTTCCGTTGCGGTTGCCGATCAGGCAAGGCGTATTTTTTGAGGTGACGGTTAAACTTCCCGATCTCATTACTAAGA
3841  ---------+---------+---------+---------+---------+---------+---------+---------+ 3920
      TTGCGAAGGCAACGCCAACGGCTAGTCCGTTCCGCATAAAAAACTCCACTGCCAATTTGAAGGGCTAGAGTAATGATTCT

      GCCACCTGCCATTAGATGATAGTATTCGAGTATGGCTGGATCAAAACAACCACTTATTGCCGTTTTCATACATCCCGGCA
3921  ---------+---------+---------+---------+---------+---------+---------+---------+ 4000
      CGGTGGACGGTAATCTACTATCATAAGCTCATACCGACCTAGTTTTGTTGGTGAATAACGGCAAAAGTATGTAGGGCCGT

      AAAAATACGTACACAGTTAGAAAATGTAACGCTGCATGATGGATGGCAGCAAATTCCCGGATTTCTGATATTACGCACAA
4001  ---------+---------+---------+---------+---------+---------+---------+---------+ 4080
      TTTTTATGCATGTGTCAATCTTTTACATTGCGACGTACTACCTACCGTCGTTTAAGGGCCTAAAGACTATAATGCGTGTT

      CCTTGCATGGCCCCGGATGGAGTCTGGTTACGCTGTACCCATACGGTAATCTACATAATCGCATCTTAAAAATTATCCTT
4081  ---------+---------+---------+---------+---------+---------+---------+---------+ 4160
      GGAACGTACCGGGGCCTACCTCAGACCAATGCGACATGGGTATGCCATTAGATGTATTAGCGTAGAATTTTTAATAGGAA

      CAACAAATCCCCTTTACATTAACAGCATTGGTGTTGATGACGTCGGCTTTTTGCTGGTTACTACATCGCTCACTGGCCAA
4161  ---------+---------+---------+---------+---------+---------+---------+---------+ 4240
      GTTGTTTAGGGGAAATGTAATTGTCGTAACCACAACTACTGCAGCCGAAAAACGACCAATGATGTAGCGAGTGACCGGTT
```

Figure 12 Sheet 3 of 5

```
      ACCGTTATGGCGTTTTGTCGATGTCATTAATAAAACCGCAACTGCACCGCTGAGCACACGTTTACCAGCACAACGACTGG
4241  ---------+---------+---------+---------+---------+---------+---------+---------+ 4320
      TGGCAATACCGCAAAACAGCTACAGTAATTATTTTGGCGTTGACGTGGCGACTCGTGTGCAAATGGTCGTGTTGCTGACC

      ATGAATTAGATAGTATTGCCGGTGCTTTTAACCAACTGCTTGATACTCTACAAGTCCAATACGACAATCTGGAAAACAAA
4321  ---------+---------+---------+---------+---------+---------+---------+---------+ 4400
      TACTTAATCTATCATAACGGCCACGAAAATTGGTTGACGAACTATGAGATGTTCAGGTTATGCTGTTAGACCTTTTGTTT

      GTCGCAGACGCACCCAGGCGCTAAATGAAGCAAAAAAACGCGCTGAGCNAGCTAACAAACGTAAAAGCATTCATCTTACG
4401  ---------+---------+---------+---------+---------+---------+---------+---------+ 4480
      CAGCGTCTGCGTGGGTCCGCGATTTACTTCGTTTTTTTGCGCGACTCGNTCGATTGTTTGCATTTTCGTAAGTAGAATGC

      GTAATAAGTCATGAGTTACGTACTCCGATGAATGGCGTACTCGGTGCAATTGAATTATTACAAACCACCCCTTTAAACAT
4481  ---------+---------+---------+---------+---------+---------+---------+---------+ 4560
      CATTATTCAGTACTCAATGCATGAGGCTACTTACCGCATGAGCCACGTTAACTTAATAATGTTTGGTGGGGAAATTTGTA

