Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2302861 A1 2001/10/10

(21) 2 302 861

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(22) Date de dépôt/Filing Date: 2000/04/10

(41) Mise à la disp. pub./Open to Public Insp.: 2001/10/10

(51) Cl.Int.^{7}/Int.Cl.^{7} C12N 15/11, A61K 38/16, A61K 39/00, C12N 5/10, C12Q 1/00, C12P 21/00, C07K 16/00, G01N 33/566

(71) Demandeurs/Applicants:
CVITKOVITCH, DENNIS G., CA;
LAU, PETER C. Y., CA;
LI, YUNG HUA, CA

(72) Inventeurs/Inventors:
CVITKOVITCH, DENNIS G., CA;
LAU, PETER C. Y., CA;
LI, YUNG HUA, CA

(74) Agent: DEETH WILLIAMS WALL LLP

(54) Titre : MOLECULES D'ACIDE NUCLEIQUE CODANT POUR DES SEQUENCES-SIGNAL ET METHODE DE TRAITEMENT DES CARIES

(54) Title: SIGNAL PEPTIDES NUCLEIC ACID MOLECULES AND METHODS FOR TREATMENT OF CARIES

Schematic layout of the arrangement of the genetic locus encoding the signal peptide precursor, the histidine kinase and the response regulator. Note that this arrangement is different from other loci in related streptococci for the following reasons: a) The comC gene is transcribed from its own promoter alone, unlike the genes thus far described in other streptococci that are arranged in an operon-like cluster with the comC/DE genes being transcribed from a single promoter.
b) The comC gene is separated from the *comD* gene by 148 nucleotides.

Streptococcus mutans
***ComCDE* Operon**



Abstract

The invention relates to a compound that competitively inhibits binding of CSP to *S.mutans* histidine kinase. The compound is preferably a peptide or an antibody. The compound is preferably a derivative of [SEQ ID NO:2], a fragment of [SEQ ID NO:2] or a derivative of a fragment of [SEQ ID NO:2].

SIGNAL PEPTIDES, NUCLEIC ACID MOLECULES AND METHODS FOR TREATMENT OF CARIES

FIELD OF THE INVENTION

The invention relates to compounds and methods that inhibit or disrupt microbial biofilms involved in infections in man and animals and in biofouling of surfaces susceptible to microbial accumulation.

BACKGROUND OF INVENTION

Bacteria often attach and accumulate on surfaces, enabling them to resist removal and killing by mechanical and chemical means. This can result in persistent and chronic infections and fouling of devices that are in contact with liquids containing the colonizing bacteria.

Controlling bacterial biofilms is desirable for almost every human enterprise in which solid surfaces are introduced into non-sterile aqueous environments. US Patent No. 6,024,958 describes peptides that attempt to control biofilm formation by preventing bacterial adherence to teeth. In addition to occurrence in dental caries, medical examples of biofilm growth include cases involving indwelling medical devices, joint implants, prostatitis, endocarditis, and respiratory infections. In fact, the Centers for Disease Control and Prevention (CDC; Atlanta, GA) estimate that 65% of human bacterial infections involve biofilms. Non-medical examples of biofilm colonization are water and beverage lines, cooling towers, radiators, aquaculture contamination, submerged pumps and impellers, hulls of commercial, fishing and military vessels and literally every situation where biofouling occurs. The potential benefits of basic research focused at biofilm physiology and genetics with the ultimate goal of controlling surface-mediated microbial growth are limitless.

Interest in the study of biofilm-grown cells has increased partly because biofilm growth provides a microenvironment for cells to exist in a physical and physiological state that can increase their resistance to antimicrobial compounds and mechanical forces (reviewed in Costerton and Lewandowski, Adv Dent Res, 11:192-195). Growth in biofilms can also facilitate the transfer of genetic information between different species (Christensen *et al.* Appl Environ Microbiol, 64:2247-2255). Recent evidence suggests that biofilm-grown cells may display a dramatically different phenotype when compared

with their siblings grown in liquid culture. In some, this altered physiological state has been shown to result from gene activation initiated by contact with surfaces (Finlay and Falkow. Microbiol Molec Rev, 61:136-169) or from signal molecules produced by the bacteria allowing them to sense the cell density (quorum sensing) (Davies *et al.* Appl Environ Microbiol, 61:860-867). Biofilms may also act as 'genotypic reservoirs', allowing persistence, transfer and selection of genetic elements conferring resistance to antimicrobial compounds.

Streptococcus mutans is the principal etiological agent of dental caries in humans. None of the known types of *S. mutans* antibiotics has satisfactorily controlled caries. There is a need to identify new ways to control *S. mutans* induced caries.

SUMMARY OF THE INVENTION

Figure 1

Schematic layout of the arrangement of the genetic locus encoding the signal peptide precursor (ComC), the histidine kinase (ComD) and the response regulator (ComE). Note that this arrangement is different from other loci in related streptococci for the following reasons: a) The comC gene is transcribed from its own unique promoter, unlike the genes thus far described in other streptococci that are arranged in an operon-like cluster with the comC/DE genes being transcribed from a single promoter. b) The comC gene is separated by 148 nucleotides from the comD gene.

Figure 2

Shows the nucleic acid molecule that is [SEQ ID NO:1]. In a preferred embodiment, the figure shows CSP (competence signal peptide). Nucleotide sequence of the locus. Figure 2 also shows histidine kinase sequences and response regulator sequences.

Figure 3

Sequence of the deduced amino acid sequence of the signal peptide, histidine kinase, and response regulator.

Figure 4

The deduced amino acid sequence of the signal peptide precursor and its predicted cleavage site, following the glycine-glycine signal.

Figure 5

Shows the peptide that is [SEQ ID NO:2]. The synthetic signal peptide that is effective at inducing competence, biofilm formation and acid tolerance in *Streptococcus mutans*.

Figure 6

The natural activity of the signal/receptor system functioning in vitro in model biofilms as determined by the ability of various strains of *S. mutans* to accept donor plasmid DNA conferring erythromycin resistance.

Figure 7

Table illustrating the effect of synthetic peptide on genetic competence in S. mutans cells.

Figure 8

List of the primers used to amplify the genes or internal regions of the target genes by polymerase chain reaction (PCR) for subsequent sequencing or inactivation.

Figure 9

ComCDE local region.

Figure 10

The comX DNA sequence, protein sequence, and the comX gene local region with 100bp included both upstream and downstream (promoter is upstream).

Figure 11

The comA and comB nucleotide and amino acid sequences. ComA and ComB are the components of the CSP exporter.

DETAILED DESCRIPTION OF THE INVENTION

In some Gram-positive bacteria (including *Streptococcus mutans*), when a specific histidine kinase receptor located in the cell membrane is disrupted, the cells become ineffective at developing a biofilm. The cells growing in this biofilm environment use a small peptide signal molecule to activate the receptor in surrounding cells, thereby communicating the message to form a biofilm. This same signal peptide and histidine

kinase are also involved in the induction of genetic competence, the cell's ability to take up and incorporate DNA from its extracellular environment, as well as that of acid tolerance, the cell's ability to survive pH levels as low as pH 3.0. A mechanism that blocks the signal molecule from activating the histidine kinase receptor molecule provides a novel method for controlling microbial biofilms, either alone or in combination with chemical or physical means.

We have identified a genetic locus in *S. mutans* consisting of three genes that encode: 1) a peptide precursor that is processed during export into a secreted 21-amino acid peptide (CSP); 2) a histidine kinase that acts as a cell surface receptor activated by the peptide; 3) a response regulator that activates a number of other genes involved in genetic competence, biofilm formation, and acid tolerance of *S. mutans*. These properties have been attributed to the bacterium's ability to cause dental caries. Inactivation of any of these three genes or impairment of interaction or activity of any of their encoded proteins will disrupt the bacterium's ability to take up foreign DNA, form biofilms, and tolerate acidic pH.

Streptococcus mutans is a resident of the biofilm environment of dental plaque, a matrix of bacteria and extracellular material that adheres to the tooth surface. Under appropriate environmental conditions populations of *S. mutans* and the pH of the surrounding plaque will drop. *S. mutans*, being among the most acid tolerant organisms residing in dental plaque, will increase it numbers in this acidic environment and eventually become a dominant member of the plaque community. This situation eventually leads to dissolution of the tooth enamel, resulting in the development of dental caries. We control the accumulation and acid tolerance of this bacterium to make it less able to cause caries. We accomplish this by using inhibitors of an extracellular signal peptide that promotes the expression of genes involved in *S. mutans* biofilm formation and acid tolerance. The invention includes compounds that inhibit the action of the peptide. These inhibitors can include peptides, antibodies, or other agents that specifically inhibit the activation of the histidine kinase and the family of genes activated as a result of the histidine kinase activation by the signal molecule. Inhibitors include: modified structures of the peptide where amino acids are removed from the N- and/or COOH terminal of the peptide and/or substitutions of internal amino acid residues. We delete, one, two to 5, 6 to 10 and 10 to 15 amino acids from the peptide (for example at either terminal) and measure competitive inhibition of signal peptide binding to histidine

kinase (1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more amino acids are deleted and inhibition measured). Inhibitors also include antibodies raised against the 21-amino acid CSP alone or coupled to a larger molecule to increase immunogenicity.

We also test inhibitors described in (Barrett et al. Proc. Natl. Acad. Sci USA 95:5317-5322) and measure competitive inhibition of signal peptide binding to histidine kinase.

In addition to identifying the genes encoding this signaling/sensing system, we have identified and chemically synthesized a 21-amino acid peptide that promotes biofilm formation and acid tolerance of *S. mutans*. A survey of the literature and genome databases reveals that genes similar to this signal-receptor system are present in most Gram-positive bacteria, and therefore an inhibitor, or family of related inhibitors may be effective at inhibiting biofilm formation among a large group of bacteria.

The invention treats or prevents dental caries by addition of compounds that inhibit the stimulatory action of the 21-amino acid peptide on biofilm formation and acid tolerance of *S. mutans*. This is accomplished by delivery of these compounds to the biofilm and/or to incorporate these inhibitors into materials to control growth on surfaces. This includes delivery by topical application, alone or in combination with other compounds including toothpaste or mouthwash.

Streptococcus mutans is also implicated in causing infective endocarditis. Inhibitors of biofilm formation, and hence aggregation are useful in the treatment of these bacterial infections as well.

Identification and characterization of Competence Signal Peptide (CSP), Histidine Kinase (HK) and Response Regulator (RR)

Competence Signal Peptide

The invention includes an isolated CSP from *S. mutans*. The invention also includes a recombinant isolated CSP peptide produced by a cell including a nucleic acid molecule encoding CSP operably linked to a promoter. The invention also includes an isolated nucleic acid molecule encoding a CSP. The peptide we work with is preferably chemically synthesized.

The invention includes CSP-encoding nucleic acid molecules and molecules having sequence identity or which hybridize to the CSP-encoding sequence and which encode a peptide having CSP activity (preferred percentages for sequence identity are described below) as well as vectors including these molecules. The invention also

includes CSP or peptides having sequence identity (preferred percentages described below) or which have CSP activity. The nucleic acid molecules and peptides of the invention may be from *S. mutans* and they may be isolated from a native source, synthetic or recombinant. The invention includes CSP or peptides having sequence identity which have CSP activity, as prepared by the processes described in this application.

Histidine Kinase

The invention includes an isolated HK from *S. mutans*. The invention also includes a recombinant isolated HK polypeptide produced by a cell including a nucleic acid molecule encoding HK operably linked to a promoter. The invention also includes an isolated nucleic acid molecule encoding a HK polypeptide.

The invention includes HK-encoding nucleic acid molecules and molecules having sequence identity or which hybridize to the HK-encoding sequence and which encode a protein having HK activity (preferred percentages for sequence identity are described below) as well as vectors including these molecules. The invention also includes HK or polypeptides having sequence identity (preferred percentages described below) or which have HK activity. The nucleic acid molecules and polypeptides of the invention may be from *S. mutans* and they may be isolated from a native source, synthetic or recombinant. The invention includes HK or polypeptides having sequence identity which have HK activity, as prepared by the processes described in this application.

Response Regulator

The invention includes an isolated RR from *S. mutans*. The invention also includes a recombinant isolated RR polypeptide produced by a cell including a nucleic acid molecule encoding RR operably linked to a promoter. The invention also includes an isolated nucleic acid molecule encoding a RR polypeptide.

The invention includes RR-encoding nucleic acid molecules and molecules having sequence identity or which hybridize to the RR-encoding sequence and which encode a polypeptide having RR activity (preferred percentages for sequence identity are described below) as well as vectors including these molecules. The invention also includes RR or polypeptides having sequence identity (preferred percentages described below) or which have RR activity. The nucleic acid molecules and polypeptides of the

invention may be from *S. mutans* and they may be isolated from a native source, synthetic or recombinant. The invention includes RR or polypeptides having sequence identity which have RR activity, as prepared by the processes described in this application.

The comA and comB nucleotide and amino acid sequences are also aspects of the invention. ComA and ComB are components of the CSP exporter. The discussion of variants, sequence identity etc. for CSP, HK, RR applies to both the full sequences shown in the figures as well as bracketed portions of sequences (coding regions). The peptides and polypeptides may be natural, recombinantly produced or synthetic.

Functionally equivalent nucleic acid molecules

The invention includes nucleic acid molecules that are functional equivalents of all or part of the CSP sequence in [SEQ ID NO:1]. (A nucleic acid molecule may also be referred to as a DNA sequence or nucleotide sequence in this application. All these terms have the same meaning as nucleic acid molecule) Functionally equivalent nucleic acid molecules are DNA and RNA (such as genomic DNA, complementary DNA, synthetic DNA, and messenger RNA molecules) that encode peptides having the same or similar CSP activity as the CSP peptide shown in [SEQ ID NO:2]. Functionally equivalent nucleic acid molecules can encode peptides that contain a region having sequence identity to a region of a CSP peptide or more preferably to the entire CSP peptide. Identity is calculated according to methods known in the art. The ClustalW program (preferably using default parameters) [Thompson, JD et al., Nucleic Acid Res. 22:4673-4680.], described below, is most preferred. For example, if a nucleic acid molecule (called "Sequence A") has 90% identity to a portion of the nucleic acid molecule in [SEQ ID NO:1], then Sequence A will preferably be identical to the referenced portion of the nucleic acid molecule in [SEQ ID NO:1], except that Sequence A may include up to 10 point mutations, such as substitutions with other nucleotides, per each 100 nucleotides of the referenced portion of the nucleic acid molecule in [SEQ ID NO:1]. Mutations described in this application preferably do not disrupt the reading frame of the coding sequence. Nucleic acid molecules functionally equivalent to the CSP sequences can occur in a variety of forms as described below.

Nucleic acid molecules may encode conservative amino acid changes in CSP peptide. The invention includes functionally equivalent nucleic acid molecules that encode conservative amino acid changes within a CSP amino acid sequence and produce silent amino acid changes in CSP.

