INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

SEQUENCE AND ANALYSIS OF *ACTINOBACILLUS ACTINOMYCETEMCOMITANS*

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirement for the

Degree of

Doctor of Philosophy

By

FARES ZOHIR NAJAR

Norman, Oklahoma

2002

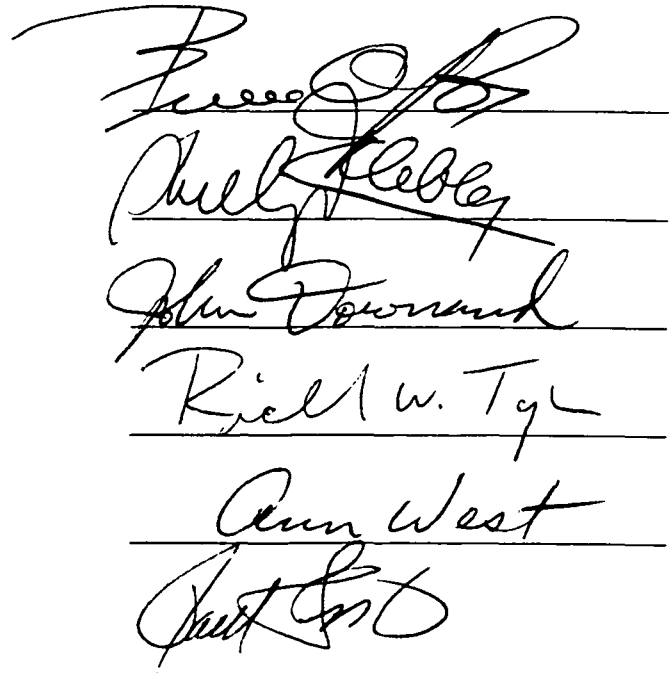
ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

SEQUENCE AND ANALYSIS OF

ACTINOBACILLUS ACTINOMYCETEMCOMITANS

A Dissertation APPROVED FOR THE

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY



Acknowledgments

I would like to thank my advisory committee, Drs. Bruce Roe, Phillip Klebba, Ann West, Richard Taylor, and John Downard for their support of my efforts to complete my dissertation work. A great deal of gratidute goes to Dr. Paul Cook for all the time and help he provided to help me understand metabolism. A very special thanks and my gratitude goes to Dr. Bruce Roe, my major professor, for his infinite support, encouragement, guidance, and patience throughout my graduate career.

I am also thankful for every member of Dr. Roe's laboratory at the Advanced Center for Genome Technology for their help and support they provided including the wonderful administrative staff, the gel room crew and the fellow graduate students. A special thanks goes to Dr. Doris Kupfer for her help, support, and invigorating scientific discussions beyond the scope of this dissertation, and to Hongshing Lai who's computer expertise was instrumental and pivotal in providing the annotation needed for the analysis of *A. actinomycetemcomitans* genes. To Jim White, who wrote and developed various scripts that helped automate and speedup analysis of the massive amount of data generated and for showing me that PERL is the salvation. My thanks also to Steve Kenton who provided an enormous help and insight to the sequence information of *A. actinomycetemcomitans* genome. To Douglas White for providing me with much computer advice that helped in my final data presentation. And I would like to thank Shaoping Lin for her wonderful insights to the world of PCR. I also would like to thank Dr. Sandra Clifton for the training she provided when I first joined the lab.

I would like to thank my family for their love, support, and advise. I simply cannot imagine this work without them on my side. My parents, for constantly pushing me to pursue my higher education. As always, I am forever thankful for them. To my sister, Ranya, who pretended to listen to me while explaining my research to her. To Bassel, my brother, my friend, and my confidant for being there when I needed him. To them I dedicate this dissertation.

Table of Content

List of Tables

List of figures

Abstract

A. actinomycetemcomitans is a Gram negative coccobacillus, facultative anaerobe that is implicated in the persistence of peridontitis, the destruction of the gum tissues and loss of teeth (**Fives-Taylor et al, 1999**). In an effort to understand the biochemical events that result in *A. actinomycetemcomitans* induced oral cavity disease, its genome was sequenced. To date over 99.8% of this ~2 MB genome have been sequenced. Analysis of this data reveals much new information about the biochemical metabolic pathways and the pathogenesis of *A. actinomycetemcomitans*.

Multiple sequencing approaches were utilized to obtain the highest accuracy of sequence data. Initially, the preliminary sequence data were collected by the shotgun sequencing method, and then, since no physical map was available for this bacterium, a multiplex PCR sequencing technique was developed to aid in gap closure and finishing. Once six genome equivalents of the sequence were obtained, the *A. actinomycetemcomitans* genome sequence was analyzed to obtain the predicted gene positions, their functions, and their structural motifs. With this information, PCR-based methods were used for sequencing gap closure and a detailed metabolic profile of *A. actinomycetemcomitans* was determined. Through homology and motif analysis, putative *A. actinomycetemcomitans* virulence factors were determined revealing a potential mode of pathogenesis.

Analysis of the predicted intermediary metabolism genes in *A. actinomycetemcomitans* reveals that this bacterium can synthesize all amino acids except leucine, histidine, and methionine and that the complete glycolytic and pentose phosphate

pathways are present to allow for production of both the reducing potentials needed for energy and for the different metabolites needed as substrates for other biosynthetic pathways. These pathways in addition to the several encoded carbohydrate transporters, allow the organism to utilize different carbohydrates from the environment as an energy sources. In contrast, fatty acids cannot be utilized from the environment by *A. acinomycetemcomitams* since several of the enzymes required for β-oxidation are not present in the genome, and the synthesis of both purines and pyrimidines only can occur via their salvage pathways. In contrast, the genes coding for the biosynthesis of several vitamins and cofactors, such as biotin, riboflavin, and folate, were present in *A. actinomycetemcomitans* genome.

During these studies, several pathogenicity-associated metabolic pathways also were identified. These include the LPS and dTDP-L-rhamanose pathways. In addition, at least 12 new putative proteins related to invasion, iron acquisition, and immunosuppresion were discovered.

With the sequence of *A. actinomycetemcomitans* essentially complete, it now is possible to develop experimental tests of these virulence mechanisms, develop other testable hypothesis of this bacteria's life style, and determine the mechanisms it uses to colonize the human oral cavity.

Chapter I

Introduction

1.1 Background

1.1.1 DNA and Genes

Deoxyribonucleic acid (DNA) is a double helix composed of two antiparallel strands of linked nucleotides. A nucleotide is composed of a nitrogenous base, a pentose sugar (2-deoxyribose), and a phosphate group. There are two types of nitrogenous bases, the pyrimidines, which consists of Thymine (T), and Cytosine (C), and the purines which are Adenine (A), and Guanine (G) (figure 1.1). The nitrogenous base is linked to position 1 on the pentose via glycosidic bond from N1 of pyrimidines or N9 of purines. By convention, the positions on the pentose ring are given a prime (') to avoid ambiguity. A nitrogenous base linked to a sugar is called nucleoside whereas a phosphorylated nucleoside is termed nucleotide. Adjacent nucleotides are linked to each other through a phosphodiester bond where the 5' position of one pentose ring is bound to the 3' position of the next pentose ring via phosphate group. Hence, the sugar-phosphate backbone consists of 5'-3' phosphodiester linkages where the nitrogenous bases protrude out from the backbone. The 5' terminal nucleotide typically has a phosphate at the 5' position while the 3' terminal nucleotide typically has a free 3' hydroxyl. It is convention to write the sequence of DNA in the 5' to 3' direction.

The DNA structure, first elucidated by Watson and Crick (**Watson et al, 1953**), involves two strands of DNA twisted around each other to form a double helix. The two strands bound via hydrogen bonds formed through specific base pairing of the nitrogenous bases where adenine pairs with thymine via two hydrogen bonds and guanine pairs with cytosine via three hydrogen bonds as seen in figure 1.1. Hence, the paired bases in DNA are said to be complementary.

The sugar-phosphate linkages constitute the backbone, which is on the outside of the double-stranded helix whereas the nitrogenous bases protrude towards each other to the inside in a perpendicular angle to the axis of the backbone. Each base pair is rotated by an angle of ~36 degrees around the axis relative the next base pair allowing for ~ 10 base pairs per a complete turn. The twisting of the strands around each other forms a narrow groove of 12 angstroms and a major groove of 22 angstroms. The double helix is right-handed as the turns run clockwise. These features correspond to the most prevalent form of DNA which is called B-form. The B-form provides an average that fits the DNA *in vivo* under physiological conditions. However, other forms of DNA do exist in the cell such as A-form and Z-form (Table 1.1) when a particular DNA sequence and /or physiological condition influences the final form of a region within the DNA molecule.

Helix type	Base pair/turn	Rotation/Base pair	Helix diameter (Angstrom)
A	11	+34.7	23
B	10	+34	19
Z	12	-30	19

Table 1.1. Different forms of DNA

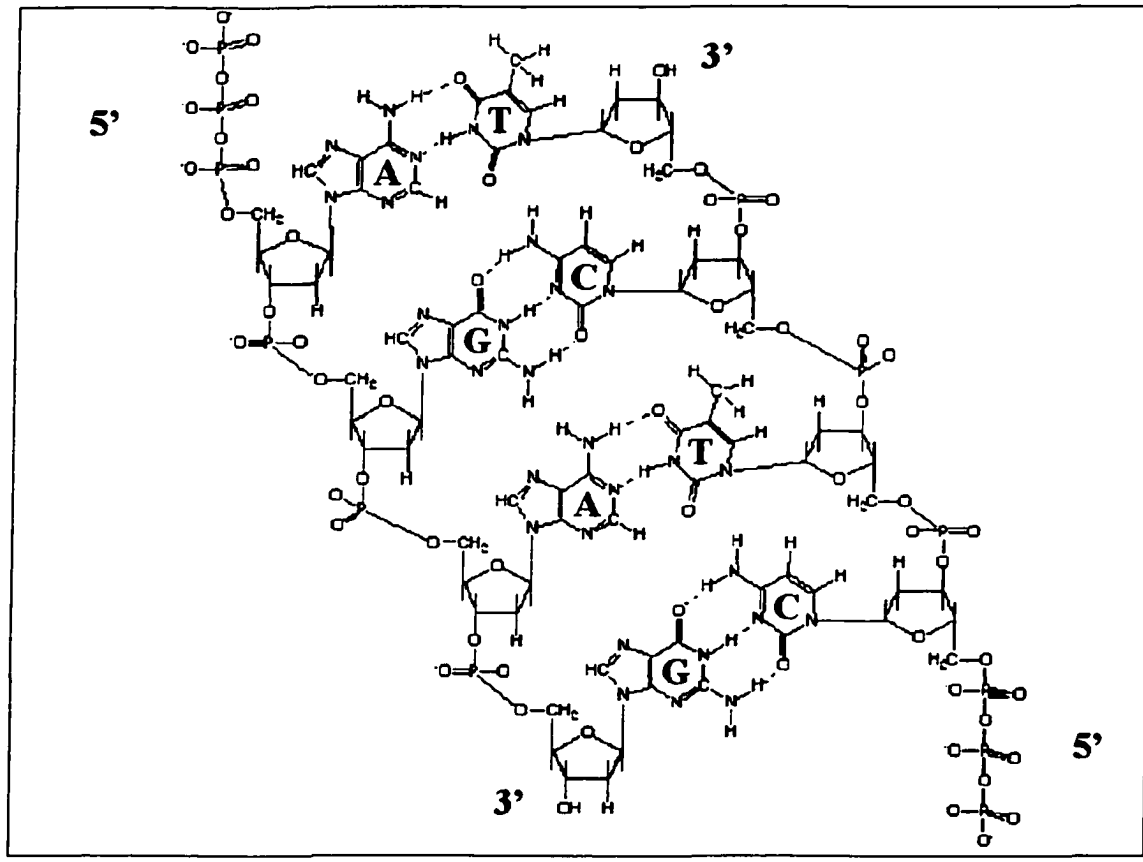


Figure 1.1. The structure of DNA.

Portions of the DNA sequence code for peptides through the genetic code. The genetic code is read in triplets which means that from 4 bases (A, C, G, and T) there are 64 different codes that can be used for coding amino acids for protein sequence.

The relationship between the DNA sequence, its primary transcript, and the synthesized protein is referred to as the central dogma of molecular biology (**Crick, 1970**) which subsequently was adjusted to include the transfer of information from RNA to DNA with the discovery of viral reverse transcriptases (**Baltimore, 1970**) as illustrated below.