      AGAGCAACAAGGATTAGCTGATACCGCCAGAAATTGTACACTGTCTTTGTTAGCTATTATTAATAATCTGCTGGATTTTT
4561  ---------+---------+---------+---------+---------+---------+---------+---------+ 4640
      TCTCGTTGTTCCTAATCGACTATGGCGGTCTTTAACATGTGACAGAAACAATCGATAATAATTATTAGACGACCTAAAAA

      CACGCATCGAGTCTGGTCATTTCACATTACATATGGAAGAAACAGCGTTACTGCCGTTACTGGACCAGGCAATGCAAACC
4641  ---------+---------+---------+---------+---------+---------+---------+---------+ 4720
      GTGCGTAGCTCAGACCAGTAAAGTGTAATGTATACCTTCTTTGTCGCAATGACGGCAATGACCTGGTCCGTTACGTTTGG

      ATCCAGGGGCCAGCGCNAAAGCAAAAAACTGTCATTACGTACTTTTGTCGGTCAACATGTCCCTCTCTATTTTCATACCG
4721  ---------+---------+---------+---------+---------+---------+---------+---------+ 4800
      TAGGTCCCCGGTCGCGNTTTCGTTTTTTGACAGTAATGCATGAAAACAGCCAGTTGTACAGGGAGAGATAAAAGTATGGC

      ACAGTATCCGTTTACNNCAAATTTTGGTTAATTTACTCGGGAACGCGGTAAAATTTACCGAAACCGGAGGATACGTCTGA
4801  ---------+---------+---------+---------+---------+---------+---------+---------+ 4880
      TGTCATAGGCAAATGNNGTTTAAAACCAATTAAATGAGCCCTTGCGCCATTTTAAATGGCTTTGGCCTCCTATGCAGACT

      CGGTCAAGCGTCATGAGGAACAATTAATATTTCTGGTTAGCGATAGCGGTAAAGGGATTGAAATACAGCAGCAGTCTCAA
4881  ---------+---------+---------+---------+---------+---------+---------+---------+ 4960
      GCCAGTTCGCAGTACTCCTTGTTAATTATAAAGACCAATCGCTATCGCCATTTCCCTAACTTTATGTCGTCGTCAGAGTT

      ATCTTTACTGCTTTTTATCAAGCAGACACAAATTCGCAAGGTACAGGAATTGGACTGACTATTGCGTCAAGCCTGGCTAA
4961  ---------+---------+---------+---------+---------+---------+---------+---------+ 5040
      TAGAAATGACGAAAAATAGTTCGTCTGTGTTTAAGCGTTCCATGTCCTTAACCTGACTGATAACGCAGTTCGGACCGATT

      AATGATGGGCGGTAATCTGACACTAAAAAGTGTCCCCGGGGTTGGAACCTGTGTCTCGCTAGTATTACCCTTACAAGAAT
5041  ---------+---------+---------+---------+---------+---------+---------+---------+ 5120
      TTACTACCCGCCATTAGACTGTGATTTTTCACAGGGGCCCCAACCTTGGACACAGAGCGATCATAATGGGAATGTTCTTA

                                  Tn insertion
                                        ⇩
      ACCAGCCGCCTCAACCAATTAAAGGGACGCTGTCAGNNNCCGTTCTGCCTGCATCGGCAACTGGCTTGCTGGGGAATACG
5121  ---------+---------+---------+---------+---------+---------+---------+---------+ 5200
      TGGTCGGCGGAGTTGGTTAATTTCCCTGCGACAGTCNNNGGCAAGACGGACGTAGCCGTTGACCGAACGACCCCTTATGC

      CGGTGAACCACCCCACCAGCAAAATGCGCTTCTCAANNCNAGAGCTTTTGTATTTCTCCGGAAAACTCTACGACCTGGCG
5201  ---------+---------+---------+---------+---------+---------+---------+---------+ 5280
      GCCACTTGGTGGGGTGGTCGTTTTACGCGAAGAGTTNNGNTCTCGAAAACATAAAGAGGCCTTTTGAGATGCTGGACCGC