Nucleic acid molecules may encode non-conservative amino acid substitutions, additions or deletions in CSP peptide. The invention includes functionally equivalent nucleic acid molecules that make non-conservative amino acid changes within the CSP amino acid sequence in [SEQ ID NO:2]. Functionally equivalent nucleic acid molecules include DNA and RNA that encode peptides, peptides and proteins having non-conservative amino acid substitutions (preferably substitution of a chemically similar amino acid), additions, or deletions but which also retain the same or similar CSP activity as the CSP peptide shown in [SEQ ID NO:2]. The DNA or RNA can encode fragments or variants of CSP. Fragments are useful as immunogens and in immunogenic compositions (U.S. Patent No. 5,837,472). The CSP or CSP-like activity of such fragments and variants is identified by assays as described below. Fragments and variants of CSP encompassed by the present invention should preferably have at least about 40%, 60%, 80% or 95% sequence identity to the naturally occurring CSP nucleic acid molecule, or a region of the sequence, such as the coding sequence or one of the conserved domains of the nucleic acid molecule, without being identical to the sequence in [SEQ ID NO:1]. Sequence identity is preferably measured with the ClustalW program (preferably using default parameters) (Thompson, JD et al., Nucleic Acid Res. 22:4673-4680)

Nucleic acid molecules functionally equivalent to the CSP nucleic acid molecule in [SEQ ID NO:1] will be apparent from the following description. For example, the sequence shown in [SEQ ID NO:1] may have its length altered by natural or artificial mutations such as partial nucleotide insertion or deletion, so that when the entire length of the coding sequence within [SEQ ID NO:1], is taken as 100%, the functional equivalent nucleic acid molecule preferably has a length of about 60-120% thereof, more preferably about 80-110% thereof. Fragments may be less than 60%.

Nucleic acid molecules containing partial (usually 80% or less, preferably 60% or less, more preferably 40% or less of the entire length) natural or artificial mutations so that some codons in these sequences code for different amino acids, but wherein the resulting peptide retains the same or similar CSP activity as that of a naturally occurring

CSP peptide. The mutated DNAs created in this manner should preferably encode a peptide having at least about 40%, preferably at least about 60%, at least about 80%, and more preferably at least about 90% or 95% sequence identity to the amino acid sequence of the CSP peptide in [SEQ ID NO:2]. Sequence identity is preferably assessed by the ClustalW program.

Since the genetic code is degenerate, the nucleic acid sequence in [SEQ ID NO:1] is not the only sequence which may code for a peptide having CSP activity. This invention includes nucleic acid molecules that have the same essential genetic information as the nucleic acid molecule described in [SEQ ID NO:1]. Nucleic acid molecules (including RNA) having one or more nucleic acid changes compared to the sequences described in this application and which result in production of a peptide shown in [SEQ ID NO:2] are within the scope of the invention.

Other functional equivalent forms of CSP-encoding nucleic acids can be isolated using conventional DNA-DNA or DNA-RNA hybridization techniques. Thus, the present invention also includes nucleic acid molecules that hybridize to one or more of the sequences in [SEQ ID NO:1] or its complementary sequence, and that encode expression for peptides, peptides and proteins exhibiting the same or similar activity as that of the CSP peptide produced by the DNA in [SEQ ID NO:1] or its variants. Such nucleic acid molecules preferably hybridize to the sequence in [SEQ ID NO:1] under moderate to high stringency conditions (see Sambrook et al. Molecular Cloning: A Laboratory Manual, Most Recent Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). High stringency washes have low salt (preferably about 0.2% SSC), and low stringency washes have high salt (preferably about 2% SSC). A temperature of about 37 °C or about 42 °C is considered low stringency, and a temperature of about 50-65 °C is high stringency. The invention also includes a method of identifying nucleic acid molecules encoding a CSP activator peptide (preferably a mammalian peptide), including contacting a sample containing nucleic acid molecules including all or part of [SEQ ID NO:1] (preferably at least about 15 or 20 nucleotides of [SEQ ID NO:1]) under moderate or high stringency hybridization conditions and identifying nucleic acid molecules which hybridize to the nucleic acid molecules including all or part of [SEQ ID NO:1].). Similar methods are described in U.S. Patent No. 5,851,788 which is incorporated by reference in its entirety.

The invention also includes methods of using all or part of the nucleic acid molecules which hybridize to all or part of [SEQ ID NO:1], for example as probes or in assays to identify antagonists or inhibitors of the peptides produced by the nucleic acid molecules (described below). The invention also includes methods of using nucleic acid molecules having sequence identity to the CSP nucleic acid molecule (as described below) in similar methods.

The invention also includes a nucleic acid molecule detection kit including, preferably in a suitable container means or attached to a surface, a nucleic acid molecule of the invention encoding CSP or a peptide having CSP activity and a detection reagent (such as a detectable label). Other variants of kits will be apparent from this description and teachings in patents such as U.S. Patent Nos. 5,837,472 and 5,801,233 which are incorporated by reference in their entirety.

A nucleic acid molecule described above is considered to have a function substantially equivalent to the CSP nucleic acid molecules of the present invention if the peptide produced by the nucleic acid molecule has CSP activity. A peptide has CSP activity if it can stimulate genetic competence and acid tolerance in *S. mutans*. Activation of the HK/RR is shown where a peptide is capable of stimulating the uptake and incorporation of foreign DNA. We describe below how the activity of these peptide-mediated processes can be measured by determining the efficiency of plasmid uptake, which is a measure of genetic competence. Since the ability to transport and incorporate foreign DNA relies on activation of the HK/RR and subsequent genes activated by the signal cascade initiated by the signal peptide, measurement of the conferment of erythromycin resistance by cells exposed to the peptide and plasmid DNA conferring erythromycin resistance indicates its level of function. Conversely if an inhibitor is capable of interfering with the action of the peptide the competence assay will indicate this by a corresponding decrease in the number of cells that acquire erythromycin resistance as described in the assays below (assays of genetic competence and assay of transformation of biofilms).

Production of CSP in eukaryotic and prokaryotic cells

The nucleic acid molecules of the invention may be obtained from a cDNA library. The nucleotide molecules can also be obtained from other sources known in the art such as expressed sequence tag analysis or *in vitro* synthesis. The DNA described in this application (including variants that are functional equivalents) can be introduced

into and expressed in a variety of eukaryotic and prokaryotic host cells. A recombinant nucleic acid molecule for the CSP contains suitable operatively linked transcriptional or translational regulatory elements. Suitable regulatory elements are derived from a variety of sources, and they may be readily selected by one with ordinary skill in the art (Sambrook, J, Fritsch, E.E. & Maniatis, T. (Most Recent Edition). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press. New York; Ausubel et al. (Most Recent Edition). Current Protocols in Molecular Biology, John Wiley & Sons, Inc.). For example, if one were to upregulate the expression of the nucleic acid molecule, one could insert a sense sequence and the appropriate promoter into the vector. Promoters can be inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific. Transcription is enhanced with promoters known in the art for expression. The CMV and SV40 promoters are commonly used to express desired peptide in cells. Other promoters known in the art may also be used (many suitable promoters and vectors are described in the applications and patents referenced in this application).

If one were to downregulate the expression of the nucleic acid molecule, one could insert the antisense sequence and the appropriate promoter into the vehicle. The nucleic acid molecule may be either isolated from a native source (in sense or antisense orientations), synthesized, or it may be a mutated native or synthetic sequence or a combination of these.

Examples of regulatory elements include a transcriptional promoter and enhancer or RNA polymerase binding sequence, a ribosomal binding sequence, including a translation initiation signal. Additionally, depending on the vector employed, other genetic elements, such as selectable markers, may be incorporated into the recombinant molecule. Other regulatory regions that may be used include an enhancer domain and a termination region. The regulatory elements may be from animal, plant, yeast, bacterial, fungal, viral, avian, insect or other sources, including synthetically produced elements and mutated elements.

In addition to using the expression vectors described above, the peptide may be expressed by inserting a recombinant nucleic acid molecule in a known expression system derived from bacteria, viruses, yeast, mammals, insects, fungi or birds. The recombinant molecule may be introduced into the cells by techniques such as *Agrobacterium tumefaciens*-mediated transformation, particle-bombardment-mediated transformation,

direct uptake, microinjection, coprecipitation, transfection and electroporation depending on the cell type. Retroviral vectors, adenoviral vectors, Adeno Associated Virus (AAV) vectors, DNA virus vectors and liposomes may be used. Suitable constructs are inserted in an expression vector, which may also include markers for selection of transformed cells. The construct may be inserted at a site created by restriction enzymes.

In one embodiment of the invention, a cell is transfected with a nucleic acid molecule of the invention inserted in an expression vector to produce cells expressing a peptide encoded by the nucleic acid molecule.

Another embodiment of the invention relates to a method of transfecting a cell with a nucleic acid molecule of the invention, inserted in an expression vector to produce a cell expressing the CSP peptide or other peptide of the invention. The invention also relates to a method of expressing the peptides of the invention in a cell. A preferred process would include culturing a cell including a recombinant DNA vector including a nucleic acid molecule encoding CSP (or another nucleic acid molecule of the invention) in a culture medium so that the peptide is expressed. The process preferably further includes recovering the peptide from the cells or culture medium.

Probes

The invention also includes oligonucleotide probes made from the cloned CSP nucleic acid molecules described in this application or other nucleic acid molecules of the invention (see Materials and Methods section). The probes may be 15 to 20 nucleotides in length. A preferred probe is at least 15 nucleotides of CSP in [SEQ ID NO:1]. The invention also includes at least 15 consecutive nucleotides of [SEQ ID NO:1]. The probes are useful to identify nucleic acids encoding CSP peptides as well as peptides functionally equivalent to CSP. The oligonucleotide probes are capable of hybridizing to the sequence shown in [SEQ ID NO:1] under stringent hybridization conditions. A nucleic acid molecule encoding a peptide of the invention may be isolated from other organisms by screening a library under moderate to high stringency hybridization conditions with a labeled probe. The activity of the peptide encoded by the nucleic acid molecule is assessed by cloning and expression of the DNA. After the expression product is isolated, the peptide is assayed for CSP activity as described in this application.

Functionally equivalent CSP nucleic acid molecules from other cells, or equivalent CSP-encoding cDNAs or synthetic DNAs, can also be isolated by amplification using Polymerase Chain Reaction (PCR) methods. Oligonucleotide primers, such as degenerate primers, based on [SEQ ID NO:1] can be prepared and used with PCR and reverse transcriptase (E. S. Kawasaki (1990), In Innis et al., Eds., PCR Protocols, Academic Press, San Diego, Chapter 3, p. 21) to amplify functional equivalent DNAs from genomic or cDNA libraries of other organisms. The oligonucleotides can also be used as probes to screen cDNA libraries.

Functionally equivalent peptides, peptides and proteins

The present invention includes not only the peptides encoded by the sequences of the invention, but also functionally equivalent peptides, peptides and proteins that exhibit the same or similar CSP peptide activity. A peptide is considered to possess a function substantially equivalent to that of the CSP peptide if it has CSP activity. CSP activity means that it is able to confer genetic competence to *S. mutans*, as measured by an increased ability to incorporate and express foreign genetic material, when added to cells as described in the assay of genetic competence below. Functionally equivalent peptides, peptides and proteins include peptides, peptides and proteins that have the same or similar protein activity as CSP when assayed, i.e. they are able to stimulate genetic competence and low pH tolerance (the ability to withstand acid challenges of pH 3.5 –pH 3.0 for up to 3 hours) in *S. mutans*. A peptide has CSP activity if it is capable of increasing the frequency of uptake and expression of foreign DNA as described in the following assay for genetic competence.

Identity refers to the similarity of two peptides or proteins that are aligned so that the highest order match is obtained. Identity is calculated according to methods known in the art, such as the ClustalW program. For example, if a peptide (called "Sequence A") has 90% identity to a portion of the peptide in [SEQ ID NO:2], then Sequence A will be identical to the referenced portion of the peptide in [SEQ ID NO:2], except that Sequence A may include up to 1 point mutations, such as substitutions with other amino acids, per each 10 amino acids of the referenced portion of the peptide in [SEQ ID NO:2]. Peptides, peptides and proteins functional equivalent to the CSP peptides can occur in a variety of forms as described below.

Peptides biologically equivalent in function to CSP peptide include amino acid sequences containing amino acid changes in the CSP sequence. The functional equivalent peptides have at least about 40% sequence identity, preferably at least about 60%, at least about 75%, at least about 80%, at least about 90% or at least about 95% sequence identity, to the natural CSP peptide or a corresponding region. "Sequence identity" is preferably determined by the ClustalW program. Most preferably, 1, 2, 3, 4, 5, 5-10, 10-15 amino acids are modified.

Variants of the CSP peptide may also be created by splicing. A combination of techniques known in the art may be used to substitute, delete or add amino acids. For example, a hydrophobic residue such as methionine can be substituted for another hydrophobic residue such as alanine. An alanine residue may be substituted with a more hydrophobic residue such as leucine, valine or isoleucine. An aromatic residue such as phenylalanine may be substituted for tyrosine. An acidic, negatively-charged amino acid such as aspartic acid may be substituted for glutamic acid. A positively-charged amino acid such as lysine may be substituted for another positively-charged amino acid such as arginine. Modifications of the peptides of the invention may also be made by treating a peptide of the invention with an agent that chemically alters a side group, for example, by converting a hydrogen group to another group such as a hydroxy or amino group.

Peptides having one or more D-amino acids are contemplated within the invention. Also contemplated are peptides where one or more amino acids are acetylated at the N-terminus. Those skilled in the art recognize that a variety of techniques are available for constructing peptide mimetics (i.e., a modified peptide or peptide or protein) with the same or similar desired biological activity as the corresponding peptide of the invention but with more favorable activity than the peptide with respect to characteristics such as solubility, stability, and/or susceptibility to hydrolysis and proteolysis. See for example, Morgan and Gainor, *Ann. Rep. Med. Chem.*, 24:243-252 (1989).

The invention also includes hybrid nucleic acid molecules and peptides, for example where a nucleic acid molecule from the nucleic acid molecule of the invention is combined with another nucleic acid molecule to produce a nucleic acid molecule which expresses a fusion peptide. One or more of the other domains of CSP described in this application could also be used to make fusion peptides. For example, a nucleotide

domain from a molecule of interest may be ligated to all or part of a nucleic acid molecule encoding CSP peptide (or a molecule having sequence identity) described in this application. Fusion nucleic acid molecules and peptides can also be chemically synthesized or produced using other known techniques. The invention includes a nucleic acid molecule encoding a fusion peptide or a recombinant vector including the nucleic acid molecule

The variants preferably retain the same or similar CSP activity as the naturally occurring CSP. The CSP activity of such variants can be assayed by techniques described in this application and known in the art.

Variants produced by combinations of the techniques described above but which retain the same or similar CSP activity as naturally occurring CSP are also included in the invention (for example, combinations of amino acid additions, and substitutions).

Variants of CSP encompassed by the present invention preferably have at least about 40% sequence identity, preferably at least about 60%, 75%, 80%, 90% or 95% sequence identity, to the naturally occurring peptide, or corresponding region or moiety of the peptide, or corresponding region. Sequence identity is preferably measured with the ClustalW.

Histidine Kinase & Response Regulator

The invention also includes sequences having identity with the histidine kinase, response regulator of the invention and comA and comB. Preferred percentages of identity (nucleic acid molecule and polypeptide) are the same as those described for the CSP.