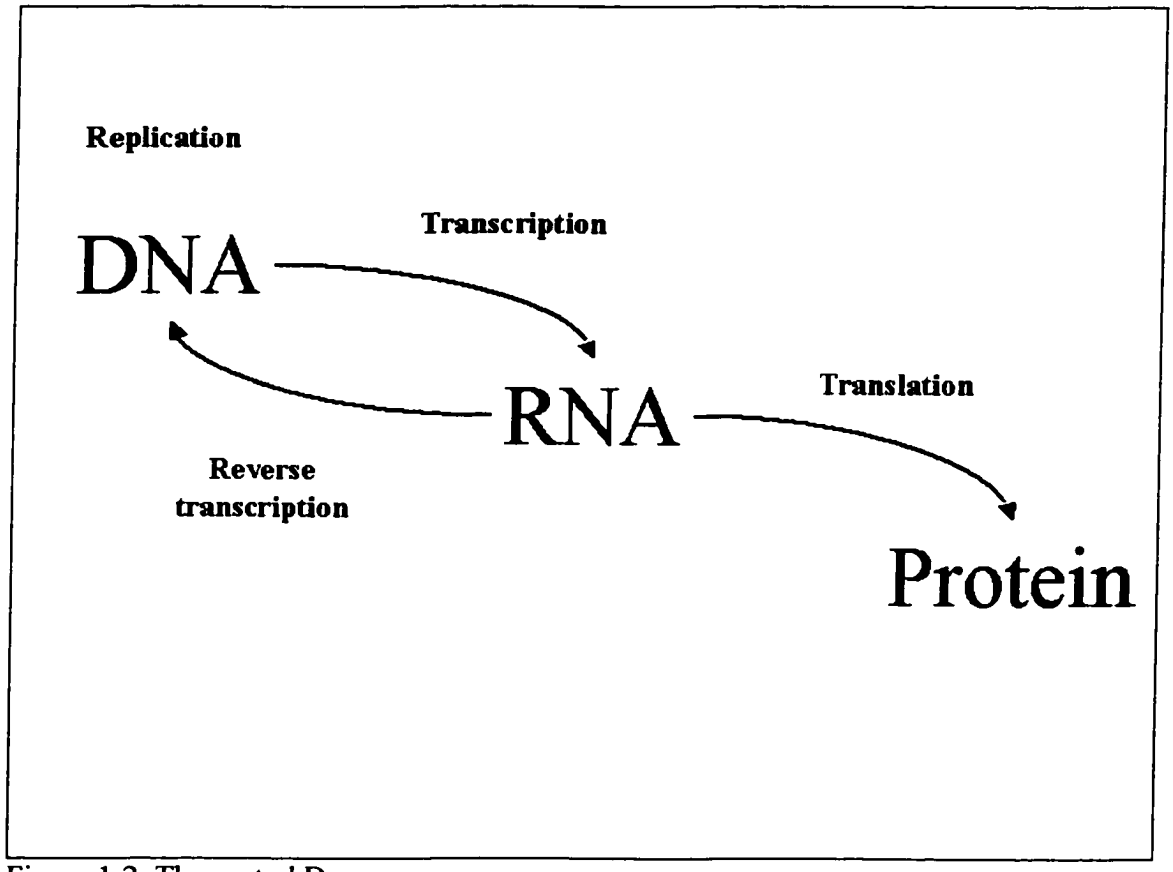


Figure 1.2. The central Dogma.

1.1.2 Organization of Prokaryotic Genomes

Genes can be classified by the type of functions their products perform. There are genes that code for enzymes, i.e. proteins which catalyze reactions are enzymes, those that code for structural proteins, and genes that encode for regulatory proteins. In addition, there are genes whose RNA transcript is not translated as they encode tRNAs and rRNAs. However, based on the total numbers, the majority of the bacterial genome encodes genes for proteins with enzymatic and structural functions. Often in bacteria, genes are organized into clusters of genes that code for proteins in the same pathway or part of the same structural apparatus such ribosomal proteins (i.e. an operon). However,

the organization of genes within a specific operon often is not conserved in distantly related organisms (**Koonin et al, 1997**). During the course of this dissertation research, a number of examples were found in *A. actinomycetemcomitans* that agree with this observation. For example, the biotin synthesis operon, has its genes in a conserved order in *A. actinomycetemcomitans* and *H. influenzae* while *E. coli* has the genes involved in this biosynthetic pathway split into two operons. In contrast, operons such as the one encoding tRNA modifying enzymes, are virtually identical in all three organisms.

1.1.3 *Actinobacillus actinomycetemcomitans* and Virulence

A pathogenic organism is an organism that is capable of causing disease (**Falkow 1996**). However, from a physiological viewpoint, pathogenicity is simply another aspect of the metabolic versatility of microorganisms which represent a specialization that allows for their long term survival (**Falkow 1996**). This versatility is demonstrated by their ability to evade the host immune system and penetrate its defenses, compete with the natural microflora residing within the host, target the tissue(s) of interest for colonization, scavenge for micronutrients from the hostile environment and replicate successfully. All these "alternate" metabolic aspects cause the resulting disease, which is necessary for their survival. In this dissertation research, the pathogenicity and modes of virulence of *A. actinomycetemcomitans* were explored.

Actinobacillus actinomycetemcomitans is a Gram-negative oral pathogen, coccobacillus, and facultative anaerobe from the family *Pasteurellaceae* (*Mannheimia*).

Although there still is some debate as to its genus, the medical community currently prefers the name actinobacillus even though the studies by De Ley and colleagues (**De Ley et al, 1990**) suggested the transfer of the organism to the genus *Haemophilus*.

Actinobacillus actinomycetemcomitans originally was isolated from patients with Localized Juvenile Periodontitis (LJP) and has been strongly implicated in a variety of periodontal diseases, including LJP, its most common form (**Sloth et al 1980**). Studies showed that there are differences in disease manifestation associated with particular genotypes of *Actinobacillus* (**DiRienzo et al, 1994**). It was shown that all progressive types of LJP are linked to a clonal species with five serotypes (a-e) (**Haubeck et al, 1995**). The most rapidly progressive type of LJP has been linked specifically to serotype b termed JP2. The enhanced leukotoxic activity found in JP2 relative to other serotypes of *Actinobacillus* is due to a 540 bp deletion in the promoter region of the leukotoxin gene operon (**Brogan et al, 1994**). In this dissertation research, the *A. actinomycetemcomitans* strain sequenced lacked this 540 bp sequence.

The mechanism of *A. actinomycetemcomitans* pathogenesis has been the focus of intensive research for the last 20 years. Many potential virulence factors have been discovered and characterized (**Zambon et al, 1985**) as shown in table 1.2, where a listing of previously studied virulence factors characterized from *A. actinomycetemcomitans* are listed. During the course of this work, a number of additional putative virulence factors were discovered.

Kiley and Holt (**Kiley et al, 1980**) originally characterized the lipopolysaccharide (LPS) in *A. actinomycetemcomitans*. More recently, studies have linked *A. actinomycetemcomitans* to the loss of alveolar bone resorption via osteoclasts and the inhibition of osteogenesis (**Loomer et al, 1995**). It also has been shown that *Actinobacillus* LPS induces the production of interleukin-1 antagonist preventing bone resorption (**Bartold et al, 1988**), affects the proteoglycan synthesis in the gingival connective tissues and binds hemoglobin suggesting a possible role of LPS in iron acquisition (**Grenier et al, 1997**).

Virulence Factor	Genes
Colonization and invasion	
Adhesin / invasion	Fimbrea (**Inoue et al, 1998**)
	Surface associated material (**Hara et al, 2000**)
	Extracelluale vesicles (**Meyer et al, 1994**)
Iron acquisition	Afu system (Hit homologue) (**Willemsen et al, 1997**)
	Hemin binding activity (**Graber et al, 1998**)
	Yfe homologue (**Graber et al, 1998**)
Evasion of host immune systems	
Leukotoxin	Leukotoxin (**Brogan et al, 1994**)
Immunosuppressive proteins	IgA protease (**Gronbaek 1999**)
Fc-binding protein	Omp29: Fc-binding protein (**Mintz et al, 1994**)
Host cells destruction	
Cytotoxin	Cytoleathal distending toxin (**Sugai et al, 1998**)
	Leukotoxin (**Brogan et al, 1994**)
Bone resorption	Surface associated material (**Hara et al, 2000**)
Inhibition of host repair	
Inhibition of fibroblast proliferation	LPS* (**Bartold et al, 1988**)
	Cytoleathal distending toxin (**Sugai et al, 1998**)
Inhibition of bone formation	Surface associated material (**Hara et al, 2000**)

Table 1.2. Known virulence factors in *A. actinomycetemcomitans*.

The ability of *A. actinomycetemcomitans* to invade cells is a phenotype that was shown almost a decade ago when Meyer and colleagues were able to recover *A. actinomycetemcomitans* from the KB human oral cell line (**Meyer et al, 1991**).

However, although the detailed mechanism of its invasion still is not known, it has been observed that *A. actinomycetemcomitans* exhibits three aspects of invasion. First, it induces the formation of *S. typhimurium*-like aperture in the host cell membrane. Then, it escapes from its encapsulating vacuole, and follows a microtubule-associated movement inside the host. Finally it exits the cell by the passage to adjacent host cells. (**Meyer et al, 1996, and Meyer et al, 1999**). Interestingly, it appears that *A. actinomycetemcomitans* has a much faster reproductive doubling time inside the cytoplasm of the host than typically is observed *in vitro*. Recently, it has been shown that *A. actinomycetemcomitans* might employ the platelet-activating factor receptor to gain entry into the epithelial cells via phosphorylcholine-decorated lipopolysaccharides present on bacterial membranes (**Schenkein et al, 2000**). Here, the phosphorylcholine usually exists within structural molecules integrated into the LPS of Gram-negative bacteria such *H. influenzae*, and the lipoteichoic acid of Gram-positive bacteria such as *Streptococcus pneumoniae* (**Tomasz et al, 1967**). A genetic locus required for phosphorylcholine synthesis has been found in *S. pneumoniae* (**Zhang et al, 1999**) which has homology to a locus in *H. influenzae* and as discovered during this dissertation research also is present in *A. actinomycetemcomitans*. Mutations in these genes have been shown to reduce virulence in *S. pneumoniae* (**Zhang et al, 1999**).

It appears that *A. actinomycetemcomitans* has the ability to acquire iron from the host as studies suggest that *A. actinomycetemcomitans* possess three siderophore-independent iron uptake systems, the *Afu*ABC system homologous to the *Haemophilus influenzae* and *Neisseria meningitidis*systems (**Willemsen et al, 1997**) , the Yfe-homolog

system similar to the *Yersinia pestis* system, and a hemin binding protein (**Graber et al, 1998**) . *A. actinomycetemcomitans* also has the capability to degrade lactoferrin, a glycoprotein a component of the innate immune system (**Allugupali et al, 1996**).

Leukotoxic activity in *A. actinomycetemcomitans* has been well known since the early 1980's (**Taichman et al, 1980**). Leukotoxin A is part of a family of toxins called RTX (repeats-in-toxin) that have a similar repeated amino acid sequence (LXGGXGNDX) which is linked to its leukotoxic activity (**Felmlee et al, 1988, and Welch et al 1991**). Studies showed that deletions in these repeats result in the production of an inactive toxin (**Felmlee et al, 1988**). Toxins of this family are encoded by bacteria such as *Escherichia coli* (*hlyA*) and *Pasturella. haemolytica* (*lktA*) (**Welch et al, 1991**) in an operon that contains three other genes, *lkxC, lkxB,* and *lkxD* . It has been suggested that LkxC protein is involved in post-translational modification of leukotoxin A to produce an active toxin. In *E. coli*, the activation of HlyA requires acyl carrier protein (**Hardie et al, 1991, and Issartel et al, 1991**). LkxB and LkxD are responsible for localizing the active LkxA on the outer membrane. Leukotoxin A affects polymorphonuclear leukocytes and macrophages by forming a cation-selective pore on the cell surface causing leakage of potassium ions and cell lysis (**Lally et al, 1989**). The overall organization of this operon is identical in all RTX toxin-producing bacteria studied.

Another toxin that recently has been discovered in *Actinobacillus* is the cytolethal distending toxin (Cdt) which is homologous to the *E. coli* enterotoxin (**Sugai et al, 1998**).