      CAACAGTTAATATTGTGTACACCAAATATGCCAGTAATAAATAATTTGTTACCACCCTGGCAGTTGCAGATTCTTTTGGT
5281  ---------+---------+---------+---------+---------+---------+---------+---------+ 5360
      GTTGTCAATTATAACACATGTGGTTTATACGGTCATTATTTATTAAACAATGGTGGGACCGTCAACGTCTAAGAAAACCA

      TGATGATGCCGATATTAATCGGGATATCATCGGCAAAATGCTTGTCAGCCTGGGCCAACACGTCACTATTGCCGCCAGTA
5361  ---------+---------+---------+---------+---------+---------+---------+---------+ 5440
      ACTACTACGGCTATAATTAGCCCTATAGTAGCCGTTTTACGAACAGTCGGACCCGGTTGTGCAGTGATAACGGCGGTCAT

      GTAACGAGGCTCTGACTTTATCACAACAGCAGCGATTCGATTTAGTACTGATTGACATTAGAATGCCAGAAATAGATGGT
5441  ---------+---------+---------+---------+---------+---------+---------+---------+ 5520
      CATTGCTCCGAGACTGAAATAGTGTTGTCGTCGCTAAGCTAAATCATGACTAACTGTAATCTTACGGTCTTTATCTACCA

      ATTGAATGTGTACGATTATGGCATGATGAGCCGAATAATTTAGATCCTGACTGCATGTTTGTGGCACTATCCGCTAGCGT
5521  ---------+---------+---------+---------+---------+---------+---------+---------+ 5600
      TAACTTACACATGCTAATACCGTACTACTCGGCTTATTAAATCTAGGACTGACGTACAAACACCGTGATAGGCGATCGCA

      ASCVNMAGAWRWTMWTCRTYGTDDAAAAAAWRDGRKDHWTCATHAYANNTTACAAAACCAGTGACATTGGCTACCTTAGC
5601  ---------+---------+---------+---------+---------+---------+---------+---------+ 5680
      TSGBNKTCTWYWAKWAGYARCAHHTTTTTTWYHCYMHDWAGTADTRTNNAATGTTTTGGTCACTGTAACCGATGGAATCG
```

Figure 12 Sheet 4 of 5