As well, probes and antibodies for a histidine kinase, response regulator comA or comB may be prepared using the description in this application and techniques known in the art. The description for preparation of CSP variants and mutants is also applicable to the histidine kinase, response regulator or comA and comB of the invention. The invention also includes fragments of HK having HK activity, fragments of RR having RR activity and fragments of comA or comB having activity.

Design of CSP peptide competitive inhibitors

The activity of the CSP peptide may be varied by carrying out selective site-directed mutagenesis. We characterize the binding domain and other critical amino acid residues in the peptide that are candidates for mutation, insertion and/or deletion. Sequence variants may be synthesized. A DNA plasmid or expression vector containing the CSP nucleic acid molecule or a nucleic acid molecule having sequence identity may be used for these studies using the U.S.E. (Unique site elimination) mutagenesis kit from Pharmacia Biotech or other mutagenesis kits that are commercially available, or using PCR. Once the mutation is created and confirmed by DNA sequence analysis, the mutant peptide is expressed using an expression system and its activity is monitored. This approach is useful to identify CSP inhibitors. All these modifications of the CSP DNA sequences presented in this application and the peptides produced by the modified sequences are encompassed by the present invention.

Pharmaceutical compositions

The CSP inhibitors are also useful when combined with a carrier in a pharmaceutical composition. The compositions are useful when administered in methods of medical treatment or prophylaxis of a disease, disorder or abnormal physical state caused by *S. mutans*. The invention also includes methods of medical treatment of a disease, disorder or abnormal physical state characterized by excessive *S. mutans* or levels or activity of CSP peptide, for example by administering a pharmaceutical composition including a carrier and a CSP inhibitor. Caries is one example of a disease which can be treated or prevented by antagonizing CSP.

The pharmaceutical compositions can be administered to humans or animals by methods such as gel, toothpaste, mouthwash, dental floss or chewing gum in methods of medical treatment. The peptides of the invention may be coupled to lipids or carbohydrates. This increases their ability to adhere to teeth, either by prolonging the duration of the adhesion or by increasing its affinity, or both. They may also be coupled to polymers, for example in dental work (eg. crowns, braces, fillings) or dental floss. The pharmaceutical compositions can be administered to humans or animals. Dosages to be administered depend on individual patient condition, indication of the drug, physical and chemical stability of the drug, toxicity of the desired effect and the chosen route of

administration (Robert Rakel, ed., Conn's Current Therapy (1995, W.B. Saunders Company, USA)). The pharmaceutical compositions are used to treat diseases caused by streptococcal infections such as caries and endocarditis.

CSP activity could be blocked by antisense mRNA or by inhibiting the activity of the exporter that secretes it from the cell. We have the sequence of these exporters. There are two copies of the genes (comAB) that are involved in export.

Nucleic acid molecules (antisense inhibitors of CSP) and competitive inhibitors of CSP may be introduced into cells using *in vivo* delivery vehicles such as liposomes. They may also be introduced into these cells using physical techniques such as microinjection and electroporation or chemical methods such as coprecipitation or using liposomes.

The pharmaceutical compositions can be prepared by known methods for the preparation of pharmaceutically acceptable compositions which can be administered to patients, and such that an effective quantity of the nucleic acid molecule or peptide is combined in a mixture with a pharmaceutically acceptable vehicle. Suitable carriers are described, for example in Remington's Pharmaceutical Sciences (Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., USA). Carriers include saline and D5W (5% dextrose and water). Excipients include additives such as a buffer, solubilizer, suspending agent, emulsifying agent, viscosity controlling agent, flavor, lactose filler, antioxidant, preservative or dye. There are preferred excipients for stabilizing peptides for parenteral and other administration. The excipients include serum albumin, glutamic or aspartic acid, phospholipids and fatty acids.

On this basis, the pharmaceutical compositions could include an active compound or substance, such as a CSP inhibitor, in association with one or more pharmaceutically acceptable vehicles or diluents, and contained in buffered solutions with a suitable pH and isoosmotic with the physiological fluids. The methods of combining the active molecules with the vehicles or combining them with diluents is well known to those skilled in the art. The compositions may also contain additives such as antioxidants, buffers, bacteriostatis, bactericidal antibiotics and solutes which render the formulation isotonic in the intended recipient; and aqueous and non-aqueous sterile suspensions which may include suspending agents and thickening agents. The composition could include a targeting agent for the transport of the active compound to specified sites.

Vaccines

Antibodies directed against the CSP would provide protection against caries. Antibodies may be manufactured as described below. Alternatively, a peptide of the invention or a fragment thereof may be used with a carrier to make a vaccine. The peptide or fragment may also be conjugated to another molecule to increase its antigenicity. Antibodies can also be coupled to the peptide (Brady, L.J. et al., "Monoclonal Antibody-Mediated Modulation of the Humoral Immune Response against Mucosally Applied *Streptococcus mutans*" (in press). In order to enhance the immune response the peptide can be coupled to KLH, ovalbumin, or thyroglobulin prior to immunization. The vaccine composition will trigger the mammal's immune system to produce antibodies. The invention includes vaccine compositions and methods of vaccinating a mammal, preferably a human, against dental caries by administering to the mammal an effective amount of a vaccine composition. Techniques for preparing and using vaccines are known in the art. To prepare the vaccine, the peptide, or a fragment of the peptide, may be mixed with other antigens (of different immunogenicity), a vehicle or an excipient. Examples of peptide vaccines are found in U.S. Patent Nos. 5,679,352, 5,194,254 and 4,950,480. Techniques for preparing vaccines involving site-directed mutagenesis are described in U.S. Patent Nos. 5,714,372, 5,543,302, 5,433,945, 5,358,868, 5,332,583, 5,244,657, 5,221,618, 5,147,643, 5,085,862 and 5,073,494. Vaccines may be administered by known techniques, such as topical or parenteral administration. Vast changes are taking place in vaccinology consequent to the introduction of new technologies. Acellular purified fractions devoid of side effects, non-pathogenic but immunogenic mutants, recombinant technology, conjugated vaccines, combination vaccines (to limit the number of injections). Vaccine delivery systems can deliver multiple doses of the vaccine at a single contact point. A genetically engineered oral vaccine is useful to impart better and longer duration of immunity. Oral vaccines are useful. The nose as a route for immunization is also useful. DNA alone can constitute the vaccines, inducing both humoral and cell-mediated immune responses. Live recombinant vaccines are also useful. Potent adjuvants add to the efficacy of the vaccines. One can also 'humanize' mouse monoclonals by genetic engineering and express these efficiently in plants. These recombinant antibodies are opening out an era of highly specific and safe therapeutic interventions. An advantage of preformed antibodies directed at a defined target and given in adequate amounts is the certainty of efficacy in every recipient, in contrast to vaccines, where the quality and quantum of immune response varies from

individual to individual. For example, nasal immunization may be done as described in C. Jespersgaard et al. "Protective Immunity against *Streptococcus mutans* Infection in Mice after Intranasal Immunization with the Glucan-Binding Region of *S. mutans* Glucosyltransferase" Infection and Immunity, December 1999, p. 6543-6549, Vol. 67, No. 12. Vaccine compositions may comprise solid or liquid formulations such as gels, sprays, inhalants, tablets, toothpastes, mouthwashes or chewing gum.

For vaccine application, cholera toxin can be used by coupling the peptide to its B-subunit to stimulate production of secretory antibody i.e., Coupling to CTB.

Screening for inhibitors of CSP

Inhibitors are preferably directed towards CSP to block *S. mutans* competence, low pH tolerance and biofilm formation.

A method of identifying a compound which reduces the interaction of CSP with HK, can include: contacting (i) CSP with (ii) HK, a CSP-binding fragment of HK or a derivative of either of the foregoing in the presence of the compound; and b)

determining whether the interaction between (i) and (ii) is reduced, thereby indicating that the compound reduces the interaction of CSP and HK. A CSP inhibitor (caries treating or preventing compound) inhibits the interaction between (i) and (ii). By way of example, one can screen a synthetic peptide library. One could also screen small non-peptide organic molecules.

In one embodiment, the invention includes an assay for evaluating whether test compounds are capable of acting as agonists or antagonists for CSP, or a peptide having CSP functional activity, including culturing cells containing DNA which expresses CSP, or a peptide having CSP activity so that the culturing is carried out in the presence of at least one compound whose ability to modulate CSP activity is sought to be determined and thereafter monitoring the cells for either an increase or decrease in the level of CSP or CSP activity. Other assays (as well as variations of the above assay) will be apparent from the description of this invention and techniques such as those disclosed in U.S. Patent No. 5,851,788, 5,736,337 and 5,767,075 which are incorporated by reference in their entirety. For example, the test compound levels may be either fixed or variable.

Preparation of antibodies

The CSP peptide is also useful as an antigen for the preparation of antibodies that can be used to purify or detect other CSP-like peptides. Antibodies may also block CSP binding to HK. Antibodies are preferably targeted to the entire CSP sequence. The CSP peptide may be conjugated to other compounds, in order to increase immunogenicity.

We generate polyclonal antibodies against the CSP which is a unique sequence. Monoclonal and polyclonal antibodies are prepared according to the description in this application and techniques known in the art. For examples of methods of preparation and uses of monoclonal antibodies, see U.S. Patent Nos. 5,688,681, 5,688,657, 5,683,693, 5,667,781, 5,665,356, 5,591,628, 5,510,241, 5,503,987, 5,501,988, 5,500,345 and 5,496,705 which are incorporated by reference in their entirety. Examples of the preparation and uses of polyclonal antibodies are disclosed in U.S. Patent Nos. 5,512,282, 4,828,985, 5,225,331 and 5,124,147 which are incorporated by reference in their entirety. Antibodies recognizing CSP can be employed to screen organisms or tissues containing CSP peptide or CSP-like peptides. The antibodies are also valuable for immuno-purification of CSP or CSP-like peptides from crude extracts.

An antibody (preferably the antibody described above) may be used to detect CSP or a similar peptide, for example, by contacting a biological sample with the antibody under conditions allowing the formation of an immunological complex between the antibody and a peptide recognized by the antibody and detecting the presence or absence of the immunological complex whereby the presence of CSP or a similar peptide is detected in the sample. The invention also includes compositions preferably including the antibody, a medium suitable for the formation of an immunological complex between the antibody and a peptide recognized by the antibody and a reagent capable of detecting the immunolgical complex to ascertain the presence of CSP or a similar peptide. The invention also includes a kit for the *in vitro* detection of the presence or absence of CSP or a similar peptide in a biological sample, wherein the kit preferably includes an antibody, a medium suitable for the formation of an immunological complex between the antibody and a peptide recognized by the antibody and a reagent capable of detecting the immunological complex to ascertain the presence of CSP or a similar peptide in a biological sample. Further background on the use of antibodies is provided,

for example in U.S. Patent Nos. 5,695,931 and 5,837,472 which are incorporated by reference in their entirety.

Research Tool

Assay of genetic competence

The ability of the peptide to activate the HK and RR and the subsequent genes involved in the conferral of the properties of genetic competence, acid tolerance and biofilm formation is determined by measuring the efficiency of uptake and expression of DNA (preferably plasmid DNA) in *S. mutans* when exposed to signal peptide and/or inhibitor. Two methods modified based on the protocols described by Perry *et al.* Infect Immun, 41:722-727 and Lindler and Macrina J Bacteriol, 166:658-665 are used to assay genetic competence. The method involves adding DNA and CSP (preferably plasmid DNA) to a S. mutans culture (or culture of a bacteria expressing CSP or a variant thereof). The rate of transformation is then determined. *S. mutans* is preferably grown in THYE plus 5% horse serum (THYE-HS). After 2-hr incubation, 1 μg/ml plasmid DNA or 10 μg/ml of chromosomal DNA is added to the culture. To assay induction of competence, competence signal peptide, (SCSP) is then added to the cultures, incubation continued for 30 minutes with a final concentration of 500 ng/ml of SCSP added to each sample. After the 30-minute incubation equal amounts of DNA is added to each well (1 μg/ml plasmid or 10 μg/ml of chromosomal DNA) and incubation continued for another 2 hrs.. Cell dilutions were immediately spread on THYE agar plates plus appropriate antibiotics. Transformation frequency was expressed as the number of transformants (antibiotic resistant cells) per number of viable recipients. This is determined by comparing the number of cells able to grow in the presence of antibiotic (conferred by the applied plasmid or chromosomal DNA) relative to the total number of cells present (i.e., that grow in the absence of antibiotic). A higher value indicates a higher rate of transformation and thus is reflective of a stimulatory effect by the peptide. Consequently, addition of a molecule that successfully acts as an inhibitor results in a lower ratio of transformants/recipients, indicating that the inhibitor is effective at blocking activity of the CSP. CSP deficient cells may also be used in a variation of these assays. One can identify compounds that inhibit CSP or variants thereof by adding a test compound to the mixture to determine if the rate of transformation is decreased by the addition of the test compound.

The activity of the system can also be measured by an *in vitro* assay that relies on the measurement of marker protein expression (such as green fluorescent protein (GFP)) via expression from a fusion to a promoter controlled by the signal cascade initiated by CSP/HK/RR. One such promoter occurs immediately 5' proximal to the *S. mutans* comX gene. *S. mutans* cells grown in microtiter wells are exposed to the CSP and/or inhibitor and the level of fluorescence of the comX::GFP strain is measured to give a quantitative measure of CSP stimulation (and conversely inhibitor activity). One can identify compounds that inhibit CSP or variants thereof by adding a test compound to the mixture to determine if the quantitative measure of CSP stimulation is decreased by the addition of the test compound.

Cells transformed with a nucleic acid molecule of the invention (histidine kinase, CSP or response regulator) are useful as research tools. For example, one may obtain a cell (or a cell line, such as an immortalized cell culture or a primary cell culture) that does not express histidine kinase, CSP or response regulator, insert a histidine kinase, CSP or response regulator nucleic acid molecule in the cell, and assess the level of expression and activity. Alternatively, histidine kinase, CSP or response regulator nucleic acid molecules may be over-expressed in a cell that expresses a histidine kinase, CSP or response regulator nucleic acid molecule. In another example, experimental groups of cells may be transformed with vectors containing different types of histidine kinase, CSP or response regulator nucleic acid molecules to assess the levels of polypeptides and peptides produced, its functionality and the phenotype of the plants (for example, phenotype in saline soil). The polypeptides and peptides are also useful for *in vitro* analysis of histidine kinase, CSP or response regulator activity or structure. For example, the polypeptides and peptides produced can be used for microscopy or X-ray crystallography studies.

The histidine kinase, CSP or response regulator nucleic acid molecules and polypeptides are also useful in assays. Assays are useful for identification and development of compounds to inhibit and/or enhance polypeptide or peptide function directly. For example, they are useful in an assay for evaluating whether test compounds are capable of acting as antagonists for histidine kinase, CSP or response regulator by: (a) culturing cells containing a nucleic acid molecule which expresses histidine kinase, CSP or response regulator peptides (or fragments or variants thereof having histidine kinase, CSP or response regulator activity) wherein the culturing is carried out in the presence of: increasing concentrations of at least one test compound

whose ability to inhibit histidine kinase, CSP or response regulator is sought to be determined; and (b) monitoring in the cells the level of inhibition as a function of the concentration of the test compound, thereby indicating the ability of the test compound to inhibit histidine kinase, CSP or response regulator activity.