This toxin is encoded by an operon where the organization of the genes is conserved in other Gram-negative bacteria (cdtA – cdtB- cdtC). The presence of all three genes is required for the cytotoxicity effect. *In vitro* studies showed that this toxin causes G2 arrest in the cell cycle of HeLa cells (**Comayras et al, 1997**). It also has been shown that in *E. coli*, Cdt causes the failure to inactivate Cdc2, which ultimately causes cell cycle arrest (**Comayras et al, 1997**), suggesting that the role of Cdt in *Actinobacillus* pathogenesis might be growth arrest of fibroblasts that ultimately results in the loss of collagen. This toxin also can induce apoptosis in murine macrophage cell lines, suggesting that it might suppress the local immune response by acting in concert with leukotoxin and LPS to inflict damage in the surrounding connective tissues in the gingiva (**Kato et al, 1995**).

Adhesion of the bacteria to host cells is the most prominent phenotype from fresh clinical isolates of *A. actinomycetemcomitans* (**Mintz et al, 1994**). This phenotype is due to many factors, among them are the RcpA/B proteins (see above discussion). Inoue and colleagues have characterized a low-molecular-weight protein, Flp that is speculated to be a major component of *Actinobacillus* fimbreae (**Inoue et al, 1998**). Recently, a cluster of seven genes down stream of *rcp*A/B genes and close to the *flp* locus has been implicated in non-specific adhesion. These genes were termed *tad*A-G (Tight Adhesion). Transposon mutation in any of the *tad* genes diminishes the rough phenotype of *Actinobacillus*. This gene is homologous to *pti*H from *Bordetella pertussis* that is required for the secretion of pertussis toxin. The other six genes, *tad*B-G have no

assigned functions, yet they are homologous to uncharacterized genes in *Y. pestis*, *H. ducreyi*, and *P. mulocida*. **(Kachlany et al, 2000)**

Other potential virulence factors of *A. actinomycetemcomitans* have been reported. These include GroEL-like protein (**Paju et al, 2000, and Goulhen et al, 1998**), OmpA family antigen (**White et al, 1998**), and trypsin-like protease (**Wang et al, 1999**), although their role in pathogenesis of *A. actinomycetemcomitans* is still to be determined.

One of the goals of this project is to gain a deeper understanding of the virulence mechanisms of this organism. There are many different approaches to evaluate candidate virulence factor (**Weiss et al, 1986**). Two approaches documented by Falkow (**Falkow 1991**) were to determine if Koch's postulate was fulfilled by administering the purified potential virulence factor in a model animal and observing whether analogous effects similar to the clinical disease are produced. Alternatively, a mutant strain lacking the virulence factor could be characterized to determine the contribution of a potential virulence factor to the development of a certain disease. For these approaches to be more efficient and accurate, the sequence of the *A. actinomycetemcomitans* genome would be extremely useful as having the genomic sequence available will allow locating its encoded genes including those involved in pathogenicity. The sequence also will enable observation of the overall genomic organization of this microbe, and then allows for comparing its genome to that of other microorganisms.

Approximately 99.9% of the roughly 2.1 Mb genome has been sequenced as part

of this dissertation research. With this information, the encoded genes and other genomic features can be determined, the metabolic functions can be reconstructed, and aspects of its pathogenicity can be elucidated.

1.2 Sequencing Strategy

1.2.1 Brief History of Sequencing

The initial DNA sequencing methods were reported independently in 1977 by Sanger and colleagues (**Sanger et al, 1977**) in England, and by Maxam and Gilbert at Harvard University (**Maxam et al, 1977**). The Maxam and Gilbert sequencing method involves chemical degradations of the DNA molecule of interest at specific bases. This degradation is accomplished by four separate chemical reagents that cleave the 5'-radiolabeled DNA preferentially at A, at G, at C, and at C and T equally. The four sets of reaction products then are electrophoresed on a polyaccrylamide gel side-by-side to resolve the resulting fragments (**Maxam et al, 1977**), which then are detected by autoradiography.

The Sanger method, which is more popular today, involves amplification of template DNA in the presence of trace amounts of the four dideoxynucleotides (**Sanger et al, 1977**). Four separate sequencing reactions corresponding to each of the four dideoxynucleotides initially were employed, but today with the introduction of four different fluorescent labeled dideoxynucleotides, all four reactions can be combined into one. Typically today, each reaction contains the DNA of interest, the four deoxynucleotides, the four fluorescently labeled dideoxynucleotide, and a modified

Thermus aquaticus DNA polymerase. After incubation, the resulting randomly terminated nested fragments are separated based on the migration through a polyacrylamide slab gel, or more recently on a capillary gel. Over the years, a number of improvements and modifications were introduced to the Sanger method that have resulted in increasing both sequencing efficiency and accuracy. For example, the distribution of these nested fragment sets has been optimized by varying the ratio of deoxynucleotides and dideoxy terminators (**Tabor et al, 1987**) and thermostable DNA polymerases have been introduced from different thermophilic bacteria such the *Bacillus stearothermophilus* Bst DNA polymerase (**Stenesh et al, 1972, Mardis, et al, 1989, and Mead, et al, 1991**) and the *Thermus aquaticus* Taq DNA polymerase (**Innis et al, 1988**). Thermophilic DNA polymerase allows the utilization of a higher extension temperature during the thermocycle reaction which, in turn, increases the accuracy and reduces background sequences and ambiguities. Also, since errors in a determined DNA sequence can be introduced by product and template secondary structure, they have been minimized by utilizing polyacrylamide gel electrophoresis in the presence of urea to denature hairpin loops in the products, and the use of dimethylsulfoxide (DMSO) in the sequencing reaction to minimize the formation of secondary structure in the templates. In addition, the utilization of nucleotide analogues such as inosine instead of guanosine to limit the hydrogen bonds between base pairs also reduces the formation of secondary structures.

A major advance in improving the Sanger sequencing method came with the use of fluorescence dyes instead of radioactive labeling, which resulted in the development of

automated detection systems (**Smith et al, 1986, Smith et al, 1987**). Here, each dye has a slightly different absorption maxima. Exciting the dye with a common source such as an argon ion laser causes four resolvable spectra that can be deciphered. The original four dyes used for sequencing were fluorescein derivatives (**Connel et al, 1987, and Lee et al, 1992**). Later, rhodamine derivatives were utilized (Figure 1.3) and more recently energy-transfer dyes (ABI BigDye and Amersham ET-terminators), have been utilized (**Rosenblum et al, 1997**) (Figure 1.4). These energy-transfer dyes consist of two moieties, a fluorescein-derivative dye that serves as a donor dye which initially is excited by the laser source, and a d-rhodamine dye which accepts the emission of the donor dye as excitation energy, subsequently emitting light at specific wavelength for detection by a CCD camera or photodiode array.

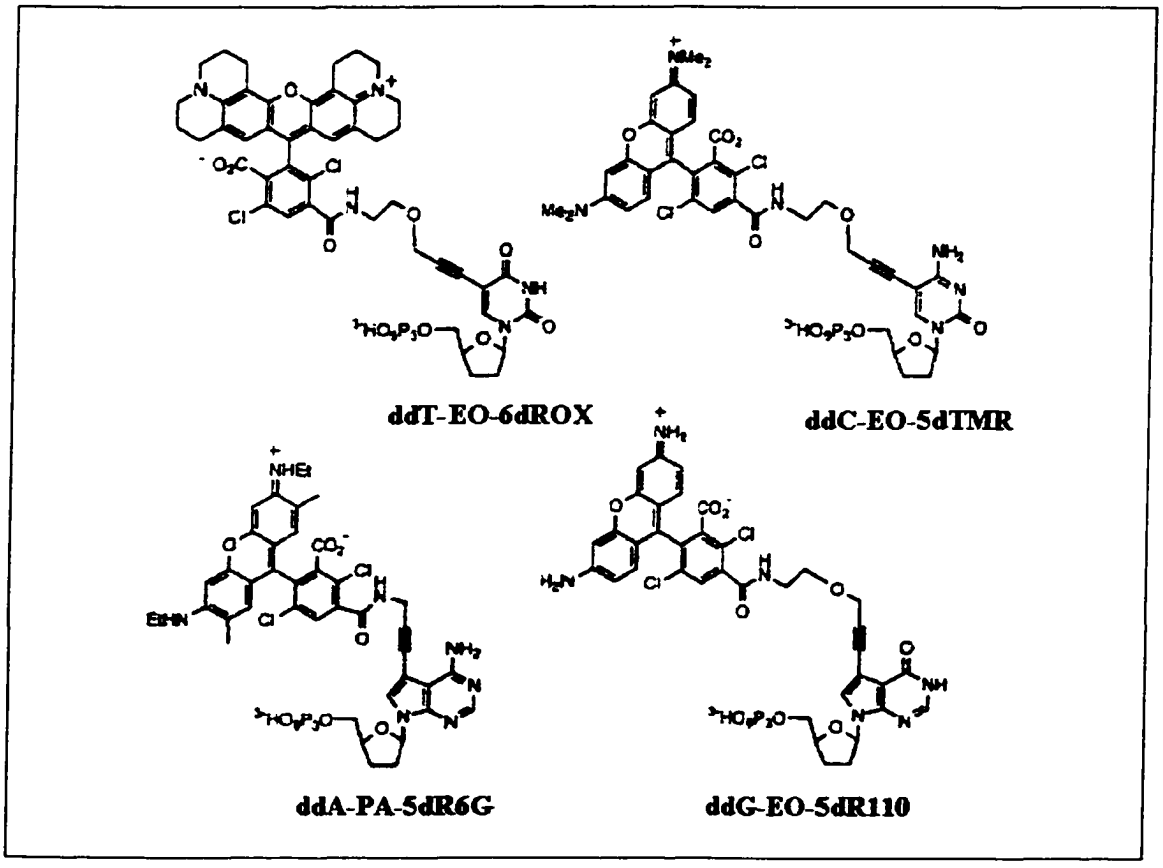


Figure 1.3. 5,7-dichloro rhodamine (d-rhodamine) terminators used for sequencing.

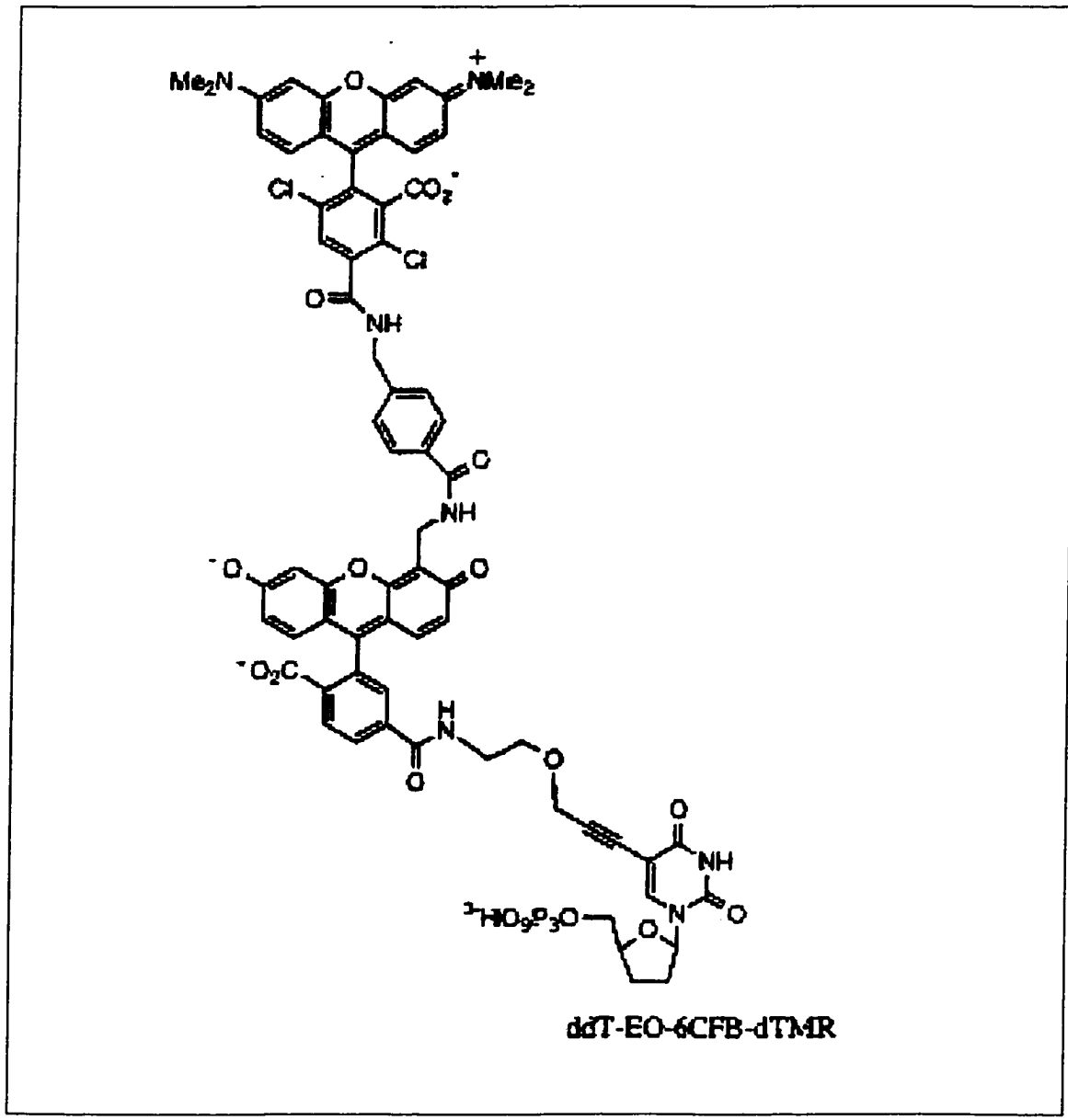


Figure 1.4. BigDye terminator (attached to thiamine).