```
      TCGCTACATCAGTATTGCCGCAGAATACCAACTTTTACGAAATATAGAGCTACAGGAGCAGGATCC
5681  ---------+---------+---------+---------+---------+---------+------ 5746
      AGCGATGTAGTCATAACGGCGTCTTATGGTTGAAAATGCTTTATATCTCGATGTCCTCGTCCTAGG
```

Figure 12

Sheet 5 of 5

Groisman *et al* (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1033-1037 describes the molecular, functional and evolutionary analysis of sequences specific to *Salmonella*.

Hensel *et al* (1995) *Science* **269**, 400-403 describes the simultaneous identification of bacterial virulence genes by negative selection. This paper was published in July 1995.

Slauch *et al* (1994) *Methods in Enzymology* **235**, 481-492 describes *in vivo* expression technology for selection of bacterial genes specifically induced in host tissues.

Pascopella *et al* (1994) *Inf. Immun.* **62**, 1313-1319 describes the use of *in vivo* complementation in *Mycobacterium tuberculosis* to identify a genomic fragment associated with virulence.

US 5,397,697 describes the identification of plant-responsive genes of bacteria.

Some virulence genes are already known for pathogenic microorganisms such as *Escherichia coli*, *Salmonella typhimurium*, *Salmonella typhi*, *Vibrio cholerae*, *Clostridium botulinum*, *Yersinia pestis*, *Shigella flexneri* and *Listeria monocytogenes* but in all cases only a relatively small number of the total have been identified.

The disease which *Salmonella typhimurium* causes in mice provides a good experimental model of typhoid fever (Carter & Collins, 1974). Approximately forty two genes affecting *Salmonella* virulence have been identified to date (Groisman & Ochman, 1994). These represent approximately one third of the total number of predicted virulence genes

(Groisman and Saier, 1990).

The object of the present invention is to identify genes involved in the adaptation of a microorganism to its environment, particularly to identify further virulence genes in pathogenic microorganisms, with increased efficiency. A further object is to reduce the number of experimental animals used in identifying virulence genes. Still further objects of the invention provide vaccines, and methods for screening for drugs which reduce virulence.

Summary of the invention

A first aspect of the invention provides a method for identifying a microorganism having a reduced adaptation to a particular environment comprising the steps of:

CLAIMS

1. A method for identifying a microorganism having a reduced adaptation to a particular environment comprising the steps of:

(1) providing a plurality of microorganisms each of which is independently mutated by the insertional inactivation of a gene with a nucleic acid comprising a unique marker sequence so that each mutant contains a different marker sequence, or clones of the said microorganism;

(2) providing individually a stored sample of each mutant produced by step (1) and providing individually stored nucleic acid comprising the unique marker sequence from each individual mutant;

(3) introducing a plurality of mutants produced by step (1) into the said particular environment and allowing those microorganisms which are able to do so to grow in the said environment;

(4) retrieving microorganisms from the said environment or a selected part thereof and isolating the nucleic acid from the retrieved microorganisms;

(5) comparing any marker sequences in the nucleic acid isolated in step (4) to the unique marker sequence of each individual mutant stored as in step (2); and

(6) selecting an individual mutant which does not contain any of the marker sequences as isolated in step (4).

2. A method according to Claim 1 wherein the plurality of microorganisms as defined in step (1) is produced from a plurality of microorganisms, each of which comprises a nucleic acid comprising a unique marker sequence, by changing their condition from a first given condition to a second given condition wherein (a) in the first given condition the said nucleic acid comprising a unique marker is maintained episomally and (b) in the second given condition the said nucleic acid comprising a



unique marker sequence insertionally inactivates a gene.

3. A method according to Claims 1 or 2 further comprising the steps:

(1A) removing auxotrophs from the plurality of mutants produced in step (1); or

(6A) determining whether the mutant selected in step (6) is an auxotroph; or

both (1A) and (6A).

4. A method of identifying a gene which allows a microorganism to adapt to a particular environment, the method comprising the method of any one of Claims 1 to 3 followed by the step:

(7) isolating the insertionally-inactivated gene from the individual mutant selected in step (6).

5. A method according to Claim 4 further comprising the step:

(8) isolating from a wild-type microorganism the corresponding wild-type gene using the insertionally-inactivated gene isolated in step (7) as a probe.

6. A method according to any one of Claims 1 to 5 wherein the particular environment is a differentiated multicellular organism.

7. A method according to Claim 6 wherein the multicellular organism is a plant.

8. A method according to Claim 6 wherein the multicellular organism is a non-human animal.

9. A method according to Claim 8 wherein the animal is a mouse, rat,