Suitable assays may be adapted from, for example, US patent no. 5,851,788.

MATERIALS AND METHODS:

Growth conditions of cells

Cells are grown in Todd Hewitt yeast extract medium at various dilutions with and without 5% horse serum and 0.01% hog gastric mucin.

Protocol for Transformation of Biofilm-grown Cells

Biofilms are developed on polystyrene microtiter plates to provide a rapid and simple method for assaying biofilm formation, and hence activity of the peptide/receptor/kinase system. Formation of biofilms is initiated by inoculating 20 ul of cell suspension into each well containing 2 ml of biofilm medium (4X diluted Todd-Hewitt Yeast Extract supplemented with final concentration of 0.01% hog gastric mucin) for overnight incubation at 37°C under an anaerobic condition. After 20-h incubation, fluid medium is removed and added with 2 ml of pre-warmed, fresh THYE plus 5% horse serum. The cultures are incubated for 30 minutes and each well is supplemented with a final concentration of 200 ng/ml of synthetic competence stimulating peptide (SCSP) and varying concentrations of the inhibitor and the incubation is continued. After 30 minutes, plasmid DNA (1 mg/ml) or chromosomal DNA (10 mg/ml) is added to each well and the cultures are incubated for an additional 2 hr. Planktonic cells are then removed and the wells are washed once with PBS buffer. Biofilm cells are collected into 2 ml fresh medium by a gentle sonication or washing the wells using a pipette. The samples are centrifuged at 12,000 xg for 5 min. Both biofilm and planktonic cells are resuspended into 200 ul of fresh medium and are immediately spread on THYE agar plus appropriate antibiotics. Transformation frequency is determined after 48-h of incubation.

Genome database analysis

Homologs of the *Streptococcus pneumoniae comD/E* genes encoding a histidine kinase/ response regulator system were identified. This sequence was used to design primers to amplify the region from a number of *S. mutans* isolates. An open reading frame consisting of 138 nucleotides was located 148 nucleotides 5' proximal from the end of the *comD* homolog in the opposite orientation (Fig 1). This ORF was found to encode a peptide of 46-amino acid in length, the precursor of the 21-amino acid CSP.

PCR amplification and nucleotide sequencing

The comCDE genes were amplified from the genomes of several *S. mutans* isolates by PCR using primers designed based on the genome database sequence and their nucleotide sequences determined. The deduced amino acid sequences are compared among the isolates by sequence alignment to confirm identity.

Gene inactivations

Genes are inactivated by integration of internal homologous fragments into the suicide vector pVA8912. Mutants defective in each of the individual genes (*comC, comD, comE*) are inactivated and their phenotypes are compared to the parent strain for their abilities to form biofilms, tolerate acidic pH (pH 2-4), and transport and incorporate DNA.

Synthesis of synthetic peptide

The sequence of the processed peptide was deduced by determining the cleavage site to be located beside the gly-gly amino acid residues (numbers 24 and 25) (fig. 4). A peptide was synthesized corresponding to amino acid sequence of residues 26-46 inclusive.

Bacterial strains and growth conditions

Seven strains of *S. mutans* were used in this study (strains include: BM71, GB14, H7, JH1005, LT11, NG8, and UAB159. All the strains were cultured from freeze-dried ampoules and routinely maintained on Todd-Hewitt Yeast Extract (THYE) plates. For selection of antibiotic resistant colonies following transformation, the medium was supplemented with either erythromycin (Em) (10 μg/ml) or kanamycin (Km) (500 μg/ml).

Assay for biofilms formed on polystyrene microtiter plates

Biofilms were developed on polystyrene microtiter plates to provide a rapid and simple method for assaying genetic transformation. A 4X diluted THYE medium supplemented with final concentration of 0.01% hog gastric mucin was used as biofilm medium (BM). Formation of biofilms was initiated by inoculating 20 µl of cell suspension into each well containing 2 ml of BM and four wells were set up: two for assaying transformation and two for quantification of biofilms. After cultures were incubated at 37^{0}C for 20 h under an anaerobic condition, fluid medium was removed for viable cell counts. The wells were rinsed once with 10mM PBS buffer (pH 7.2) and biofilm cells were collected in 2 ml PBS by a gentle sonication for 15 seconds. Both biofilm and the planktonic cells were immediately spread on THYE plates using a spiral system (Spriral Plater, Model D, Cincinnati, OH) and incubated at 37^{0}C under an anaerobic condition. Formation of biofilms was quantified by viable cell counts after 48 h of incubation.

Assay for "steady-state" biofilms

Biofilms were also grown in a chemostat-based biofilm fermentor to define and optimize the conditions for genetic competence of biofilm-grown cells. The biofilm fermentor was modified in the Mechanical Engineering and Glass Blowing Shops, University of Toronto, based on a similar system described previously (Li and Bowden, 1994). The vessel was made of glass with a working volume of 400 ml. The vessel lip was constructed of stainless steel with 10 sampling ports, which allowed sterile insertion and retrieval of glass rods (0.5 cm in diameter, approximately 4.0 cm^2 area immersed in fluid medium), providing abiotic surfaces for accumulation of biofilms. Temperature in the chemostat vessel was maintained at 37^{0}C ± 0.1 by a temperature controller (Model R-600F, Cole Parmer Instrument Cop., Vernon Hill, IL). The culture pH was controlled by a pH control unit (Digital pH Meter/Controller, Model 501-3400, Barnant Corp. Barrington, IL) through the addition of 1M KOH or 1M HCl. The vessel was placed on a magnetic stirrer (Fisher Scientific) and the culture was stirred at 200 rpm by a polypropylene coated magnetic stirrer bar (3cm in length). Continuous cultures were obtained by pumping fresh 4X diluted THYE medium supplemented with a final concentration of 0.01% hog gastric mucin (Type III, Sigma) into the vessel (400 ml) at the desired dilution rates. Daily maintenance of the chemostat included optical density reading, viable cell counts and pH measurement in fluid cultures. When the cultures reached "steady-state" (at least 10 mean generation times), glass rods were aseptically inserted into the chemostat for the

initiation of biofilm formation. Then, biofilms of different ages were removed from the cultures for both genetic transformation and quantification of biofilms using viable cell counts.

Transformation protocol

Two methods modified based on the protocols described by Perry *et al* (Infect Immun, 41:722-727) and Lindler and Macrina (J Bacteriol, 166:658-665) were used to assay natural transformation of biofilm cells. Biofilms formed on polystyrene microtiter plates were added with 2 ml of pre-warmed, fresh THYE plus 5% horse serum (THYE-HS) immediately following removal of the BM medium, and the incubation continued at 37^{0}C. After 2-hr incubation, a final concentration of 1 μg/ml plasmid DNA or 10 μg/ml of chromosomal DNA was added to each well. The cultures were incubated for an additional 2 hr before collection of the cells for plating. To assay induction of competence by synthetic competence stimulating peptide (SCSP), the cultures were incubated for 30 minutes and a final concentration of 500 ng/ml of SCSP was added to each well. After a 30-minute incubation, equal amounts of DNA was added to each well (1 μg/ml plasmid or 10 μg/ml of chromosomal DNA) and incubation continued for another 2 hr. Fluid medium was then removed from individual wells and the wells were washed once with PBS buffer. Biofilm cells were collected into 2 ml PBS buffer by gentle sonication or by washing the wells using a pipette. The samples were centrifuged at 12,000 X g for 5 minutes. Both biofilm and planktonic cells were resuspended into 200 μl of fresh medium and were immediately spread on THYE agar plates plus appropriate antibiotics. For the biofilms developed in the chemostat, rods with biofilm cells were removed and placed into 2 ml of pre-warmed, fresh THYE-HS medium for 30-min incubation. Transformation was then initiated by using the same methods as described above. The planktonic cells were also removed to compare the transformation frequency. After completion of the transformation procedures, both biofilm and planktonic cells were spread on THYE agar plus appropriate antibiotic. Transformation frequency was assessed after 48-h incubation. Transformation frequency was expressed as the number of transformants per μg DNA per viable recipient at the time of DNA added.

Donor DNA

Both plasmid and chromosomal DNA were used as donor DNA to assay genetic transformation in this study. Plasmid DNA included an integrative plasmid, pVAGTFA carrying an erythromycin resistance (Em^r) determinant and a fragment of the *S. mutans* *gtfA* gene. The replicative plasmid, pDL289 carrying a kanamycin resistance gene (Km^r) was also used. Chromosomal DNA harboring an Em^r gene was prepared from a recombinant *S. mutans* strain harboring a chromosomally integrated copy of pVAGTFA.

The present invention has been described in detail and with particular reference to the preferred embodiments; however, it will be understood by one having ordinary skill in the art that changes can be made without departing from the spirit and scope thereof. For example, where the application refers to peptides, it is clear that polypeptides may often be used. Likewise, where a gene is described in the application, it is clear that nucleic acid molecules or gene fragments may often be used.

All publications (including GenBank entries), patents and patent applications are incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

We claim:

1. A compound that competitively inhibits binding of CSP to *S. mutans* histidine kinase.

2. The compound of claim 1, comprising a peptide or an antibody.

3. The compound of claim 2, comprising a derivative of [SEQ ID NO:2], a fragment of [SEQ ID NO:2] or a derivative of a fragment of [SEQ ID NO:2].

4. The compound of claim 3, wherein amino acids are removed from the N-terminus and/or C-terminus of [SEQ ID NO:2].

5. A pharmaceutical composition comprising all or part of the peptide of claims 1 to 4 and a carrier.

6. A method of medical treatment or prophylaxis of caries or endocarditis, comprising administering the compound of any of claims 1 to 4 or the pharmaceutical composition of claim 5.

7. An isolated nucleic acid molecule encoding a *S. mutans* competence signal peptide, or a fragment of a peptide having CSP activity.

8. An isolated nucleic acid molecule encoding a competence signal peptide, or a fragment of a competence signal peptide having *S. mutans* competence signal peptide activity, comprising a nucleic acid molecule selected from the group consisting of:

   (a) a nucleic acid molecule that hybridizes to all or part of a nucleic acid molecule shown in [SEQ ID NO:1], the fragment of [SEQ. ID NO:1] encoding [SEQ ID NO:2] or a complement thereof under moderate or high stringency hybridization conditions;

   (b) a nucleic acid molecule degenerate with respect to (a).

9. An isolated nucleic acid molecule encoding a competence signal peptide, or a fragment of a competence signal peptide having *S. mutans* competence signal peptide activity, comprising a nucleic acid molecule selected from the group consisting of:

   (a) the nucleic acid molecule of the coding strand shown in [SEQ ID NO:1], or a complement thereof;

(b) a nucleic acid molecule encoding the same amino acid sequence as a nucleotide sequence of (a); and

(c) a nucleic acid molecule having at least 50% or 60% identity with the nucleotide sequence of (a) or the fragment of [SEQ. ID NO:1] encoding [SEQ ID NO:2].

10. The nucleic acid molecule of claim 1, comprising all or part of a nucleotide sequence shown in [SEQ ID NO:1], the fragment of [SEQ. ID NO:1] encoding [SEQ ID NO:2], or a complement thereof.

11. A CSP nucleic acid molecule isolated from *S. mutans*, or a fragment thereof having CSP activity.

12. A recombinant nucleic acid molecule comprising a nucleic acid molecule of any of claims 7 to 11 and a constitutive promoter sequence or an inducible promoter sequence, operatively linked so that the promoter enhances transcription of the nucleic acid molecule in a host cell.

13. The nucleic acid molecule of any of claims 7 to 11, wherein the molecule comprises genomic DNA, or cDNA.

14. The nucleic acid molecule of any of claims 7 to 11, wherein the nucleic acid molecule is chemically synthesized.

15. The nucleic acid molecule of any of claims 7 to 11, wherein the CSP is involved in genetic competence, biofilm formation, and acid tolerance of *S. mutans*.

16. A vector comprising the nucleic acid molecule of any of claims 7 to 11.

17. A host cell comprising the recombinant nucleic acid molecule of claim 12 or the vector of claim 16, or progeny of the host cell.

18. The host cell of claim 17, selected from the group consisting of a fungal cell, a yeast cell, a bacterial cell, a microorganism cell and a plant cell.

19. A method for producing a recombinant host cell capable of expressing the nucleic acid molecule of any of claims 7 to 11, the method comprising introducing into the host cell a vector of claim 16.

20. A method for expressing a peptide in the host cell of claim 19, the method comprising culturing the host cell under conditions suitable for gene expression.

21. An isolated polypeptide encoded by and/or produced from the nucleic acid molecule of any of claims 7 to 11, or the vector of claim 16.

22. An isolated CSP or a fragment thereof having *S. mutans* CSP activity.

23. The polypeptide of claim 22 comprising a *S. mutans* CSP.

24. The polypeptide of claim 23 comprising all or part of an amino acid sequence in [SEQ ID NO:2].

25. A polypeptide fragment of the peptide of claim 24, or a peptide mimetic of the CSP.

26. The polypeptide of claim 24 which is recombinantly produced.

27. A polypeptide comprising a sequence having greater than 30%, 50% or 60% sequence identity to the polypeptide of claim 24.

28. The polypeptide of claim 24, isolated from *S. mutans*.

29. An isolated nucleic acid molecule encoding the polypeptide of any of claims 21 to 28.

30. An antibody directed against the polypeptide of any of claims 21 to 28.

31. The antibody of claim 30, comprising a monoclonal antibody or a polyclonal antibody.

32. A vaccine composition comprising all or part of the peptide of any of claims 21 to 28 and a carrier.

33. The vaccine composition of claim 32, wherein the peptide is coupled to a compound comprising all or part of KLH, ovalbumin, or thyroglobulin.

34. A method of evaluating caries-reducing properties of a compound comprising contacting the compound with:

a) CSP, a HK -binding fragment of CSP or a derivative of either of the foregoing; and

b) HK, a CSP binding fragment of HK or a derivative of either of the foregoing;

wherein (a) and (b) are capable of binding; and determining the ability of the compound to interfere with the binding of a) with b), the ability to interfere with binding indicating that the compound reduces caries.

35. A method of evaluating caries-reducing properties of a compound comprising contacting the compound with:

a) a DNA vector encoding a marker gene; and

b) a S. mutans culture;

by determining whether the compound reduces uptake of the DNA vector into the *S. mutans* culture, the reduced uptake of the DNA vector indicating that the compound reduces caries.

36. The method of claim 35, wherein reduction of caries is indicted by reduced transformation efficiency in S. mutans.

37. The method of claim 35, wherein reduction of caries is indicted by determining changes in the physiological characteristics of biofilm formation and acid tolerance in *S. mutans*.

Figure 1
Schematic layout of the arrangement of the genetic locus encoding the signal peptide precursor, the histidine kinase and the response regulator. Note that this arrangement is different from other loci in related streptococci for the following reasons: a) The comC gene is transcribed from its own promoter alone, unlike the genes thus far described in other streptococci that are arranged in an operon-like cluster with the comC/DE genes being transcribed from a single promoter.
b) The comC gene is separated from the *comD* gene by 148 nucleotides.