Two variants of the Sanger sequencing method have been used, one with 5'-labeled fluorescent primers, and more recently an approach where the dideoxy terminators are fluorescent-labeled **(Rosenblum et al, 1997)**. The major advantage of the dye-labeled

terminators is that a single reaction can be used to sequence each DNA template sample, as opposed to the dye-labeled primers, where four primers, each with the same oligonucleotide sequence and different fluorescent dyes, are used. In the later case, four separate reactions, each containing one of the four dideoxy terminators are needed, each with a different labeled primer. At the end of the cycling reaction, the four samples are pooled and loaded on one lane for electrophoresis. Since each dye must be associated with the dideoxy terminator used in separate reaction, this disadvantage can be overcome with terminator approach, as there is no need to label primers prior to sequencing, which is a time-consuming process, limiting the choice of priming sites for sequencing (**Rosenblum et al, 1997**).

1.2.2 Shotgun Phase of DNA sequencing

The shotgun sequencing strategy employs an initial shotgun phase, followed by a closure and finishing phase. The shotgun phase involves the random shearing DNA to smaller clonable fragments (1-6 kb). Different strategies have been developed to generate such fragments as randomly as possible including sonication (**Deininger, 1983**), mechanical shearing (**Schriefer et al, 1990**), partial restriction digestion (**Fitzgerald et al, 1992, Bankier et al, 1987**), and nebulization (**Bodenteich et al, 1993**) (Figure 1.5). We have implemented the nebulization procedure using low temperatures (~ -20^{0}C) at 6-8 psi to achieve extremely random shearing and that produces fragments in the 2-4 Kb range.

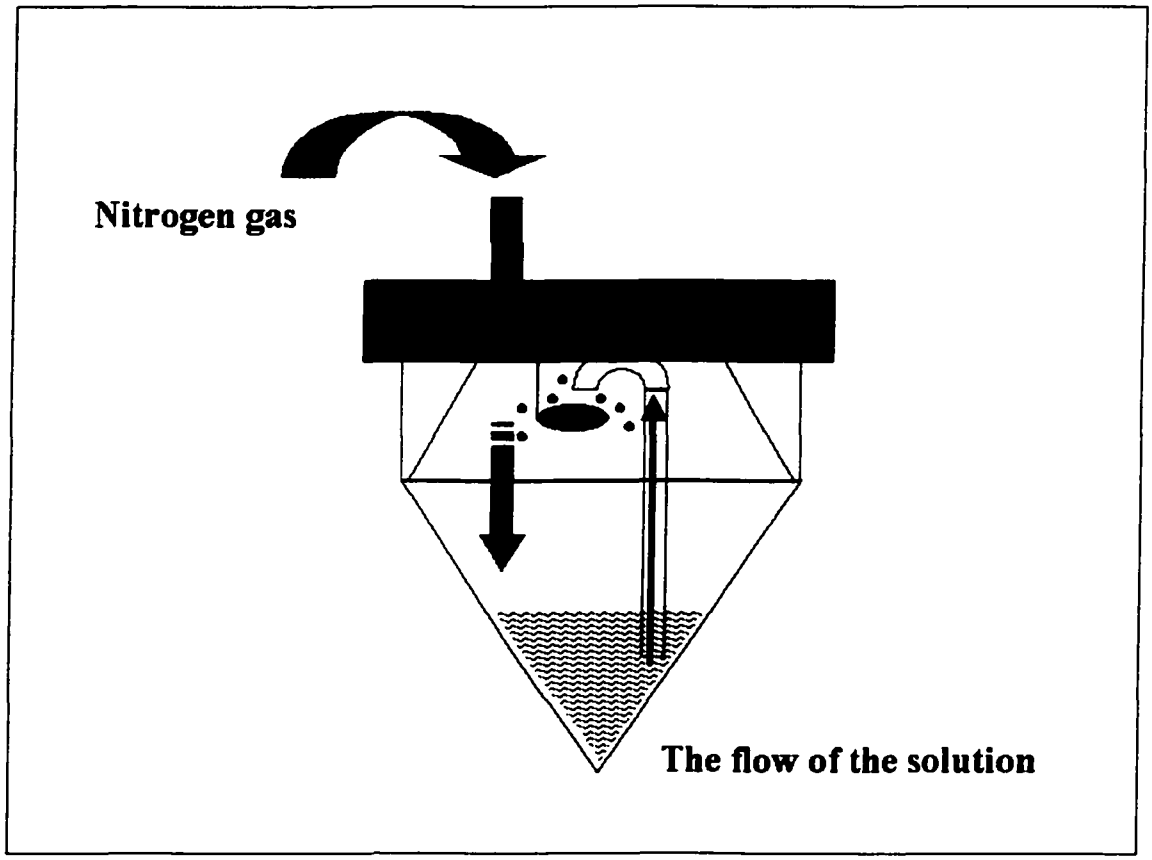


Figure 1.5. The nebulizer. The green arrow represent the flow of the nitrogen gas into the chamber. The narrow arrow represents the flow of the DNA solution.

These fragments then are subcloned into a cloning vector such as pUC18 or M13 and the resulting plasmids are transformed into *E. coli*. Following isolation of the plasmid DNA, sequencing is accomplished using standard universal primers that prime at vector sites flanking the subcloned fragments. Typically, 6-10 equivalents of the genome (6-10 x coverage) are sequenced before beginning the final phase of gap closure. The advantage of the random shotgun sequencing method is its suitability for high throughput sequencing and speed, since the primers used are standard throughout the entire project.

However, regions in the genome typically remain undetermined during the shotgun phase because of either cloning bias and low sequencing efficiency. Other unsequenced regions usually are due to sequence assembly difficulties caused by the presence of repetitive elements in the genome which are larger than the average sequencing read. These repeats cause a branching point in the assembly and ultimately must be resolved using a different sequencing approach such as those discussed below.

1.2.3 Sequence Assembly – Phred, Phrap, and Consed

One of the first DNA sequence assembly programs, written by Roger Staden (**Dear et al, 1991**), was designed for manual data entry and editing of the trace files generated by a user reading directly from electrophoretograms of radio labeled sequencing gels. As the data collection process became more automated, the sequence assembly software also became more robust. The most common programs currently used in high throughput genomic sequencing laboratories for quality assessment, assembly, and editing are Phred, Phrap, and Consed (**Gordon et al, 1998**). Phred (**Ewing et al, 1998**) reads DNA trace files, calls bases, and assigns quality values to them after the vector sequences are masked out by the crossmatch program. The base calls and their quality values then are written to output files. This software can read trace files from different commercial DNA sequencing instruments providing flexibility of utilizing different chemistries in a sequencing project. Quality values for the bases are written to files, which can be used by the Phrap assembly program to increase the accuracy of the assembled sequence (**Gordon et al, 1998**). Phrap ("phragment" assembly program) then

employs the data generated by Phred and assembles the individual sequence reads into contiguous sequences using the Phred quality values. Then, Consed (**Gordon et al, 1998**), a viewer program, can display phrap output interactively, providing extensive information about the assembled sequence including the original trace files and the quality of the data and the presence of repetitive sequences. Viewing the data aids in designing the experiments needed to close and finish different regions. Then, a primer calculation program, such as PimOU, can be used to generate an output file containing suggested primers for extending the ends of each contiguous sequence using user-defined parameters that includes melting temperature (Tm), base composition, and base sequence constraints.

1.2.4 Closure Strategies

As discussed above, gaps in the DNA sequence remain after the shotgun phase for various reasons. There are physical gaps that have no sub-clone representation. There are gaps caused by difficult to sequence regions which have been sub-cloned, and there are gaps caused by the inability of the sequence assembly programs to correctly assemble large repeated sequences. Usually physical gaps can be closed by generating polymerase chain reaction (PCR) products from genomic DNA to cover the physical gaps and sequencing them, while the existing pUC sub-clones usually can be used as templates for gaps resulting from difficult to sequence regions using a variety of techniques as discussed below.

1.2.4.1 Repetitive sequence and the use of large-insert clones

Repetitive sequences can vary in complexity from the very simple such as homopolymeric tracts of single nucleotides (poly(A), or poly(C)) to the more complex composed of several multimeric repeats that are homogeneous, heterogeneous or degenerate repeats (**Van Belkum et al, 1998**).

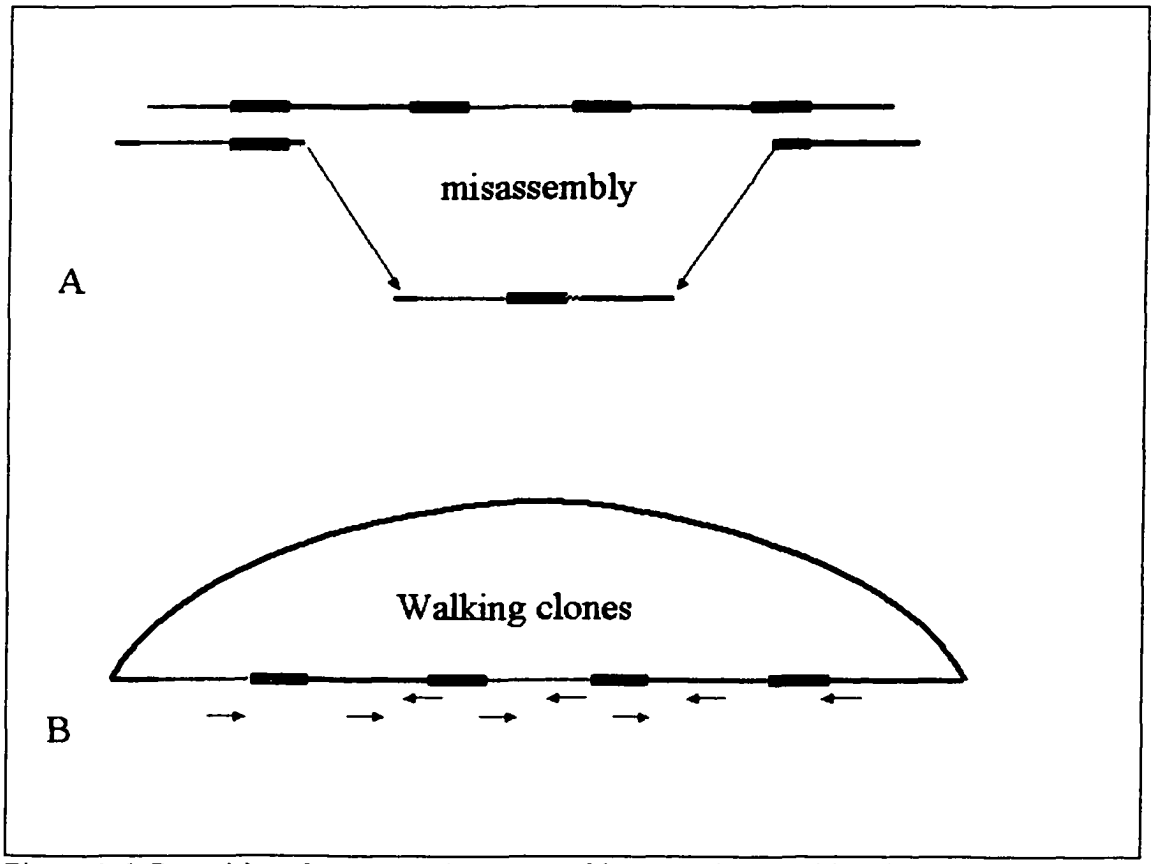


Figure 1.6. Repetitive elements and misassembly. A. Misassembly of a sequence with four repeat sequences (boxes) intervened with unique sequences. Misassembly occurs when the assembly program mistakes the repeated sequences as an overlap and assemble them. B. The role of walking clone is to span such repeats.