* C L M S *

rabbit, dog or monkey.

10. A method according to Claim 9 wherein the animal is a mouse.

11. A method according to any one of Claims 6 to 10 wherein in step (4) the microorganisms are retrieved from the said environment at a site remote from the site of introduction in step (3).

12. A method according to any one of Claims 8 to 10 wherein in step (3) the microorganism is introduced orally or intraperitoneally.

13. A method according to Claim 12 when dependent on Claims 8 or 9 wherein in step (4) the microorganisms are retrieved from the spleen.

14. A method according to any one of the preceding claims wherein the microorganism is a bacterium.

15. A method according to any one of Claims 1 to 13 wherein the microorganism is a fungus.

16. A method according to Claim 7 wherein the microorganism is a bacterium pathogenic to plants.

17. A method according to Claim 7 wherein the microorganism is a fungus pathogenic to plants.

18. A method according to any one of Claims 8 to 10 wherein the microorganism is a bacterium pathogenic to animals.

19. A method according to any one of Claims 8 to 10 wherein the



microorganism is a fungus pathogenic to animals.

20. A method according to Claim 18 wherein the bacterium is any one of *Bordetella pertussis*, *Campylobacter jejuni*, *Clostridium botulinum*, *Escherichia coli*, *Haemophilus ducreyi*, *Haemophilus influenzae*, *Helicobacter pylori*, *Klebsiella pneumoniae*, *Legionella pneumophila*, *Listeria* spp., *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Pseudomonas* spp., *Salmonella* spp., *Shigella* spp., *Staphylococcus aureus*, *Streptococcus pyogenes*, *Streptococcus pneumoniae*, *Vibrio* spp., and *Yersinia pestis*.

21. A method according to Claim 19 wherein the fungus is any one of *Aspergillus* spp., *Cryptococcus neoformans* and *Histoplasma capsulatum*.

22. A method according to any one of the preceding claims wherein in step (1) the gene is insertionally inactivated using a transposon or transposon like element or other DNA sequence carrying a unique marker sequence.

23. A method according to any one of the preceding claims wherein in step (1) each different marker sequence is flanked on either side by sequences common to each said nucleic acid.

24. A method according to Claim 23 wherein in step (2) the nucleic acid comprising the unique marker is isolated using DNA amplification techniques and oligonucleotide primers which hybridise to the said common sequences.

25. A method according to Claim 23 or 24 wherein in step (4) the nucleic acid comprising a plurality of said marker sequences is isolated using DNA amplification techniques and oligonucleotide primers which



hybridise to the said common sequences.

26. A method according to Claim 1 or 4 wherein the particular environment is not a human body.

5

AMENDED SHEET

* C L M S *

9703322-9

RTMPS copy

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/04, 1/68, C12N 15/10 — A2

(11) International Publication Number: WO 96/17951

(43) International Publication Date: 13 June 1996 (13.06.96)

(21) International Application Number: PCT/GB95/02875

(22) International Filing Date: 11 December 1995 (11.12.95)

(30) Priority Data:

9424921.6	9 December 1994 (09.12.94)	GB
9501881.8	31 January 1995 (31.01.95)	GB
9509239.1	5 May 1995 (05.05.95)	GB

(81) Designated States: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, LS, MW, SD, SZ, UG).

Published
Without international search report and to be republished upon receipt of that report.

(71) Applicant *(for all designated States except US)*: RPMS TECHNOLOGY LIMITED [GB/GB]; Commonwealth Building, Du Cane Road, London W12 0NN (GB). see PF24 of 5/11/98

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: HOLDEN, David, William [GB/GB]; Dept. of Infectious Diseases and Bacteriology, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London W12 0NN (GB).

(74) Agent: BASSETT, Richard; Eric Potter Clarkson, St. Mary's Court, St. Mary's Gate, Nottingham NG1 1LE (GB).

(54) Title: IDENTIFICATION OF GENES

(57) Abstract

A method for identifying a microorganism having a reduced adaptation to a particular environment comprising the steps of: (1) providing a plurality of microorganisms each of which is independently mutated by the insertional inactivation of a gene with a nucleic acid comprising a unique marker sequence so that each mutant contains a different marker sequence, or clones of the said microorganism; (2) providing individually a stored sample of each mutant produced by step (1) and providing individually stored nucleic acid comprising the unique marker sequence from each individual mutant; (3) introducing a plurality of mutants produced by step (1) into the said particular environment and allowing those microorganisms which are able to do so to grow in the said environment; (4) retrieving microorganisms from the said environment or a selected part thereof and isolating the nucleic acid from the retrieved microorganisms; (5) comparing any marker sequences in the nucleic acid isolated in step (4) to the unique marker sequence of each individual mutant stored as in step (2); and (6) selecting an individual mutant which does not contain any of the marker sequences as isolated in step (4).