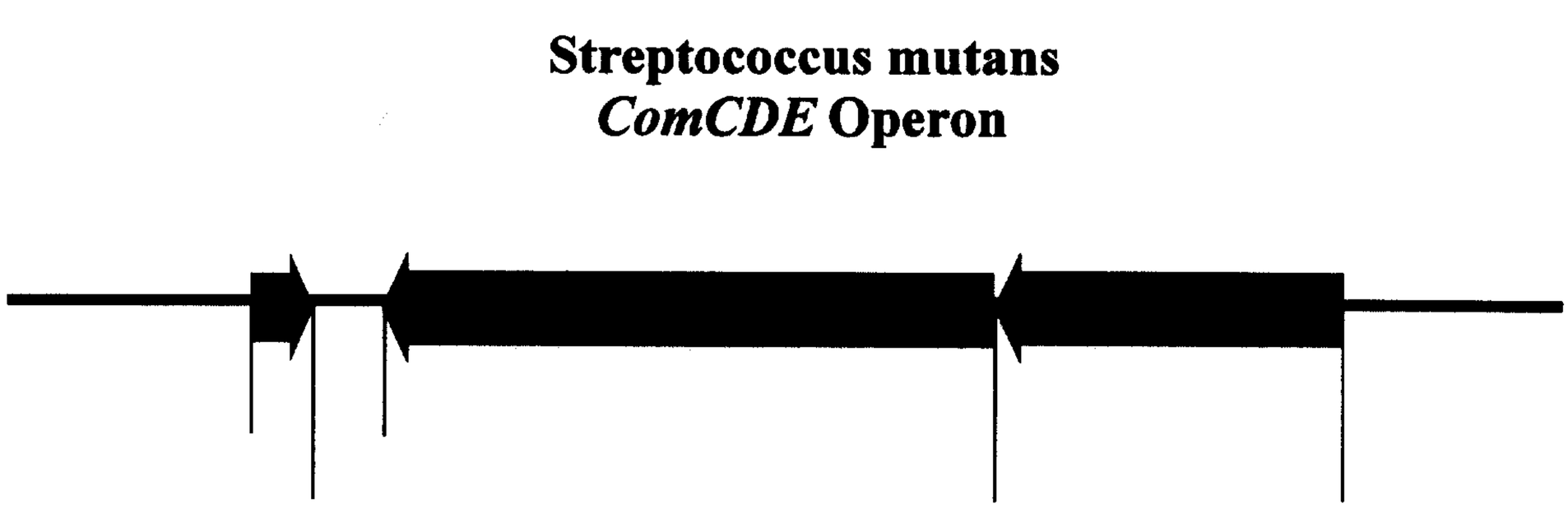


Figure 2
Sequences of the open reading frames encoding the signal peptide precursor (ComC), the histidine kinase (ComD), and the response regulator (ComE).

➢ S. mutans comC gene
Encodes a precursor to a signal peptide
[ATGAAAAAAACACTATCATTAAAAAATGACTTTAAAGAAATTAAGACTGATGAATTAG
AGATTATCATTGGCGGA(AGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAG
TTTTACACAAGCTTTGGGAAAA)]TAA

➢ S. mutans CSP encoding sequence
Competence Signal Peptide
AGCGGAAGCCTATCAACATTTTTCCGGCTGTTTAACAGAAGTTTTACACAAGCTTTGGG
AAAA [SEQ ID NO:1]

➢ S. mutans comD gene
Encodes a protein that functions as a histidine kinase receptor
[ATGAATGAAGCCTTAATGATACTTTCAAATGGTTTATTAACTTATCTAACCGTTCTAT
TTCTCTTGTTTCTATTTTCTAAGGTAAGTAATGTCACTTTATCGAAAAAGGAATTAACT
CTTTTTTCGATAAGCAATTTTCTGATAATGATTGCTGTTACGATGGTGAACGTAAACCT
GTTTTATCCTGCAGAGCCTCTTTATTTTATAGCTTTATCAATTTATCTTAATAGACAGA
ATAGTCTTTCTCTAAATATATTTTATGGTCTGCTGCCTGTTGCCAGTTCTGACTTGTTT
AGGCGGGCAATCATATTCTTTATCTTGGATGGAACTCAAGGAATTGTAATGGGCAGTAG
CATTATAACCACCTATATGATCGAGTTTGCAGGAATAGCGCTAAGTTACCTCTTTCTCA
GTGTGTTCAATGTTGATATTGGTCGACTTAAAGATAGTTTGACCAAGATGAAGGTCAAA
AAACGCTTGATTCCAATGAATATTACTATGCTTCTATACTACCTTTTAATACAGGTATT
GTATGTTATAGAGAGTTATAATGTGATACCGACTTTAAAATTTCGTAAATTTGTCGTTA
TTGTCTATCTTATTTTATTTTTGATTCTGATCTCATTTTTAAGCCAATATACCAAACAA
AAGGTTCAAAATGAGATAATGGCACAAAAGGAAGCTCAGATTCGAAATATCACCCAGTA
TAGTCAGCAAATAGAATCTCTTTACAAGGATATTCGAAGTTTCCGCCATGATTATCTGA
ATATTTTAACTAGCCTCAGATTAGGCATTGAAAATAAAGATTTAGCTAGTATTGAAAAG
ATTTACCATCAAATCTTAGAAAAAACAGGACATCAATTGCAGGATACCCGTTATAATAT
CGGCCATCTAGCTAATATTCAAAACGATGCTGTCAAGGGTATCTTGTCAGCAAAAATCT
TAGAAGCTCAGAATAAAAAGATTGCTGTCAATGTAGAAGTCTCAAGTAAAATACAACTG
CCTGAGATGGAGTTGCTTGATTTCATTACCATACTTTCTATCTTGTGTGATAATGCCAT
TGAGGCTGCTTTCGAATCATTAAATCCTGAAATTCAGTTAGCCTTTTTTAAGAAAAATG
GCAGTATAGTCTTTATCATTCAGAATTCCACCAAAGAAAAACAAATAGATGTGAGTAAA
ATTTTTAAAGAAACTATTCCACTAAAGGCTCCAATCGCGGTATTGGTTTAGCAAAGGTG
AATCATATTCTTGAACATTATCCCAAAACCAGTTTACAAACAAGCAATCATCATCATTT
ATTCAAGCAACTCCTAATAATAAAA]TAG

➢ S. mutans comE gene
Encodes a response regulator that activates transcription of a number of genes

[ATGATTTCTATTTTTGTATTGGAAGATGATTTTTTACAACAAGGACGTCTTGAAACCA
CCATTGCAGCTATCATGAAAGAAAAAAATTGGTCTTATAAAGAATTGACTATTTTTGGA
AAACCACAACAACTTATTGACGCTATCCCTGAAAAGGGCAATCACCAGATTTTCTTTTT
GGATATTGAAATCAAAAAAGAGGAAAAGAAAGGACTGGAAGTAGCCAATCAGATTAGAC
AGCATAATCCTAGTGCAGTTATTGTCTTTGTCACGACACATTCTGAGTTTATGCCCCTC
ACTTTTCAGTATCAGGTATCTGCTTTGGATTTTATTGATAAATCTTTGAATCCTGAGGA
GTTCTCCCACCGCATTGAATCAGCGCTGTATTATGCTATGGAAAACAGCCAGAAGAATG
GTCAATCAGAGGAACTTTTTATTTTCCATTCATCTGAAACTCAGTTTCAGGTCCCTTTT
GCTGAGATTCTGTATTTTGAAACATCTTCAACAGCCCATAAGCTCTGCCTTTATACTTA
TGATGAACGGATTGAATTCTACGGCAGTATGACTGACATTGTTAAAATGGATAAGAGAC
TTTTTCAGTGCCATCGCTCTTTTATTGTCAATCCTGCCAATATTACCCGTATTGATCGG
AAAAAACGCTTGGCCTATTTTCGAAATAATAAGTCTTGTCTTATTTCACGAACTAAGTT
AACAAAACTGAGAGCTGTGATTGCTGATCAAAGGAGAGCAAAA]TGA

Figure 3
The amino acid sequences of the signal peptide precursor (ComC), the histidine kinase (ComD), and the response regulator (ComE).

➢ S. mutans ComC protein (CSP Precursor)
MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK

➢ S. mutans ComD protein (Histidine Kinase)
MNEALMILSNGLLTYLTVLFLLFLFSKVSNVTLSKKELTLFSISNFLIMIAVTMVNVNL
FYPAEPLYFIALSIYLNRQNSLSLNIFYGLLPVASSDLFRRAIIFFILDGTQGIVMGSS
IITTYMIEFAGIALSYLFLSVFNVDIGRLKDSLTKMKVKKRLIPMNITMLLYYLLIQVL
YVIESYNVIPTLKFRKFVVIVYLILFLILISFLSQYTKQKVQNEIMAQKEAQIRNITQY
SQQIESLYKDIRSFRHDYLNILTSLRLGIENKDLASIEKIYHQILEKTGHQLQDTRYNI
GHLANIQNDAVKGILSAKILEAQNKKIAVNVEVSSKIQLPEMELLDFITILSILCDNAI
EAAFESLNPEIQLAFFKKNGSIVFIIQNSTKEKQIDVSKIFKENYSTKGSNRGIGLAKV
NHILEHYPKTSLQTSNHHHLFKQLLIIK

➢ S. mutans ComE protein (Response Regulator)
MISIFVLEDDFLQQGRLETTIAAIMKEKNWSYKELTIFGKPQQLIDAIPEKGNHQIFFL
DIEIKKEEKKGLEVANQIRQHNPSAVIVFVTTHSEFMPLTFQYQVSALDFIDKSLNPEE
FSHRIESALYYAMENSQKNGQSEELFIFHSSETQFQVPFAEILYFETSSTAHKLCLYTY
DERIEFYGSMTDIVKMDKRLFQCHRSFIVNPANITRIDRKKRLAYFRNNKSCLISRTKL
TKLRAVIADQRRAK

Figure 4
The deduced amino acid sequence of the signal peptide precursor in various strains and its predicted cleavage site. The original peptide is expressed as a 46-amino acid peptide that is cleaved after the glycine-glycine residues to generate an active signal peptide.

```
BM71 CSP      1 MKKTPSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
GB14 CSP      1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
H7 CSP        1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
JH1005 CSP    1 MKKTLSLKNDFKEIKTDELEIIIGGSGTLSTFFRLFNRSFTQA    43
LT11 CSP      1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
NG8 CSP       1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
UAB159 CSP    1 MKKTLSLKNDFKEIKTDELEIIIGGSGSLSTFFRLFNRSFTQALGK 46
                **** **********************.***************

consensus: 1 MKKTLSLKNDFKEIKTDELEIIIGG SGSLSTFFRLFNRSFTQALGK 46
predicted cleavage site:                ^
```

Figure 5
The synthetic signal peptide that is effective at inducing competence, biofilm formation and acid tolerance in *Streptococcus mutans*.

SGSLSTFFRLFNRSFTQALGK [SEQ ID NO:2]

Figure 6
The natural activity of the signal/receptor system functioning *in vitro* in model biofilms as determined by the ability of various strains of *S. mutans* to accept donor plasmid DNA conferring erythromycin resistance.

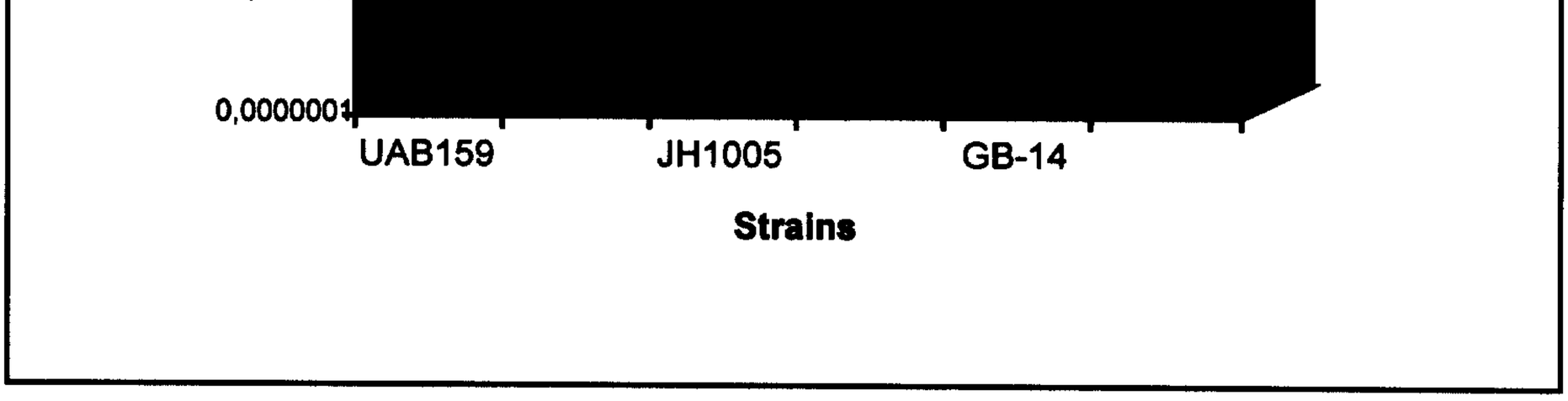


Figure 7
Induction of genetic transformation in *Streptococcus mutans* by synthetic competence stimulating peptide (SCSP)[1]

Strain	Peptide added Number of Transformants/Recipients	No peptide Number of Transformants/Recipients
UAB15	4.65×10^{-1}	1.78×10^{-6}
JH1005[2]	6.98×10^{-2}	0

[1]The final concentration of SCSP used was 500 ng/ml.
The strain contains a nonsense mutation in the *comC* gene encoding the CSP.

Figure 8
List of the primers used to amplify the genes or internal regions of the target genes by polymerase chain reaction (PCR) for subsequent sequencing or inactivation.