To resolve the sequence of these regions, it often is useful to obtain large-insert (4-6 Kb) "walking" clones (Figure 1.6) where the repeat region is completely contained within a single clone which then can be sequenced by multiple rounds of custom-synthesized primer walking.

1.2.4.2 Incompatibility and rearrangement – PCR and Multiplex Polymerase Chain Reaction (MPCR)-based approaches

In large DNA sequencing projects, it sometimes is possible to observe regions of the genome that may be toxic after they are cloned into a vector and then transformed into another organism. In addition, there may be instances where long repeats in walking clones might undergo rearrangements or deletions and hence not accurately represent the original sequence (**Behnke et al, 1979**). Thus, a PCR-based approach to sequencing physical gaps (**Claustres et al, 1989**), which does not require a cloning step, can reduce the possibility of toxicity and rearrangements associated with cloning (**Tettelin et al, 1999**). Although PCR is powerful when the order of unjoined sequences is known, when the order is not available due to, for example, the lack of a physical map, it is more practical to use a more robust version of PCR. One such method is the multiplex PCR (MPCR) initially described by Claustres and colleagues for diagnosing patients with Duchenne muscular dystrophy by generating MPCR products (**Claustres et al, 1989**) for regions where mutations occur (**Hejtmancik et al, 1986**). By pooling multiple primer sets that flank all potentially mutated regions in a simple amplification reaction, products can be generated which then can be sequenced with each of the individual primers

(**Clauster et al, 1989**). During the course of this Ph.D. dissertation research, this multiplex PCR was investigated and was adapted for high-throughput sequence gap closure to join and finish unordered contigs (contiguous sequences). Here, primers were synthesized from the ends of all assembled contigs, pooled, and used in MPCR. Products of the MPCR then were sequenced using the individual primers that were used in the original MPCR reaction.

To increase the efficiency of this technique, contiguous sequences can be hypothetically aligned by searching for "split" genes or operons, or by alignment with the contiguous sequences with completed genomes of other bacteria when available. This is achieved by subjecting the ends of the contiguous sequences to a homology search using BlastX program (**Altschul et al, 1990**) and then locating the contig ends which share homology to a GenBank entry. The ends of these contigs, then, likely are from adjacent regions separated by the gap. This process has been automated by an in-house Perl script, written by Jim White, that extracts the sequences from the ends of the contigs based on user-defined parameters, submits them to a BlastX search against the GenBank database (Figure1.7), and produces a listing of possible "split" genes.

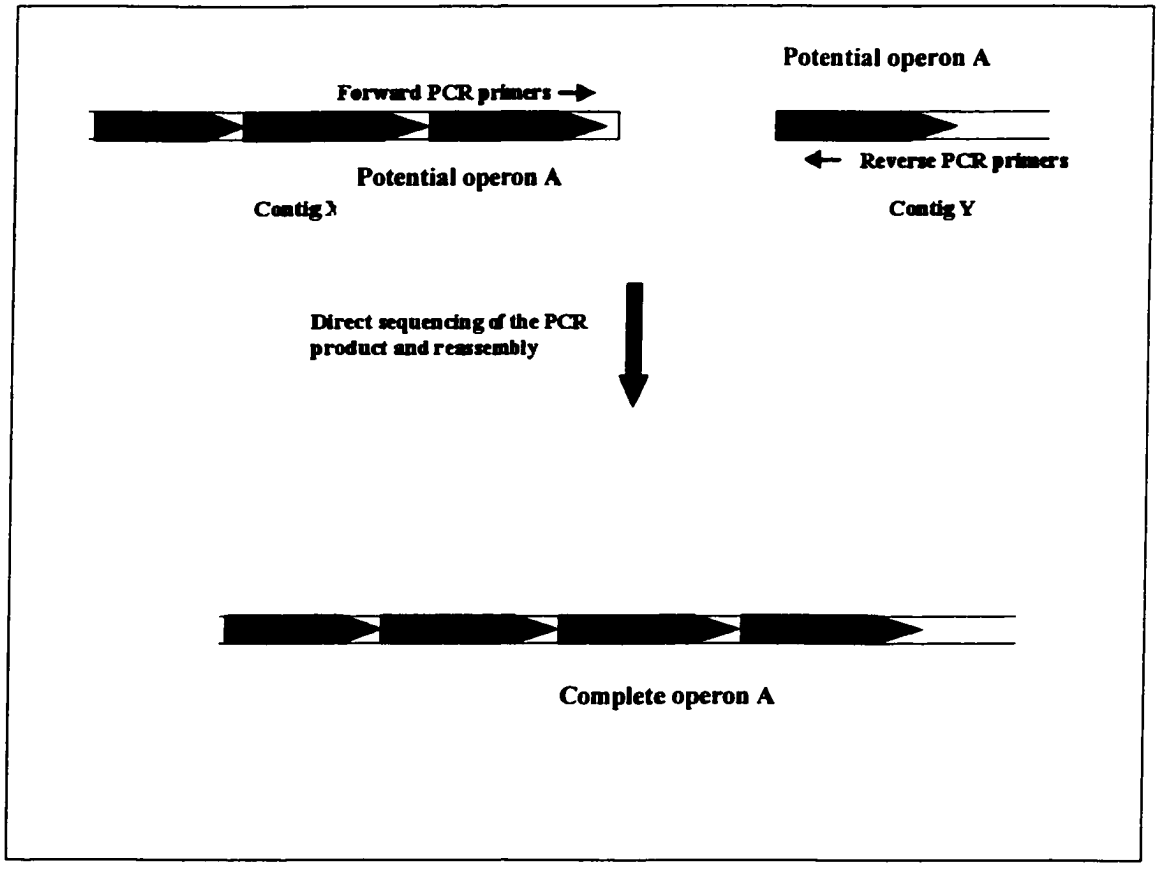


Figure 1.7. Recognizing the theoretical order of contigs based on knowledge in gene organizations in operons or the existence of split genes between contigs for PCRing the gaps.

1.2.4.3 GC-rich regions and nucleotide analogs

GC-rich sequences often are prone to form secondary structures such as hair-pin loops which are difficult to sequence and can cause "band compressions" that are observed as DNA sequencing artifacts (**Mills et al, 1979**). Different approaches have been utilized, such as the addition of DMSO or betaine to inhibit such structures from forming during the thermal cycling and the PCR reactions. Other approaches focused on

using nucleotide analogs, including 7-deaza-dGTP, that can only form two hydrogen bonds with cytosine due to a twist of the base from the normal base plain thereby reducing the possibility of G-C base pairing under the denaturing conditions of sequencing (**Dierick et al, 1993**).

Sequencing DNA templates with secondary structures often times is difficult to accomplish. To create a more accessible DNA sequencing template, it is possible to perform PCR on such a region using the nucleotide analogue 7-deaza-dGTP (figure 1.8) instead of dGTP in the PCR mix to generate PCR products that will ultimately yield longer, more accurate sequence reads with one of the commercially available DNA sequencing reagent mixes (i.e. ABI BigDye, ABI BigDye with dGTP replacing dITP, ABI dRhodamine mix, or the Amersham ET dye terminator mix).

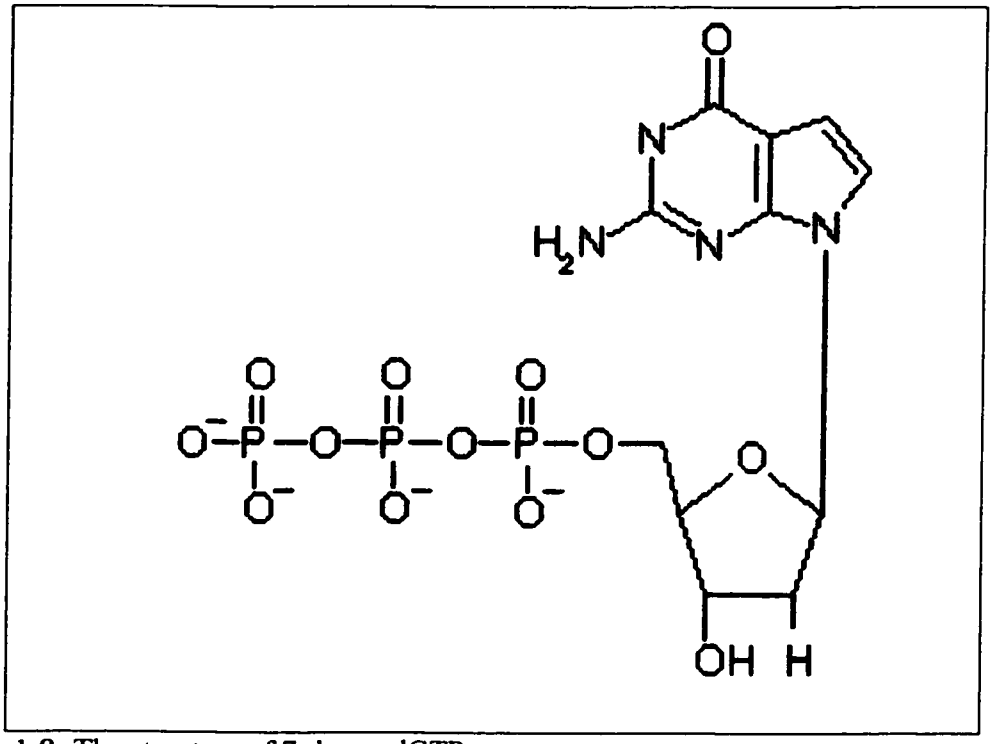


Figure 1.8. The structure of 7-deaza-dGTP.

1.2.5 Analysis and Annotation of a DNA sequence

Analysis and annotation provides information about the various biological functions encoded by a DNA sequence. In the past, this process usually did not occur until a DNA sequence was complete. However, analysis and annotation of the *A. actinomycetemcomitans* genome began once 6X coverage of the genomic sequence was obtained because it was thought that preliminary genomic information might be useful for the closure progress as described above. In addition, since 6X coverage yields over 95% of the sequence (**Lander et al, 1988**), this preliminary annotation might provide early insight into *A. actinomycetemcomitans* virulence factors, unique metabolic pathway characteristics, genomic organization, and the sequence of other regions of biological interest. Therefore, the *A. actinomycetemcomitans* initially was searched for open reading frames (ORFs). Once biological functions are assigned to each ORF, a preliminary metabolic pathway scheme was constructed. This information then was analyzed to find unique gene sets, such as virulence factors, that possibly could give *A. actinomycetemcomitans* its niche in the environment. In addition, putative regulatory elements, regulatory proteins, promoters and repetitive sequences also could be located and assigned putative functions.

To analyze a genome, a battery of programs are needed to work collectively in steps to extract as much information as possible from the sequence data (see figure1.9). The first step is to identify potential ORFs in the genome which then are assigned

putative functions based on the results from homology search against the GenBank database. Functional protein coding genes then are grouped into physiological and metabolism classes. ORFs with no homology in the database are analyzed further to determine if they harbor known functional or structural motifs. For example, a predicted protein sequence can be analyzed to discern if it contains membrane-associated, or trans-membrane associated sequences, or other features. Statistical information, such as, the number of genes associated with each specific metabolic function, then can be compiled.

Data visualization is a rapidly developing area in bioinformatics since it allows users to view the data in a useful comparative way. A large number of visualization tools have been developed which can be used to view the results from various computational analysis methods, such as Blast homology searches. These visualization tools can handle different data formats and provide a simple visual interface of large amounts of data that is internet accessible. It also is possible to link data sets from different sources graphically and make the data available without the need for powerful computers or programming expertise. To this end, a number of tools have been developed to display DNA sequence analysis results. Although these tools differ slightly in their user interfaces, they essentially assign ORFs with their putative functions and possible regulatory information. One of the most popular tools is the Kyoto Encyclopedia of Genes and Genomes (KEGG) (**Kanehisa et al,, 2000**) which has been developed and maintained by the Institute for Chemical Research Kyoto University. This well structured web-based visualization and annotation tool contains three main sections: a metabolic pathways database, genomic information, and computational tools such as homology searches and metabolic reconstruction. One of the unique features of KEGG

is that it contains a computational tool for building alternative routes between metabolic products. This tool allows searching for alternative enzymes that may be needed to complete a pathway when no BLAST homologue was detected. Here, KEGG provides views of the metabolic pathways as a series of connections of reactions that lead to final products. Hence it can arrange the reactions in different routes that can lead to the same product as long as the reactions it utilizes for the construction of these routes are biochemically valid, although filling in missing enzymes is strictly hypothetical until it is experimentally validated. Use of KEGG was very helpful in generating an overall visual system to annotate predicted metabolic pathways in *A. actinomycetemcomitans*. Another powerful analysis and visualization tool is the What Is There? (WIT) program created by Ross Overbeek and colleagues (**Selkov et al, 1997**). This web-based metabolic reconstruction viewer, also is a useful source of information about metabolic pathways that previously have been described for different organizations (**Selkov et al, 1997**). WIT has the capacity to allow editing of the functional assignments of any ORF in any organism in the database under a user-specified user-id. Additionally, the user can view other assignments made by peers, thereby allowing users to update annotation based on their scientific experiences. WIT also provides comparative diagrams of regions of the different genomes to facilitate comparative genomics. By utilizing other viewing programs such as Jalview (http://www.ebi.ac.uk/~michele/jalview/contents.html) and Boxshade (http://huge.eng.uiowa.edu/~tscheetz/sequence-analysis/examples/BoxShade/BOX_form.html), WIT offers a variety of options for visualizations. Although WIT and KEGG, both are valuable tools for metabolic reconstructions, the Reily metabolic schema (**Neidhardt et al, 1999**), offers a wider

spectrum of protein functions, catalytic and regulatory, with fairly detailed description. For example although KEGG utilizes a more straight forward approach to organization of enzymes, it omits many of the regulatory proteins. In contrast, WIT has a broader organization of metabolic pathways and includes regulatory protein that are derived from the Reily schema.

EcoCyc is another very powerful analysis-visualization tool (**Karp et al, 1997, http://ecocyc.PangeaSystems.com/ecocyc/ecocyc.html**) which is tailored to the *E. coli* genome. EcoCyc utilizes an interface that allows users to view the overall metabolic and physiological functions of the organism and selectively zoom in on the function of interest. HinCyc (**Karp et al, 1996, http://citeseer.nj.nec.com/374789.html**), is a similar visualization tool specifically tailored for *H. influenzae*. Finally, TIGR (**The Institute for Genome Research; www.tigr.org**) provides a user-friendly web site that contains analysis information on the organisms that they have sequenced.

In addition to the above, an alternative general visualization program, Artemis (**Rutherford et al, 2000**) which allows Blast and other analysis programs to be linked to the sequence data, contains both viewing and manual editing functions.

To search for secretory proteins, the amino acid sequences of all ORFs are generated and then analyzed through SignalP (**Nielsen et al, 1997**). This program, which is at http://www.cbs.dtu.dk/services/SignalP-2.0/#submission, predicts whether a protein contains a signal peptide using artificial neural networks trained on the identification of signal peptides from Gram-negative, Gram–positive, and eukaryotic organisms (**Nielsen et al, 1997**). This type of analysis is very useful for discovering possible virulence

factors, key elements for microbe-host interactions which also could be possible therapeutic vaccine targets. The table below shows the training sets for the SignalP neural network.

	Gram-negative	Gram-positive
Average length	25.1 aa	32.0 aa
N-regions	K+R-rich	
H-region	Short, hydrophobic	Long, hydrophobic
C-region	short, S+A-rich	longer, P+T-rich
-3,-1 positions	almost exclusively Ala	
+1 to +5 region	rich in A, D/E, S/T	

Table 1.3. General characteristics of prokaryotic signal peptides as utilized by SignalP algorithm.

As mentioned above, motif searches can be utilized to study putative ORFs for possible functional domains by comparison with a number of publicly available motif databases. Since a protein sequence may be too distantly related to a known protein to be revealed by a regular BLAST search, motif-based searches may reveal a cluster of amino acid residues that are important for certain functions such as binding properties or catalytic activities. To this end, a web-based search tool such as Motif at the Bioinformatics Center Institute for Chemical Research at Kyoto University (**http://motif.genome.ad.jp/**) can be useful to search protein sequences for conserved patterns. Motifs utilizes several protein pattern libraries such as Pfam (**Bateman et al, 1999**), ProDom (**Sonnhammer et al, 1992**), Blocks (**Henikoff et al, 1999**), and Prosite (**Hofmann et al, 1999**) all of which have different algorithms for collecting protein patterns and profiles.

Another powerful method for assigning functions to proteins, particularly if the protein has no known function, is the Cluster of Orthologus Groups (COGs) (**Koonin et al,, 1998**). Each COG contains several orthologous proteins from at least three different lineages. Since orthologous proteins typically perform the same function and are related by vertical descent or speciation, it is assumed that orthologous proteins have a common ancestor. The construction of COGs is based on the assumption that any group of at least 3 proteins from distantly related organisms that are more similar to one another than to any other protein from the same genome are likely to belong to an orthologous group. This allows for building groups of proteins with similar functions (**Koonin et al,, 1998**). In the course of this dissertation research, all ORFs with no significant BLAST homology to a protein of known function were analyzed through the COGnitor program (**http://www.ncbi.nlm.nih.gov/COG/xognitor.html**).

Finally, the tRNAscan-SE program was used to detect tRNA genes (**Eddy et al, 1994, and Fichant et al, 1991**). This program predicts tRNA genes based on both conserved nucleotides at specific positions for H-bonding, i.e. secondary structure for the tRNA stem regions, and the size range for the four loop regions. The output of tRNAscan-SE was integrated into the Artemis file where an overall representation of the genome can be viewed.

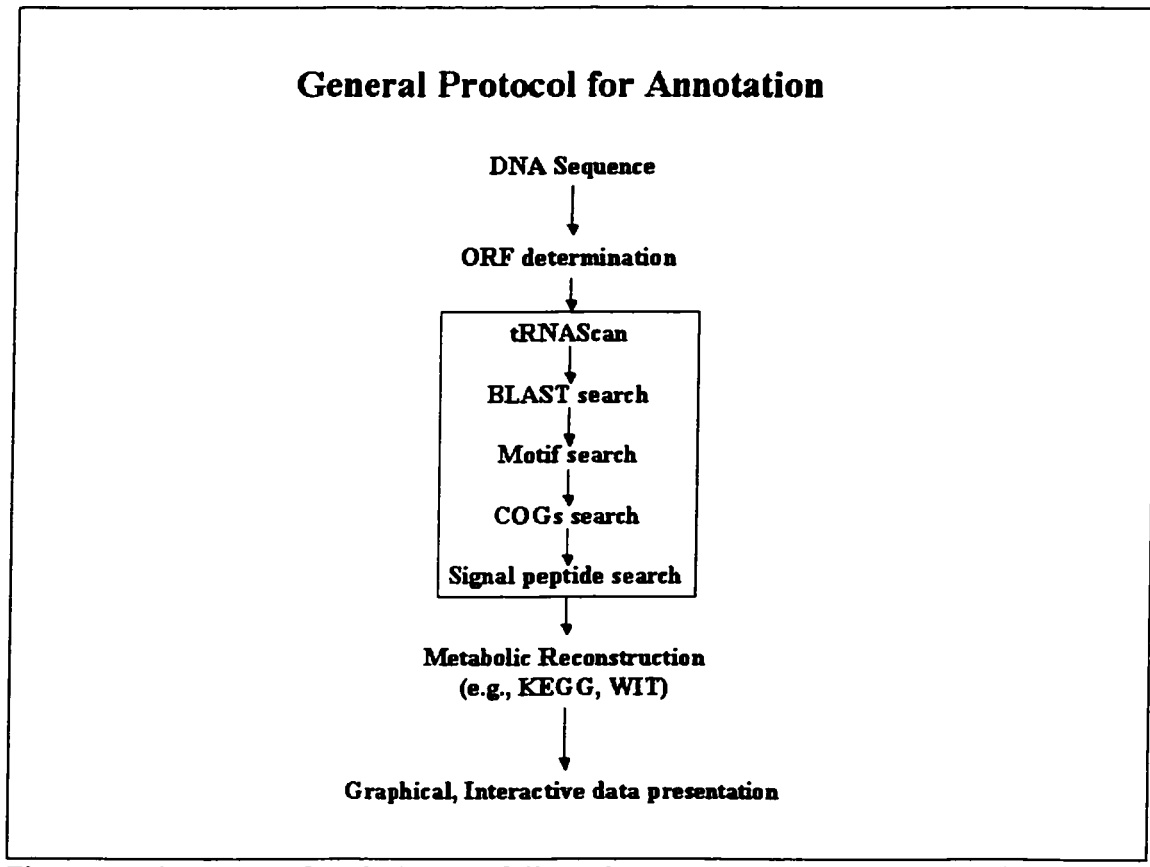


Figure 1.9. Overview of analysis steps followed to annotate and generate metabolic profile for bacteria.

In summary, during this dissertation research, the genome of *A. actinomycetemcomitans* has been sequenced and a general process has been developed to utilize the different gene feature finding programs to annotate this genome. As illustrated in Figure1.9 above, the first step is to determine ORFs in the genome. Next, the protein sequence of the ORFs is inferred from the DNA sequence. After a subsequent BLASTP search (basic local alignment and search tool for protein sequences), ORFs with 70% or more homology to known proteins are assigned to biological functions. Proteins with less homology are analyzed further for common patterns in several motif-based databases

associated with biological functions. Then, searches against the COGs database can be used to group proteins into functional categories that assist in explaining missing enzymatic functions in pathways. To obtain tRNA genes, the tRNAscan prgram was utilized. Finally, the data generated from these procedures are combined, categorized into a metabolic schema, visually represented and linked in a graphical interface that allows modification of annotation and interpretation of the data to facilitate further analysis.

Chapter II

Materials and Methods

2.1 *Actinobacillus actinomycetemcomitans* subclone library construction and shotgun sequencing

Chromosomal DNA of *Actinobacillus actinomycetemcomitans* strain HK1651 was supplied by Dr. David Dyer at The University of Oklahoma-Health Science Center using the Puregene™ DNA Purification Kit for Gram-negative bacteria (D-70KA). The strain is a clinical isolate obtained from Dr. Morgan Kilian at the University of Aarhus, Denmark **(Poulsen et al. 1994)**.

Shotgun sequencing requires shearing the genome to fragments, typically 2-6 Kb, and end-repair to create blunt ends. The fragments were 5' phosphorylated prior to ligation into the vector. The vector DNA used in this project was pUC18 that was pretreated with *Sma*I restriction enzyme to generate blunt ends, and dephosphorylated to prevent spontaneous recirculization of the vector (Amersham 27-4860-00).

2.1.1. Nebulization

Large-insert and small-insert libraries were created for the shotgun phase. The large-insert library was prepared by nebulizing the chromosomal DNA at 6 psi, and -15^{0}C. While the small-insert library was prepared by nebulizing the chromosomal DNA at 8 psi, and -15^{0}C. At the later stages of this work, a third library, generated by PCR-based methods, was developed.

Nebulization was performed by dissolving 100-200 μg chromosomal DNA in 2.5 ml of 50% sterile glycerol-sterile-distilled deionized water (v/v) in an IPI nebulizer (Medical Products, Inc., Chicago). After placing the nebulizer in a -15^0C dry ice-ethanol bath, nitrogen gas was introduced from the top of the nebulizer at the appropriate pressure for two and a half minutes. The gas flow causes the solution to be drawn up the tube to the upper chamber where it is spread, splashing it onto a small plastic protrusion to generate fine droplets. The size of the droplet, hence the size of the fragments of DNA, is inversely proportional to the pressure applied as illustrated in figure 1.5.

The solution containing the sheared genomic DNA then was divided into four 1.5 ml snap-cap tubes (0.5 ml solution/tube) and the DNA fragments were precipitated by adding two volumes of ethanol-acetate solution (95% ethanol/0.12M sodium acetate). After the DNA fragments were collected by centrifugation, the pellet was washed with 70% ethanol. Each pellet then was suspended in 30 μl of 1X TM buffer (50 mM Tris-HCl, pH 8.0, 15 mM $MgCl_2$).

2.1.2 End Fill-in, kinase reactions

The fragments generated from nebulization were end-repaired and phosphorylated by treating them with the Klenow fragment of DNA polymerase and T4 polynucleotide kinase after adding 5 μl of 10X kinase buffer (500 mM Tris-HCl, pH7.6, 100 mM $MgCl_2$, 10 mM DTT, and 50 μg/ml BSA), 5 μl of 10 mM rATP, and 7 μl of 0.25 mM dNTPs.