ComC region

ComC Primer Pair: F5-B5

[F5] 23406-23424 5'- AGTTTTTTGTCTGGCTGCG -3'
19 nt forward primer
pct G+C: 47.4 Tm: 50.5

[B5] 24056-24037 5'- TCCACTAAAGGCTCCAATCG -3'
20 nt backward primer
pct G+C: 50.0 Tm: 51.9

651 nt product for F5-B5 pair (23406-24056)
Optimal annealing temp: 50.3
pct G+C: 30.9 Tm: 71.5

ComD region

ComD Primer Pair: F1-B1

[F1] 392-415 5'- CGCTAAGTTACCTCTTTCTCAGTG -3'
24 nt forward primer
pct G+C: 45.8 Tm: 51.6

[B1] 683-663 5'- GCTTCCTTTTGTGCCATTATC -3'
21 nt backward primer
pct G+C: 42.9 Tm: 50.8

292 nt product for F1-B1 pair (392-683)
Optimal annealing temp: 49.5
pct G+C: 30.8 Tm: 70.2

ComE region

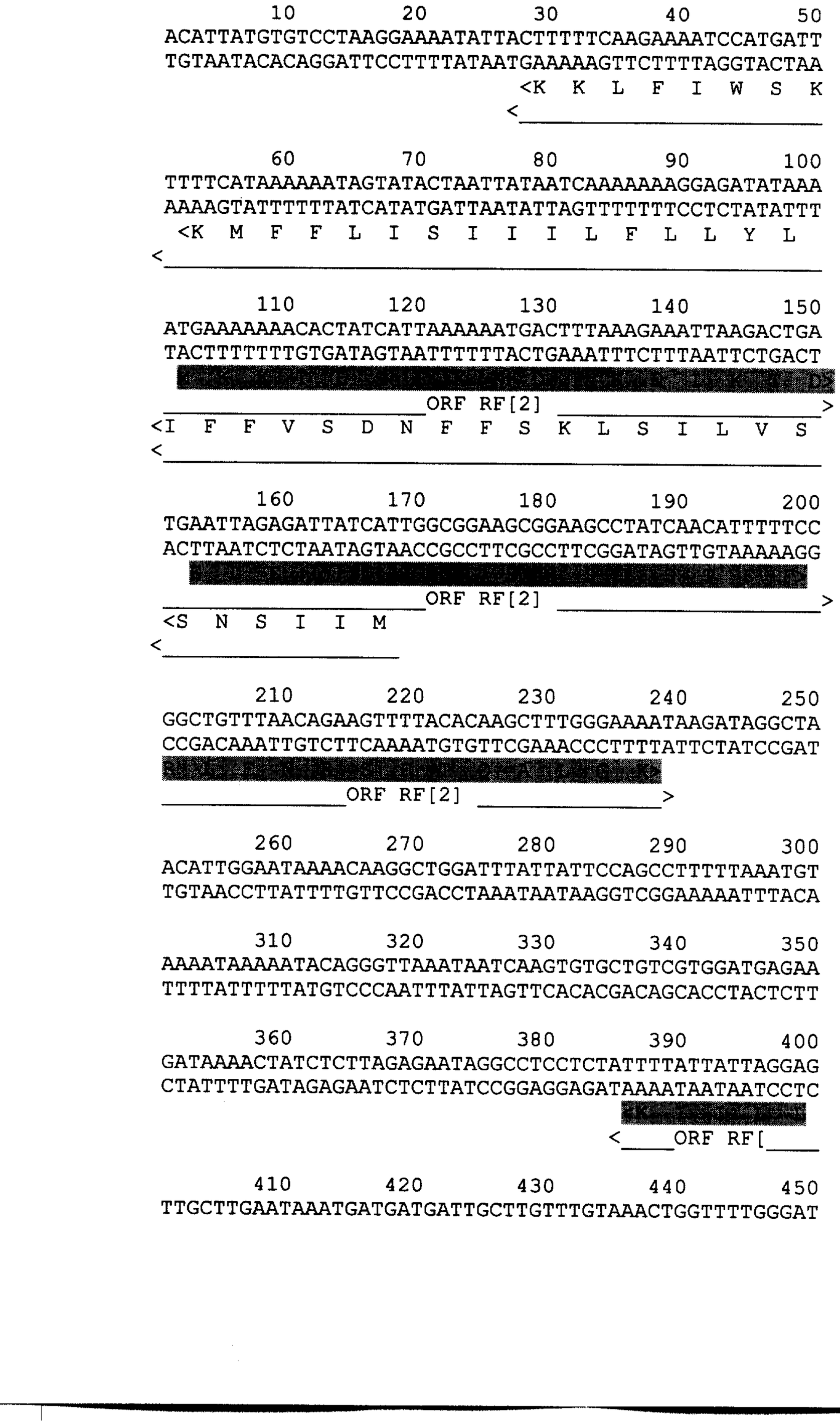
ComE Primer Pair: F1-B1

[F1] 145-165 5'- CCTGAAAAGGGCAATCACCAG -3'
21 nt forward primer
pct G+C: 52.4 Tm: 55.9

[B1] 606-585 5'- GCGATGGCACTGAAAAAGTCTC -3'
22 nt backward primer
pct G+C: 50.0 Tm: 55.4

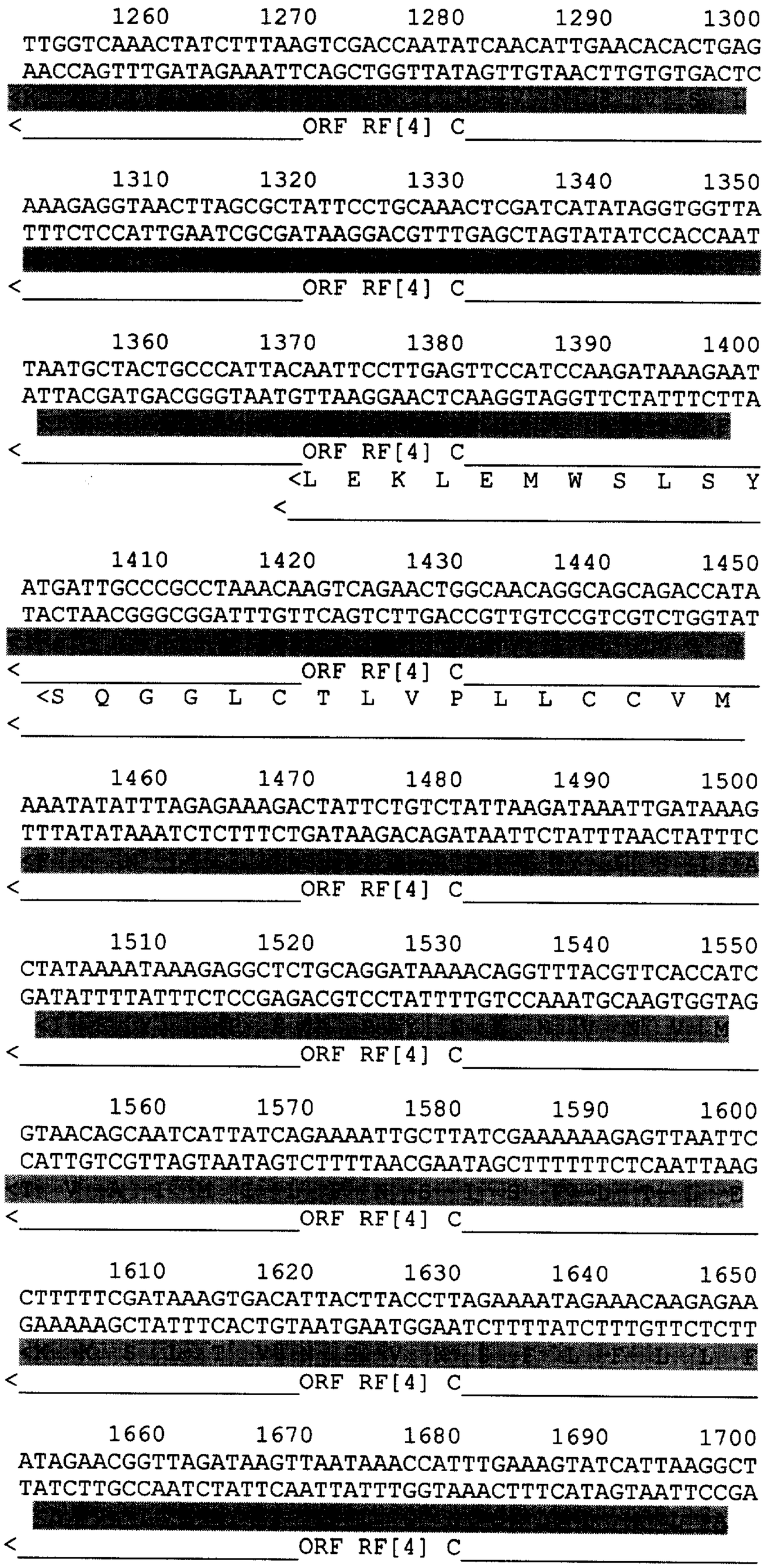
462 nt product for F1-B1 pair (145-606)
Optimal annealing temp: 53.6
pct G+C: 38.3 Tm: 74.1

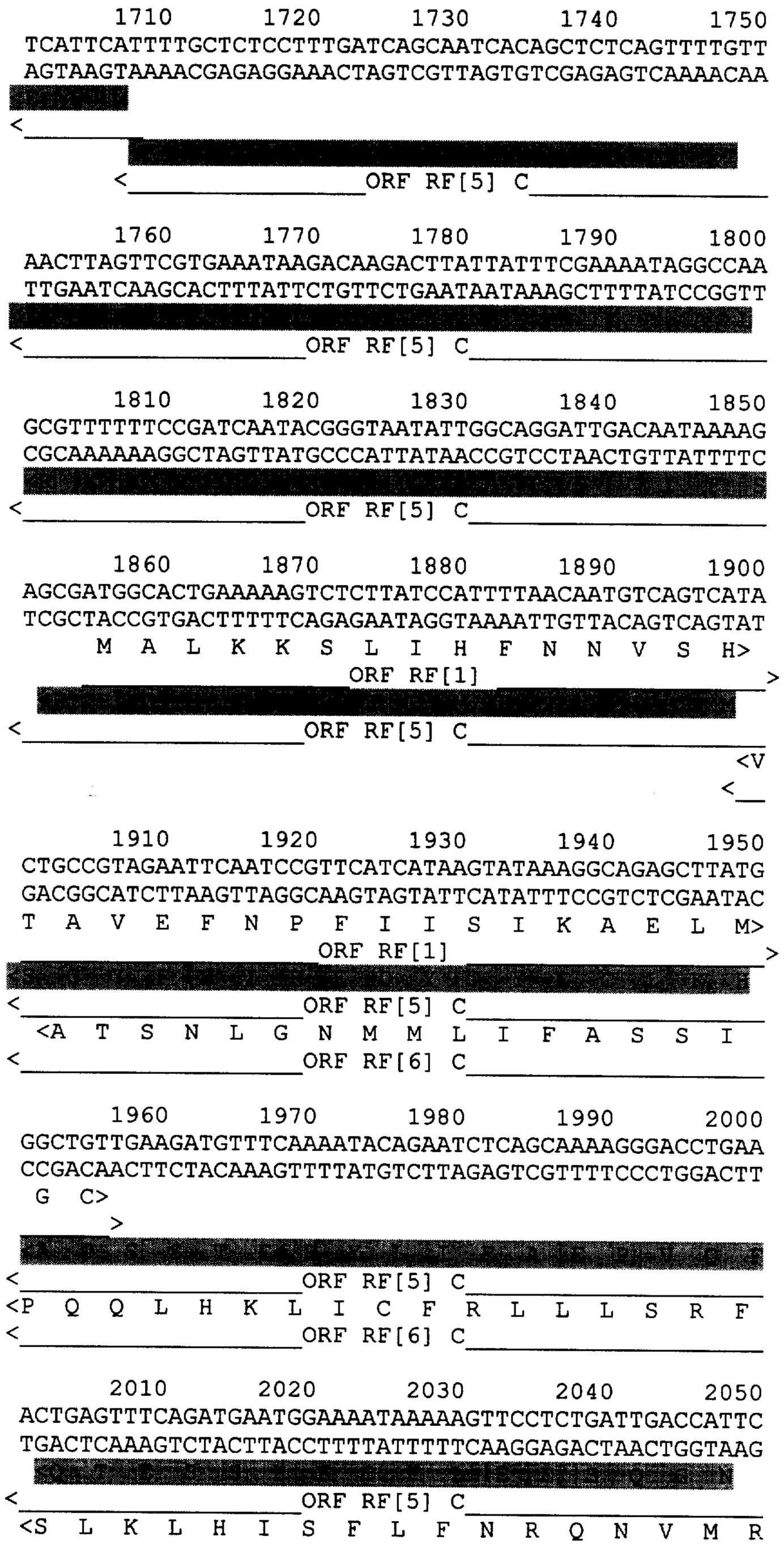
Figure 9
ComCDE local region. The ComC (first highlighted region), ComD (second highlighted region) and ComE (third highlighted region) proteins are highlighted.

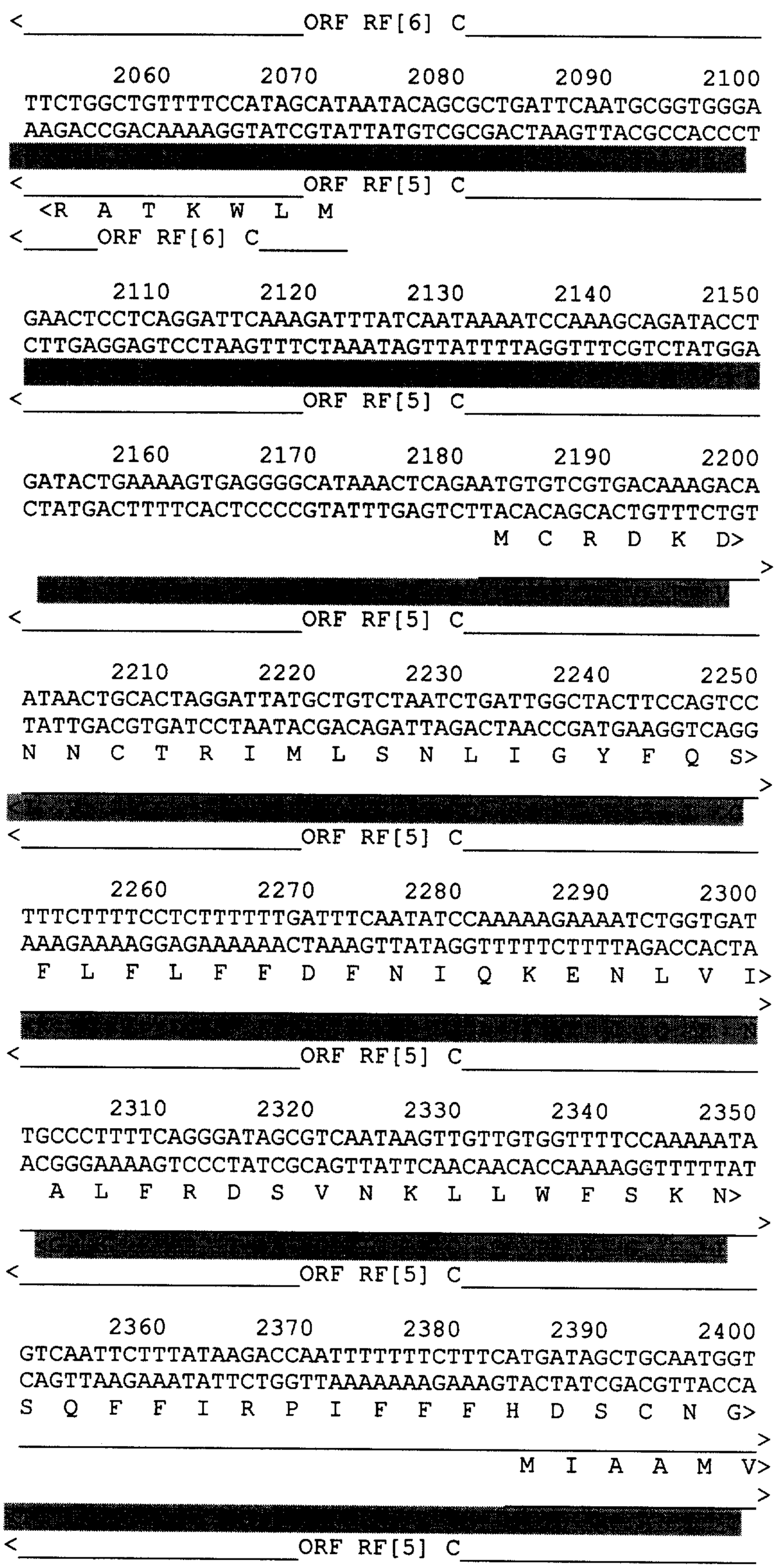


AACGAACTTATTTACTACTACTAACGAACAAACATTTGACCAAAACCCTA
ORF RF[4] C
460 470 480 490 500
AATGTTCAAGAATATGATTCACCTTTGCTAAACCAATACCGCGATTGGAG
TTACAAGTTCTTATACTAAGTGGAAACGATTTGGTTATGGCGCTAACCTC
<H E L I H N V K A L G I G R N S
ORF RF[4] C
510 520 530 540 550
CCTTTAGTGGAATAGTTTTCTTTAAAAATTTTACTCACATCTATTTGTTT
GGAAATCACCTTATCAAAAGAAATTTTTAAAATGAGTGTAGATAAACAAA
ORF RF[4] C
560 570 580 590 600
TTCTTTGGTGGAATTCTGAATGATAAAGACTATACTGCCATTTTTCTTAA
AAGAAACCACCTTAAGACTTACTATTTCTGATATGACGGTAAAAAGAATT
ORF RF[4] C
610 620 630 640 650
AAAAGGCTAACTGAATTTCAGGATTTAATGATTCGAAAGCAGCCTCAATG
TTTTCCGATTGACTTAAAGTCCTAAATTACTAAGCTTTCGTCGGAGTTAC
M>
ORF RF[4] C
660 670 680 690 700
GCATTATCACACAAGATAGAAAGTATGGTAATGAAATCAAGCAACTCCAT
CGTAATAGTGTGTTCTATCTTTCATACCATTACTTTAGTTCGTTGAGGTA
A L S H K I E S M V M K S S N S I>
ORF RF[3]
ORF RF[4] C
710 720 730 740 750
CTCAGGCAGTTGTATTTTACTTGAGACTTCTACATTGACAGCAATCTTTT
GAGTCCGTCAACATAAAATGAACTCTGAAGATGTAACTGTCGTTAGAAAA
S G S C I L L E T S T L T A I F>
ORF RF[3]
ORF RF[4] C
760 770 780 790 800
TATTCTGAGCTTCTAAGATTTTTGCTGACAAGATACCCTTGACAGCATCG
ATAAGACTCGAAGATTCTAAAAACGACTGTTCTATGGGAACTGTCGTAGC
L F>
ORF RF[4] C
810 820 830 840 850
TTTTGAATATTAGCTAGATGGCCGATATTATAACGGGTATCCTGCAATTG
AAAACTTATAATCGATCTACCGGCTATAATATTGCCCATAGGACGTTAAC


ORF RF[4] C
860 870 880 890 900
ATGTCCTGTTTTTTCTAAGATTTGATGGTAAATCTTTTCAATACTAGCTA
TACAGGACAAAAAAGATTCTAAACTACCATTTAGAAAAGTTATGATCGAT
ORF RF[4] C
910 920 930 940 950
AATCTTTATTTTCAATGCCTAATCTGAGGCTAGTTAAAATATTCAGATAA
TTAGAAATAAAAGTTACGGATTAGACTCCGATCAATTTTATAAGTCTATT
ORF RF[4] C
960 970 980 990 1000
TCATGGCGGAAACTTCGAATATCCTTGTAAAGAGATTCTATTTGCTGACT
AGTACCGCCTTTGAAGCTTATAGGAACATTTCTCTAAGATAAACGACTGA
M A E T S N I L V K R F Y L L T>
ORF RF[4] C
1010 1020 1030 1040 1050
ATACTGGGTGATATTTCGAATCTGAGCTTCCTTTTGTGCCATTATCTCAT
TATGACCCACTATAAAGCTTAGACTCGAAGGAAAACACGGTAATAGAGTA
I L G D I S N L S F L L C H Y L I>
ORF RF[4] C
1060 1070 1080 1090 1100
TTTGAACCTTTTGTTTGGTATATTGGCTTAAAAATGAGATCAGAATCAAA
AAACTTGGAAAACAAACCATATAACCGAATTTTTACTCTAGTCTTAGTTT
L N L L F G I L A>
ORF RF[4] C
1110 1120 1130 1140 1150
AATAAAATAAGATAGACAATAACGACAAATTTACGAAATTTTAAAGTCGG
TTATTTTATTCTATCTGTTATTGCTGTTTAAATGCTTTAAAATTTCAGCC
ORF RF[4] C
1160 1170 1180 1190 1200
TATCACATTATAACTCTCTATAACATACAATACCTGTATTAAAAGGTAGT
ATAGTGTAATATTGAGAGATATTGTATGTTATGGACATAATTTTCCATCA
ORF RF[4] C
1210 1220 1230 1240 1250
ATAGAAGCATAGTAATATTCATTGGAATCAAGCGTTTTTTGACCTTCATC
TATCTTCGTATCATTATAAGTAACCTTAGTTCGCAAAAAACTGGAAGTAG
ORF RF[4] C


1260 1270 1280 1290 1300
TTGGTCAAACTATCTTTAAGTCGACCAATATCAACATTGAACACACTGAG
AACCAGTTTGATAGAAATTCAGCTGGTTATAGTTGTAACTTGTGTGACTC
<ORF RF[4] C
1310 1320 1330 1340 1350
AAAGAGGTAACTTAGCGCTATTCCTGCAAACTCGATCATATAGGTGGTTA
TTTCTCCATTGAATCGCGATAAGGACGTTTGAGCTAGTATATCCACCAAT
<ORF RF[4] C
1360 1370 1380 1390 1400
TAATGCTACTGCCCATTACAATTCCTTGAGTTCCATCCAAGATAAAGAAT
ATTACGATGACGGGTAATGTTAAGGAACTCAAGGTAGGTTCTATTTCTTA
<ORF RF[4] C
<L E K L E M W S L S Y
<
1410 1420 1430 1440 1450
ATGATTGCCCGCCTAAACAAGTCAGAACTGGCAACAGGCAGCAGACCATA
TACTAACGGGCGGATTTGTTCAGTCTTGACCGTTGTCCGTCGTCTGGTAT
<ORF RF[4] C
<S Q G G L C T L V P L L C C V M
<
1460 1470 1480 1490 1500
AAATATATTTAGAGAAAGACTATTCTGTCTATTAAGATAAATTGATAAAG
TTTATATAAATCTCTTTCTGATAAGACAGATAATTCTATTTAACTATTTC
<ORF RF[4] C
1510 1520 1530 1540 1550
CTATAAAATAAAGAGGCTCTGCAGGATAAAACAGGTTTACGTTCACCATC
GATATTTTATTTCTCCGAGACGTCCTATTTTGTCCAAATGCAAGTGGTAG
<ORF RF[4] C
1560 1570 1580 1590 1600
GTAACAGCAATCATTATCAGAAAATTGCTTATCGAAAAAAGAGTTAATTC
CATTGTCGTTAGTAATAGTCTTTTAACGAATAGCTTTTTTCTCAATTAAG
<ORF RF[4] C
1610 1620 1630 1640 1650
CTTTTTCGATAAAGTGACATTACTTACCTTAGAAAATAGAAACAAGAGAA
GAAAAAGCTATTTCACTGTAATGAATGGAATCTTTTATCTTTGTTCTCTT
<ORF RF[4] C
1660 1670 1680 1690 1700
ATAGAACGGTTAGATAAGTTAATAAACCATTTGAAAGTATCATTAAGGCT
TATCTTGCCAATCTATTCAATTATTTGGTAAACTTTCATAGTAATTCCGA
<ORF RF[4] C


1710 1720 1730 1740 1750
TCATTCATTTTGCTCTCCTTTGATCAGCAATCACAGCTCTCAGTTTTGTT
AGTAAGTAAAACGAGAGGAAACTAGTCGTTAGTGTCGAGAGTCAAAACAA
ORF RF[5] C
1760 1770 1780 1790 1800
AACTTAGTTCGTGAAATAAGACAAGACTTATTATTTCGAAAATAGGCCAA
TTGAATCAAGCACTTTATTCTGTTCTGAATAATAAAGCTTTTATCCGGTT
ORF RF[5] C
1810 1820 1830 1840 1850
GCGTTTTTTCCGATCAATACGGGTAATATTGGCAGGATTGACAATAAAAG
CGCAAAAAAGGCTAGTTATGCCCATTATAACCGTCCTAACTGTTATTTTC
ORF RF[5] C
1860 1870 1880 1890 1900
AGCGATGGCACTGAAAAAGTCTCTTATCCATTTTAACAATGTCAGTCATA
TCGCTACCGTGACTTTTTCAGAGAATAGGTAAAATTGTTACAGTCAGTAT
M A L K K S L I H F N N V S H>
ORF RF[1]
ORF RF[5] C
1910 1920 1930 1940 1950
CTGCCGTAGAATTCAATCCGTTCATCATAAGTATAAAGGCAGAGCTTATG
GACGGCATCTTAAGTTAGGCAAGTAGTATTCATATTTCCGTCTCGAATAC
T A V E F N P F I I S I K A E L M>
ORF RF[1]
ORF RF[5] C
<A T S N L G N M M L I F A S S I
ORF RF[6] C
1960 1970 1980 1990 2000
GGCTGTTGAAGATGTTTCAAAATACAGAATCTCAGCAAAAGGGACCTGAA
CCGACAACTTCTACAAAGTTTTATGTCTTAGAGTCGTTTTCCCTGGACTT
G C>
ORF RF[5] C
<P Q Q L H K L I C F R L L L S R F
ORF RF[6] C
2010 2020 2030 2040 2050
ACTGAGTTTCAGATGAATGGAAAATAAAAAGTTCCTCTGATTGACCATTC
TGACTCAAAGTCTACTTACCTTTTATTTTTCAAGGAGACTAACTGGTAAG
ORF RF[5] C
<S L K L H I S F L F N R Q N V M R


ORF RF[6] C
2060 2070 2080 2090 2100
TTCTGGCTGTTTTCCATAGCATAATACAGCGCTGATTCAATGCGGTGGGA
AAGACCGACAAAAGGTATCGTATTATGTCGCGACTAAGTTACGCCACCCT
ORF RF[5] C
<R A T K W L M
ORF RF[6] C
2110 2120 2130 2140 2150
GAACTCCTCAGGATTCAAAGATTTATCAATAAAATCCAAAGCAGATACCT
CTTGAGGAGTCCTAAGTTTCTAAATAGTTATTTTAGGTTTCGTCTATGGA
ORF RF[5] C
2160 2170 2180 2190 2200
GATACTGAAAAGTGAGGGGCATAAACTCAGAATGTGTCGTGACAAAGACA
CTATGACTTTTCACTCCCCGTATTTGAGTCTTACACAGCACTGTTTCTGT
M C R D K D>
ORF RF[5] C
2210 2220 2230 2240 2250
ATAACTGCACTAGGATTATGCTGTCTAATCTGATTGGCTACTTCCAGTCC
TATTGACGTGATCCTAATACGACAGATTAGACTAACCGATGAAGGTCAGG
N N C T R I M L S N L I G Y F Q S>
ORF RF[5] C
2260 2270 2280 2290 2300
TTTCTTTTCCTCTTTTTTGATTTCAATATCCAAAAAGAAAATCTGGTGAT
AAAGAAAAGGAGAAAAAACTAAAGTTATAGGTTTTTCTTTTAGACCACTA
F L F L F F D F N I Q K E N L V I>
ORF RF[5] C
2310 2320 2330 2340 2350
TGCCCTTTTCAGGGATAGCGTCAATAAGTTGTTGTGGTTTTCCAAAAATA
ACGGGAAAAGTCCCTATCGCAGTTATTCAACAACACCAAAAGGTTTTTAT
A L F R D S V N K L L W F S K N>
ORF RF[5] C
2360 2370 2380 2390 2400
GTCAATTCTTTATAAGACCAATTTTTTTCTTTCATGATAGCTGCAATGGT
CAGTTAAGAAATATTCTGGTTAAAAAAAGAAAGTACTATCGACGTTACCA
S Q F F I R P I F F F H D S C N G>
M I A A M V>
ORF RF[5] C

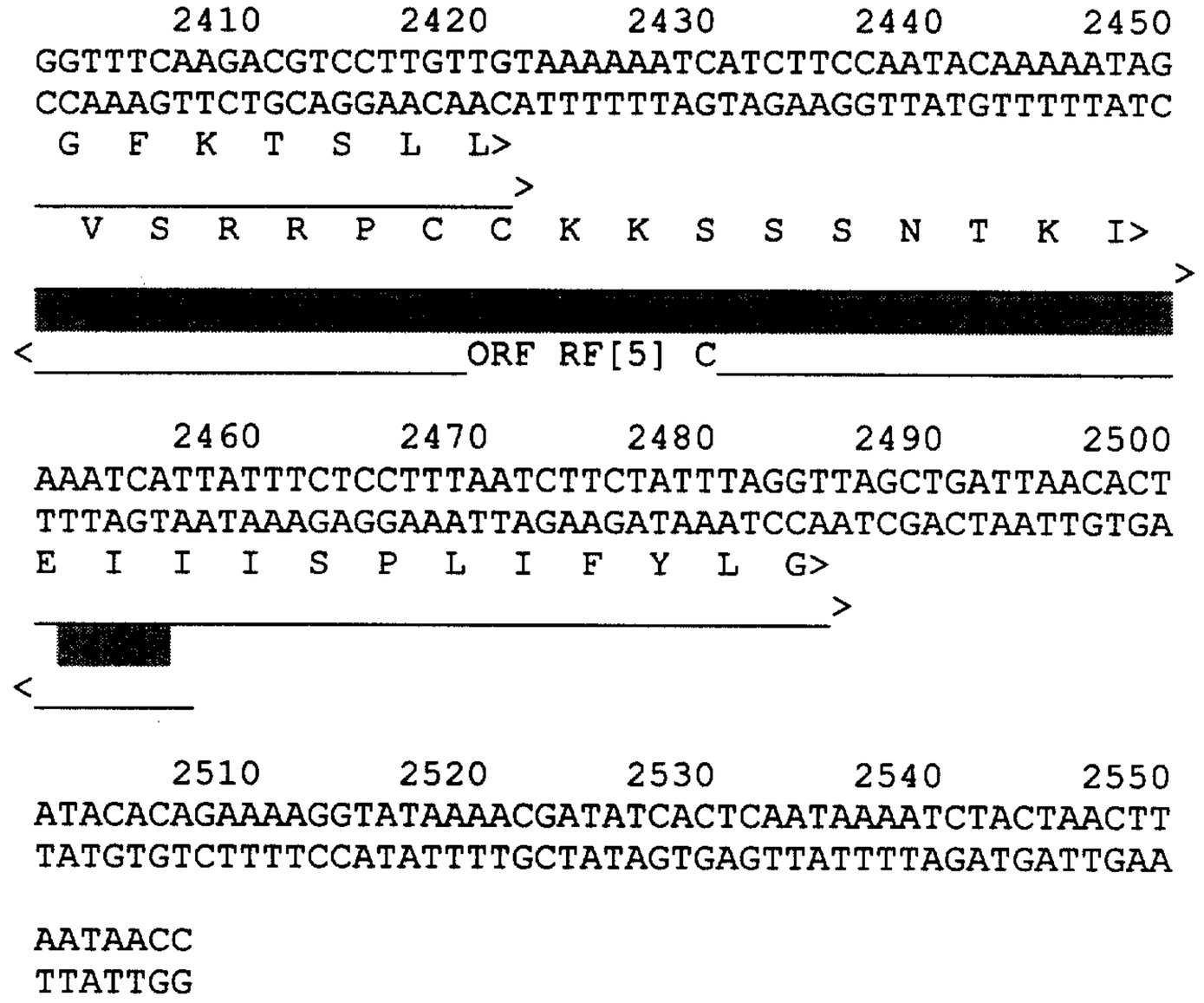

2410 2420 2430 2440 2450
GGTTTCAAGACGTCCTTGTTGTAAAAAATCATCTTCCAATACAAAAATAG
CCAAAGTTCTGCAGGAACAACATTTTTTAGTAGAAGGTTATGTTTTTATC
G F K T S L L>
V S R R P C C K K S S S N T K I>
ORF RF[5] C
2460 2470 2480 2490 2500
AAATCATTATTTCTCCTTTAATCTTCTATTTAGGTTAGCTGATTAACACT
TTTAGTAATAAAGAGGAAATTAGAAGATAAATCCAATCGACTAATTGTGA
E I I I S P L I F Y L G>
2510 2520 2530 2540 2550
ATACACAGAAAAGGTATAAAACGATATCACTCAATAAAATCTACTAACTT
TATGTGTCTTTTCCATATTTTGCTATAGTGAGTTATTTTAGATGATTGAA
AATAACC
TTATTGG

Figure 10

The comX nucleotide sequence, amino acid sequence, and its local region with 100bp included both upstream and downstream (promoter is upstream).