The reaction was incubated for 30 minutes at 37^{0}C and then halted by adding 10 µl of agarose loading dye (0.02% bromophenol blue, 5 mM EDTA, pH 8.0, 50% glycerol).

2.1.3 Fragment size selection

The end-repaired, phosphorylated DNA was loaded on to a 1% low-melt agarose gel in parallel with molecular size markers (*Hind*III-treated λ-DNA mixed with *Hae*I-treated ϕX174-DNA from Gibco BRL). After electrophoresis (Figure 2.1 above), the appropriate size-containing bands were excised and placed into a 1.5 ml snap-cap microfuge tube which in turn was frozen at –70^{0}C.

After thawing and centrifugation in a table-top microfuge at 13,000 rpm to pellet the agarose, the supernatant containing the DNA fragments in the supernatant was transferred into a new 1.5 ml snap-top microfuge tube. This "freeze and squeeze" step was repeated two more times and each time the supernatants were pooled. Ethanol/acetate (95% ethanol/0.12 M sodium acetate) then was added to the pooled supernatant to precipitate the DNA fragments followed by collection by centrifugation. After the DNA pellet was washed twice with 70% ethanol, it was dissolved in 15 µl of sterile-distilled-deionized water and stored frozen at –20^{0}C.

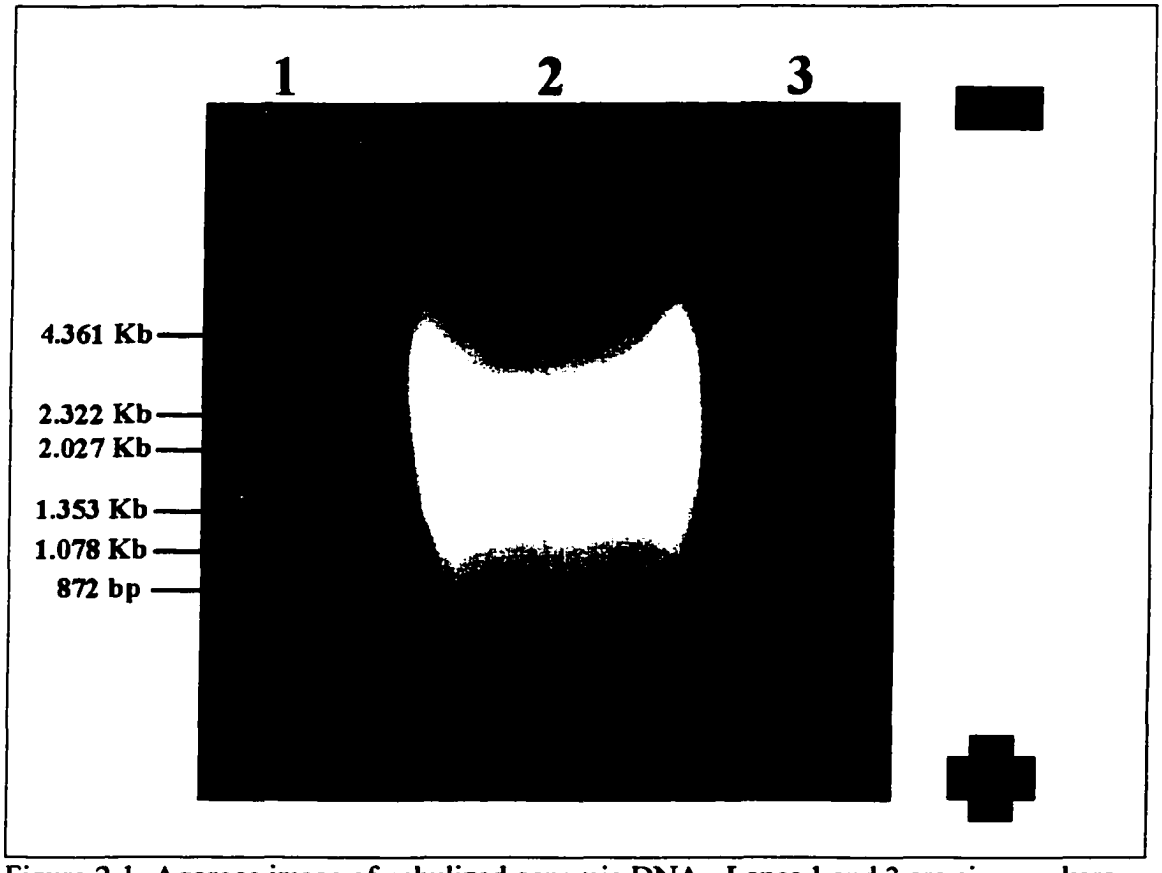


Figure 2.1. Agarose image of nebulized genomic DNA. Lanes 1 and 3 are size markers (λ-*HindIII* mixed with ϕX174-*HaeIII*). Lane 2 is *A. actinomycetemcomitans* DNA nebulized at 8 psi. Generally, fragments in the 1-2 and 2-4 Kb range were excised from the gel.

2.1.4 Subcloning the fragments and transformation

After five µl of the extracted fragments were examined on a 1% agarose gel, a series of ligation reactions were set up according to the following table:

	1X	2X	4X
DNA fragments	1 µl	2 µl	4 µl
pUC18 (10 ng/µl)	2 µl	2 µl	2 µl
Ligation buffer	1 µl	1 µl	1 µl
sdd-water	5 µl	4 µl	2 µl
T4-DNA ligase (400U/µl)	1 µl	1 µl	1 µl

The reactions were incubated at 4^0C for 18 hours.

Following the ligation, 2 µl of each ligation mix was suspended in 40 µl of electrocompetent *E. coli* XL1-Blue MRF' for transformation using an electroporator (Bio-Rad *E. coli* Pulser). Electroporation was accomplished by subjecting this cell-DNA mixture to an electrical pulse of 2.5 KV for 5 microseconds at 4^0C, after which time the cells immediately were diluted with 1 ml of cold YENB medium (7.5 grams of Bacto Yeast Extract, and 8 grams of Bacto Nutrient Broth brought to 1 liter with distilled water and autoclaved), and allowed to recover by incubation at 37^0C for 20 minutes with shaking at 250 rpm. The transformed cells then were collected by centrifugation at 2500 rpm for 5 minutes and resuspended in 200 µl of fresh YENB medium supplemented with 30 µl of 25 mg/ml 5-bromo, 4-chloro, 3-indolyl,β-D-galactoside (Xgal; Sigma B-4252) and 30 µl of 25 mg/ml isopropyl thiogalactoside (IPTG; Sigma I-5502). IPTG was included as a lac operon inducer analog of galactose which can not be cleaved by the enzyme β-galactosidase on pUC18, and Xgal was present as a chromogenic substrate that is hydrolyzed by intact β-galactosidase to form an intense blue color. One hundred and thirty µl of the resuspended cells were plated on 10 cm diameter plastic Petri dishes containing 20 ml of LB agar supplemented with 100 µg /ml ampicillin, which then were incubated at 37^0C for 18 hours. White colonies were picked, using Flexys colony picker™, into 1.5 ml of TB broth (12 grams of Bacto-tryptone, and 24 grams of yeast extract were brought to 900 ml with distilled water, autoclaved and cooled. Then, 100 ml of 10X TB salts (2.31 g KH_2PO_4, and 12.54 g K_2HPO_4 autoclaved and cooled) were added to yield 1 L of TB broth containing salts). Ampicillin then was added to the final

concentration of 100 µl/ml to the TB broth with salts, and 1.5 ml of this media was added to each well of 96-deep-well blocks (Beckman #140504). After inoculation of each well with individual colonies, the blocks were incubated for 20 hours at 37^{0}C with shaking at 350 rpm for aeration. Cell pellets were collected at 3000 rpm for 7 minutes and frozen at –20^{0}C. Later in the course of this dissertation research, protocols were developed for utilization of smaller cell-growth media volumes. Here, colonies were picked into flat-bottom microtiter plates (Dynatech Cat. No. 001-012-9200 12.7 cm x 8.5 cm) containing 150 µl of TB broth supplemented with 100 µg/ml ampicillin. These smaller microtiter plates then were incubated in a HiGro™ incubator (Gene Machines, Inc.) at 37^{0}C for 18 hours with shaking at 520 rpm with an oxygen flow set to begin 3.5 hours after shaking started, and a full open flow rate with the HiGro oxygen flow setting at 0.5 second on and 0.5 minutes off. After this 18 hours incubation, the cells were collected by centrifugation at 2500 rpm for 10 minutes and frozen at –80^{0}C.

2.1.5 Semi-automated isolation of subclone template DNA for sequencing

Two methods were developed for template isolation during the course of this research. The first utilizes deep-well blocks for growth and subsequent isolation while the second utilizes flat-bottom microtiter plate for growth and isolation. For deep-well blocks, cell pellets were suspended in 200 µl of TE-RNase solution (50 mM Tris-HCl, pH 7.6, 0.5 M EDTA-Na, 40 µg/ml RNase A and 0.04 U/ul RNase T1), and the blocks were placed on a table-top shaker at a setting of 7. After 10 minutes of shaking, 200 µl lysis buffer (1% SDS, 0.2 M NaOH) was added and the blocks were further incubated on the shaker at a setting of 4 for 30 minutes. Then, 200 µl of 3 M KOAc (294.45 g KOAc

in a total volume of 1 L adjusted to pH 4.5 with acetic acid) was added and the blocks were incubated at 37^{0}C with shaking at 350 rpm for 20 minutes and frozen at –70^{0}C for 10 hours at which time they were centrifuged at 4500 rpm for 45 minutes. Two hundred µl of the resulting supernatant was transferred to a new block, and DNA was precipitated by adding 500 µl of 95% ethanol. After centrifugation for 30 minutes at 3000 rpm, the DNA pellet was washed with 1 ml of 70% ethanol and dissolved in100 µl sterile-distilled deionized water (sdd-water). The DNA then was evaluated for insert size and the presence of contaminating materials by electrophoresis on 1% agarose gel. At a later stage of this dissertation research, colonies were picked into flat-bottom microtiter plates containing 150 µl instead of 1.5 ml of TB broth. Consequently, 60 µl instead of 200 µl of each of the above three solutions were added and treated as described for the deep-well plates. After freezing overnight, the plates were centrifuged for 30 minutes at 3000 rpm, and then 60 µl of the supernatant was transferred into each well of a V-bottom microtiter plate (Dynex brand VWR Cat.No. 62402-914) and 130 µl of 95% ethanol was added using the Hydra (Robbins Scientific, Inc.) to precipitate the DNA. The plates then were centrifuged at 3200 rpm for 30 minutes, the supernatant was decanted, and the pellets were washed with 150 µl of 70% ethanol. After centrifuging the plates for 10 minutes and decanting the excess ethanol, the DNA was dried and then dissolved in 50 µl of sdd-water and evaluated by electrophoresis on a 1% agarose gel, as above.

2.1.6 Cycle sequencing conditions

Sequencing was performed using the BigDye terminator kit (PE Applied Biosystems cat.# 430151) which contains Taq polymerase buffer (composition

unknown), AmpliTaq FS DNA polymerase, the four dNTPs, and the four dye-labeled dideoxynucleotide terminators. The AmpliTaq enzyme contains two modifications, an N-terminal deletion which eliminates the 5'-3' editing activity of the enzyme to allow more processivity, and a phenylalanine to tyrrosine mutation in the active site which enhances its affinity for the fluorescent-labeled nucleotide terminators (**Tabor et al. 1995**). As explained in the introduction, each terminator contains attached fluorescent dyes, a fluorescine-derivative (donor dye) connected to a dRhodamine-derivative (acceptor dye). The donor dye can be excited by an argon ion laser source contained in the PE Applied Biosystems DNA sequencer. Energy emitted from the first dye in turn excites the electrons in the dRhodamine acceptor which emits light that is detected, by a PMT in the ABI-377 or a CCD camera in the ABI-3700, when they return to their ground state (**Rosenblum et al. 1997**). The dRhodamine-derivative acceptor dyes, dR6G, dROX, dTAMARA, were attached to dR110 for ddATP, ddCTP, ddTTP, and ddTTP respectively. The structures of the d-rhodamine dyes are given in figure 1.3. Approximately 150-200 ng of template DNA, 1 μl of 13 μM universal forward (GACGTTGTAAAACGACGGCC) or universal reverse (CACAGGAAACAGCTATGACC) primers, dimethoxysulfoxide (DMSO) at a final concentration of 5% (V/V) and 2 μl of the ABI BigDye™ reaction kit premix containing AmpliTaq FS, thermostable pyrophosphatase, dATP, dCTP, dTTP (100 μM each), dITP (500 μM), ddATP, ddCTP, ddTTP, and ddGTP (~ 0.11 μM each) at a final concentration of $1/12^{th}$, $1/16^{th}$, or $1/20^{th}$ that recommended by the supplier were added to each well in a 96 (Robbins Scientific, Inc, cat.# 1055-00-0) or 384 well thermocycle plate (Robbins Scientific, Inc, cat.# 1047-00-0). The reactions then were thermocycled for 60 cycles of

denaturation at 95^{0}C for 30 seconds, annealing at 50^{0}C for 20 seconds, extension at 60^{0}C for 4 minutes. During the latter gap closure phase of this dissertation research, cycle sequencing reactions in which ABI-BigDye mix was replaced with either the ABI-BigDye-dGTP (ABI# 4307175) or ABI-dRhodamine mix (ABI# 4303143) were performed for regions that could not be successfully sequences using the standard ABI-BigDye mix, and incubated as described above.