➢ S. mutans comX gene

ATGGAAGAAGATTTTGAAATTGTTTTTAATAAGGTTAAGCCAATTGTATGGAAATTAAG
CCGTTATTACTTTATTAAAATGTGGACTCGTGAAGATTGGCAACAAGAGGGAATGTTGA
TTTTGCACCAATTATTAAGGGAACATCCAGAATTAGAAGAGGATGATACAAAATTGTAT
ATCTATTTTAAGACACGTTTTTCTAATTACATTAAAGATGTTTTGCGTCAGCAAGAAAG
TCAGAAACGTCGTTTTAATAGAATGTCTTATGAAGAAGTCGGTGAGATTGAACACTGTT
TGTCAAGTGGCGGTATGCAATTGGATGAATATATTTTATTTCGTGATAGTTTGCTTGCA
TATAAACAAGGTCTGAGTACTGAAAAGCAAGAGCTGTTTGAGCGCTTGGTAGCAGGAGA
GCACTTTTTGGGAAGGCAAAGTATGCTGAAAGATTTACGTAAAAAATTAAGTGATTTTA
AGGAAAAA

➢ S. mutans ComX protein

MEEDFEIVFNKVKPIVWKLSRYYFIKMWTREDWQQEGMLILHQLLREHPELEEDDTKLY
IYFKTRFSNYIKDVLRQQESQKRRFNRMSYEEVGEIEHCLSSGGMQLDEYILFRDSLLA
YKQGLSTEKQELFERLVAGEHFLGRQSMLKDLRKKLSDFKEK

➢ S. mutans comX gene local region

GTAAATAAAACAGCCAGTTAAGATGGGACATTTATGTCCTGTTCTTAAAGTCTTTTTCG
TTTTATAATAATTTTATTATAAAAGGAGGTCATCGTAATAGATGGAAGAAGATTTTGAA
ATTGTTTTTAATAAGGTTAAGCCAATTGTATGGAAATTAAGCCGTTATTACTTTATTAA
AATGTGGACTCGTGAAGATTGGCAACAAGAGGGAATGTTGATTTTGCACCAATTATTAA
GGGAACATCCAGAATTAGAAGAGGATGATACAAAATTGTATATCTATTTTAAGACACGT
TTTTCTAATTACATTAAAGATGTTTTGCGTCAGCAAGAAAGTCAGAAACGTCGTTTTAA
TAGAATGTCTTATGAAGAAGTCGGTGAGATTGAACACTGTTTGTCAAGTGGCGGTATGC
AATTGGATGAATATATTTTATTTCGTGATAGTTTGCTTGCATATAAACAAGGTCTGAGT
ACTGAAAAGCAAGAGCTGTTTGAGCGCTTGGTAGCAGGAGAGCACTTTTTGGGAAGGCA
AAGTATGCTGAAAGATTTACGTAAAAAATTAAGTGATTTTAAGGAAAAATAGTTAAAAA
GGGAAAGAATGGAACATGTGATTGTACCATTCTTTTTGGTTGAAAATTAAGAAAAGTTA
TTATAAATTATTGGTTTAACATGCCATATTA

Figure 11
The comA and comB nucleotide and amino acid sequences. ComA and ComB are the components of the CSP exporter.

➢ S. mutans comA gene

ATGAAACAAGTTATTTATGTTGTTTTAATCGTCATAGCCGTTAACATTCTCTTAGAGAT
TATCAAAAGAGTAACAAAAAGGGGAGGGACAGTTTCGTCATCTAATCCTTTACCAGATG
GGCAGTCTAAGTTGTTTTGGCGCAGACATTATAAGCTAGTACCTCAGATTGATACCAGA
GACTGTGGGCCGGCAGTGCTGGCATCTGTTGCAAAGCATTACGGATCTAATTACTCTAT
CGCTTATCTGCGGGAACTCTCAAAGACTAACAAGCAGGGAACAACAGCTCTTGGCATTG
TTGAAGCTGCTAAAAAGTTAGGCTTTGAAACACGCTCTATCAAGGCGGATATGACGCTT
TTTGATTATAATGATTTGACCTATCCTTTTATCGTCCATGTGATTAAAGGAAAACGTCT
GCAGCATTATTATGTCGTCTATGGCAGCCAGAATAATCAGCTGATTATTGGAGATCCTG
ATCCTTCAGTTAAGGTGACTAGGATGAGTAAGGAACGCTTTCAATCAGAGTGGACAGGC
CTTGCAATTTTCCTAGCTCCTCAGCCTAACTATAAGCCTCATAAAGGTGAAAAAAATGG
TTTGTCTAATTTCTTCCCGTTGATCTTTAAGCAGAAAGCTTTGATGACTTATATTATCA
TAGCTAGCTTGATTGTGACGCTCATTGATATTGTCGGATCATACTATCTCCAAGGAATA
TTGGACGAGTACATTCCTGATCAGCTGATTTCAACTTTAGGAATGATTACGATTGGTCT
GATAATAACCTATATTATCCAGCAGGTCATGGCTTTTGCAAAAGAATACCTCTTGGCCG
TACTCAGTTTGCGTTTAGTCATTGATGTTATCCTGTCTTATATCAAACATATTTTTACG
CTTCCTATGTCTTTCTTTGCGACAAGGCGAACAGGAGAAATCACGTCTCGTTTTACAGA
TGCCAATCAGATTATTGATGCTGTAGCGTCAACCATCTTTTCAATCTTTTTAGATATGA
CTATGGTAATTTTGGTTGGTGGGGTTTTGTTGGCGCAAAACAATAACCTTTTCTTTCTA
ACCTTGCTCTCCATTCCGATTTATGCCATCATTATTTTTGCTTTCTTGAAACCCTTTGA
GAAAATGAATCACGAAGTGATGGAAAGCAATGCTGTGGTAAGTTCTTCTATCATTGAAG
ATATCAATGGGATGGAAACCATTAAATCACTCACAAGTGAGTCCGCTCGTTATCAAAAC
ATTGATAGTGAATTTGTTGATTATTTGGAGAAAAACTTTAAGCTACACAAGTATAGTGC
CATTCAAACCGCATTAAAAAGCGGTGCTAAGCTTATCCTCAATGTTGTCATTCTCTGGT
ATGGCTCTCGTCTAGTTATGGATAATAAAATCTCAGTTGGTCAGCTTATCACCTTTAAT
GCTTTGCTGTCTTATTTCTCAAATCCAATTGAAAATATTATCAATCTGCAATCCAAACT
GCAGTCAGCTCGCGTTGCCAATACACGTCTTAATGAGGTCTATCTTGTCGAATCTGAAT
TTGAAAAAGACGGCGATTTATCAGAAAATAGCTTTTTAGATGGTGATATTTCGTTTGAA
AATCTTTCTTATAAATATGGATTTGGGCGAGATACCTTATCAGATATTAATTTATCAAT
CAAAAAAGGCTCCAAGGTCAGTCTAGTTGGAGCCAGTGGTTCTGGTAAAACAACTTTGG
CTAAACTGATTGTCAATTTCTACGAGCCTAACAAGGGGATTGTTCGAATCAATGGCAAT
GATTTAAAAGTTATTGATAAGACAGCTTTGCGGCGGCATATTAGCTATTTGCCGCAACA
GGCCTATGTTTTTAGTGGCTCTATTATGGATAATCTCGTTTTAGGAGCTAAAGAAGGAA
CGAGTCAGGAAGACATTATTCGTGCTTGTGAAATTGCTGAAATCCGCTCGGACATTGAA
CAAATGCCTCAGGGCTATCAGACAGAGTTATCAGATGGTGCCGGTATTTCTGGCGGTCA
AAAACAGCGGATTGCTTTAGCTAGGGCCTTATTAACACAGGCACCGGTTTTGATTCTGG
ATGAAGCCACCAGCAGTCTTGATATTTTGACAGAAAAGAAAATTATCAGCAATCTCTTA
CAGATGACGGAGAAAACAATAATTTTTGTTGCCCACCGCTTAAGCATTTCACAGCGTAC
TGACGAAGTCATTGTCATGGATCAGGGAAAAATTGTTGAACAAGGCACTCATAAGGAAC
TTTTAGCTAAGCAAGGTTTCTATTATAACCTGTTTAAT

➢ S. mutans ComA protein

MKQVIYVVLIVIAVNILLEIIKRVTKRGGTVSSSNPLPDGQSKLFWRRHYKLVPQIDTR
DCGPAVLASVAKHYGSNYSIAYLRELSKTNKQGTTALGIVEAAKKLGFETRSIKADMTL
FDYNDLTYPFIVHVIKGKRLQHYYVVYGSQNNQLIIGDPDPSVKVTRMSKERFQSEWTG
LAIFLAPQPNYKPHKGEKNGLSNFFPLIFKQKALMTYIIIASLIVTLIDIVGSYYLQGI
LDEYIPDQLISTLGMITIGLIITYIIQQVMAFAKEYLLAVLSLRLVIDVILSYIKHIFT
LPMSFFATRRTGEITSRFTDANQIIDAVASTIFSIFLDMTMVILVGGVLLAQNNNLFFL
TLLSIPIYAIIIFAFLKPFEKMNHEVMESNAVVSSSIIEDINGMETIKSLTSESARYQN
IDSEFVDYLEKNFKLHKYSAIQTALKSGAKLILNVVILWYGSRLVMDNKISVGQLITFN
ALLSYFSNPIENIINLQSKLQSARVANTRLNEVYLVESEFEKDGDLSENSFLDGDISFE
NLSYKYGFGRDTLSDINLSIKKGSKVSLVGASGSGKTTLAKLIVNFYEPNKGIVRINGN
DLKVIDKTALRRHISYLPQQAYVFSGSIMDNLVLGAKEGTSQEDIIRACEIAEIRSDIE
QMPQGYQTELSDGAGISGGQKQRIALARALLTQAPVLILDEATSSLDILTEKKIISNLL
QMTEKTIIFVAHRLSISQRTDEVIVMDQGKIVEQGTHKELLAKQGFYYNLFN

➢ S. mutans comB gene

ATGGATCCTAAATTTTTACAAAGTGCAGAATTTTATAGGAGACGCTATCATAATTTTGC
GACACTATTAATTGTTCCTTTGGTCTGCTTGATTATCTTCTTGGTCATATTCCTTTGTT
TTGCTAAAAAAGAAATTACAGTGATTTCTACTGGTGAAGTTGCACCAACAAAGGTTGTA
GATGTTATCCAATCTTACAGTGACAGTTCAATCATTAAAAATAATTTAGATAATAATGC
AGCTGTTGAGAAGGGAGACGTTTTAATTGAATATTCAGAAAATGCCAGTCCAAACCGTC
AGACTGAACAAAAGAATATTATAAAAGAAAGACAAAAACGAGAAGAGAAGGAAAAGAAA
AAACACCAAAAGAGCAAGAAAAAGAAGAAGTCTAAGAGCAAGAAAGCTTCCAAAGATAA
GAAAAAGAAATCGAAAGACAAGGAAAGCAGCTCTGACGATGAAAATGAGACAAAAAAGG
TTTCGATTTTTGCTTCAGAAGATGGTATTATTCATACCAATCCCAAATATGATGGTGCC
AATATTATTCCGAAGCAAACCGAGATTGCTCAAATCTATCCTGATATTCAAAAAACAAG
AAAAGTGTTAATCACCTATTATGCTTCTTCTGATGATGTTGTTTCTATGAAAAAGGGGC
AAACCGCTCGTCTTTCCTTGGAAAAAAAGGGAAATGACAAGGTTGTTATTGAAGGAAAA
ATTAACAATGTCGCTTCATCAGCAACTACTACTAAAAAAGGAAATCTCTTTAAGGTTAC
TGCCAAAGTAAAGGTTTCTAAGAAAAATAGCAAACTCATCAAGTATGGTATGACAGGCA
AGACAGTCACTGTCATTGATAAAAAGACTTATTTTGATTATTTCAAAGATAAATTACTG
CATAAAATGGATAAT

➢ S. mutans ComB protein

MDPKFLQSAEFYRRRYHNFATLLIVPLVCLIIFLVIFLCFAKKEITVISTGEVAPTKVV
DVIQSYSDSSIIKNNLDNNAAVEKGDVLIEYSENASPNRQTEQKNIIKERQKREEKEKK
KHQKSKKKKKSKSKKASKDKKKKSKDKESSSDDENETKKVSIFASEDGIIHTNPKYDGA
NIIPKQTEIAQIYPDIQKTRKVLITYYASSDDVVSMKKGQTARLSLEKKGNDKVVIEGK
INNVASSATTTKKGNLFKVTAKVKVSKKNSKLIKYGMTGKTVTVIDKKTYFDYFKDKLL
HKMDN

Schematic layout of the arrangement of the genetic locus encoding the signal peptide precursor, the histidine kinase and the response regulator. Note that this arrangement is different from other loci in related streptococci for the following reasons: a) The comC gene is transcribed from its own promoter alone, unlike the genes thus far described in other streptococci that are arranged in an operon-like cluster with the comC/DE genes being transcribed from a single promoter.
b) The comC gene is separated from the *comD* gene by 148 nucleotides.

Streptococcus mutans
***ComCDE* Operon**