2.1.7 Removal of unincorporated terminators

Unincorporated terminators and buffer salts were removed from the sequencing reactions either by filtering through Sephadex G50 mini-columns or by ethanol precipitation. For Sephadex removal, columns were prepared in a 96-well filter plates (Millipore, cat.# madvn6550) by adding dry Sephadex G-50 (Amersham, cat.# 17-0043-02) to each well using a 45 µl column loader (Millipore, cat.# MACL 096 45). After hydration with 300 µl of distilled-deionized water and incubating for at least 3 hours at 4^{0}C, the Sephadex containing plates were centrifuged at 1500 rpm for 3 minutes and an additional 100 µl dd-water was added to each well and centrifuged as above.

After the cycle sequencing reaction was completed, 10 µl of dd-water was added to each reaction using a Robbins Hydra96. This mixture then was transferred to the top of the Sephadex columns, that were placed over a 96 well collection plate and centrifuged at 1500 rpm for 3 minutes. The collected, purified reactions were dried and stored at –20^{0}C until loading.

At a later stage of this project, the dye was removed from the sequencing reactions by ethanol precipitation. Here, 10 µl of dd-water was added to each well in the

thermocyle plate bringing the final volume to about 20 µl, and two volumes (40 µl) of 95% ethanol-acetate then were added to each well. The thermocycle plate then was centrifuged at 3200 rpm for 30 minutes and after the ethanol-acetate was decanted, the precipitated reaction products were washed with 100 µl 70% ethanol and centrifuged for 10 minutes at 3200 rpm. After decanting the ethanol wash, plates were dried for 5 minutes at room temperature and stored at -20^0C until ready for loading onto the sequencer.

2.1.8 Sequencing

The shotgun sequencing products were electrophoretically resolved on either a ABI PRISM 377 slab gel, or the ABI PRISM 3700, a capillary sequencer. For the ABI PRISM 377, 5.3% acrylamide (FMC) containing 8 M urea was poured on a 48 cm, low fluorescent glass plates separated with 0.2 mm spacers. After polymerization, each sample was suspended in 1 µl loading buffer (0.02% blue dextran, 0.2 mM EDTA in deionized formamide) and loaded. Electrophoresis then was performed at 2.5 kV and 52 degrees C for 10 hours in 1X TBE buffer (10.8 g Tris, 5.5 g boric acid, and 8.4 g EDTA in total volume of 1 liter).

For the ABI PRISM 3700, capillaries were automatically filled with flowable gel mix POP-5 (ABI, Inc., Performance Optimized Polymer type 5 cat.# 4313087) and the reactions were electrokinetically injected after suspending them in 20 µl dd-water. Electrophoresis was performed at 6.5 Kv for 2.5 hours. Because of the small inner diameter (50 µm), the temperature gradient within the tube is minimized and the heat dissipation is rapid allowing application of high electric field.

The data was collected automatically and analyzed using the ABI base caller on the Macintosh (for ABI377) or Dell-PC computer (for ABI3700). The data then were transferred to a Unix-based SUN work station.

2.1.9 Sequence assembly

The data collected from the sequencers was transferred via the file transfer protocol (ftp) to the project directory on the SUN station. There, the trace files (the chromatograms generated from the sequencers for each reaction) were analyzed by Phred software (**Ewing et al. 1998**) that writes the base sequence from and assigns quality values to each base to generate output files in PHD format that are passed to Phrap software for assembly (**Gordon et al. 1998**).

Phred analyzes the data by first determining where the peaks would be centered if there were no compressions, or other factors shifting the peaks from their theoretical calculated locations. Then, it examines each trace to find the centers of the actual, or observed peaks and matches the observed peaks detected in the second step with the predicted peak locations found in the first step (**Ewing et al. 1998**). Finally, Phred evaluates the traces surrounding each called base using quality value parameters to quantify the trace quality (**Ewing et al. 1998**) . The quality value (QV) is related to the base call error probability (P_e) by the formula

$$QV = -10 * \log_{10}(P_e)$$

Phred uses data from a chemistry parameter file called 'phredpar.dat' in order to identify dye primer data. This information then is written into individual PHD files named according to each original trace file and passed along to the Phrap assembly program **(Ewing et al. 1998)**. FastA files containing the formatted base sequence then are generated from the PHD files. These are used by the cross_match program to screen out the vector sequence. This program compares the sequence obtained from Phred with sequences in "seqs_fasta". Phrap (Fragment Assembly Program) uses the data generated from Phred to assemble the reads and generate a contiguous consensus sequence integrating the quality scores assessed by phred to construct theses consensus sequences or contigs. This in turn can be viewed and edited by a Phrap viewer program Consed **(Gordon et al. 1998)**. Consed, a program intended for interactively viewing and editing assemblies generated from the Phrap assembly program, utilizes the quality files generated by Phred and the output assembly files generated by Phrap to create an accurate and interactive view of the data. The Consed view consists of the consensus sequence of each contig showing the individual reads that compose it. The individual reads are linked to both the fluorescent chromat trace files, generated from the DNA sequencer, and the quality files from Phred **(Gordon et al. 1998)**. The latter files are represented with different coloring schemes, with white background and capital letters representing high quality to a dark shade of gray and lower case letters representing poor quality data. Additionally, Consed presents the repeats in the sequence data represented by different shades of green which can be linked directly to repeats present on different contigs **(Gordon et al. 1998)** . The high-quality discrepancies can be viewed as red color which can indicates chimeric clones. These too are linked to different contigs.

2.2 Gap Closure Phase of Sequencing

2.2.1 Primer-walking. Large-insert clones

The principle behind primer walking is to use large-insert clones sequences generated using universal forward and reverse primers to point to potential "bridging or spanning clones" covering any gaps between contigs in the assembled sequence. Repeat sequences often are present at the regions flanked by the gaps. The Phred assembly program often misassembles these regions by overlapping repeats into one contiguous sequence shortening the original sequence or by misassembling regions that flank different copies of repeated sequences. Sequencing bridging or spanning clones (as above) to completion, via a direct custom-synthesized primer walking approach, ensures that such misassembly does not occur.

Walking primers were picked using the PrimOU program, which is an improved adaptation of the Primo program developed at the University of Texas-Southwest Medical Center. PrimOU identifies unique primer sequences that have been screened against either the contiguous sequence from which it originated, or against the all contig sequences collected. It is automatically activated when Phrap finishes assembling the data. The output of the primer sequences is stored as a text file containing the contig number, the direction, and the sequence of the primer. Additionally, PrimOU utilizes the quality assessment for the bases ensuring accuracy. Output files generated by PrimOU were sent electronically to the MerMade oligonucleotide synthesizer (Avantec, Inc.). For

A. actinomycetemcomitans, the primers picked were according to the parameters shown in the table below:

Primer Length (bp)	20
Melting Temperature (Tm)	60°C +/- 1°C
GC%	40
Uniqueness	14/20

All walking primers were synthesized on a Mermade oligonucleotide synthesizer, the commercial version of an instrument, originally designed at the University of Texas Southwestern Medical Center which can synthesize 192 primers in two 96-well filter plates in about 16 hours. The standard phosphoramidite chemistry for the synthesis process as illustrated in figure 2.2 below (**Beaucage et al. 1981**). This instrument contains an XY platform mounted inside an isolated chamber filled with argon gas to provide a dry and inert atmosphere for synthesis (synthesis chamber) where the two 96-well filter plates are set. The gas and the reagents are dispensed into each well and a vacuum pump is used to transfer the reagents into a waste container after each synthesis step.

The DNA synthesis process is composed of four steps plus the final deprotection step. The synthesis progresses from the 3'-5' direction utilizing fully protected deoxynucleotides as illustrated in figure 2.3. Briefly, the first protected deoxynucleotide (the 3' end of the oligonucleotide sequence) which is linked to a solid support called CPG

(Controlled Pore Glass) is detritylated using trichloroacetic acid. Next, the CPG linked deoxynucleotide is coupled to the second detritylated deoxynucleotide via tetrazole. This produces a phosphite nucleotide intermediate and a small percentage of deoxynucleoside-phosphoramidite that will remain as truncated sequences at the end of the synthesis. The capping process occurs through acetylation of the free 5'hydroxyl group and then the unstable trivalent phosphite triester is oxidized using iodine to a more stable pentavalent phosphate triester.

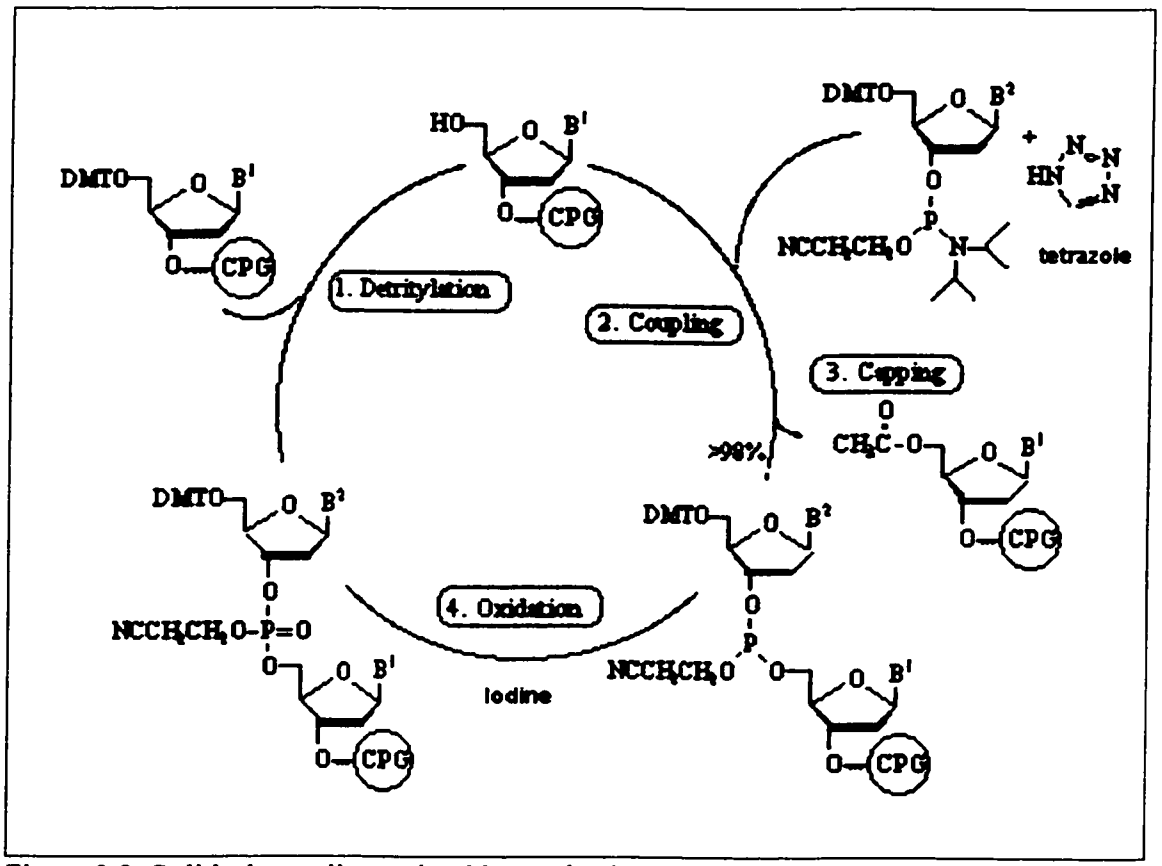


Figure 2.2. Solid-phase oligonucleotide synthesis.

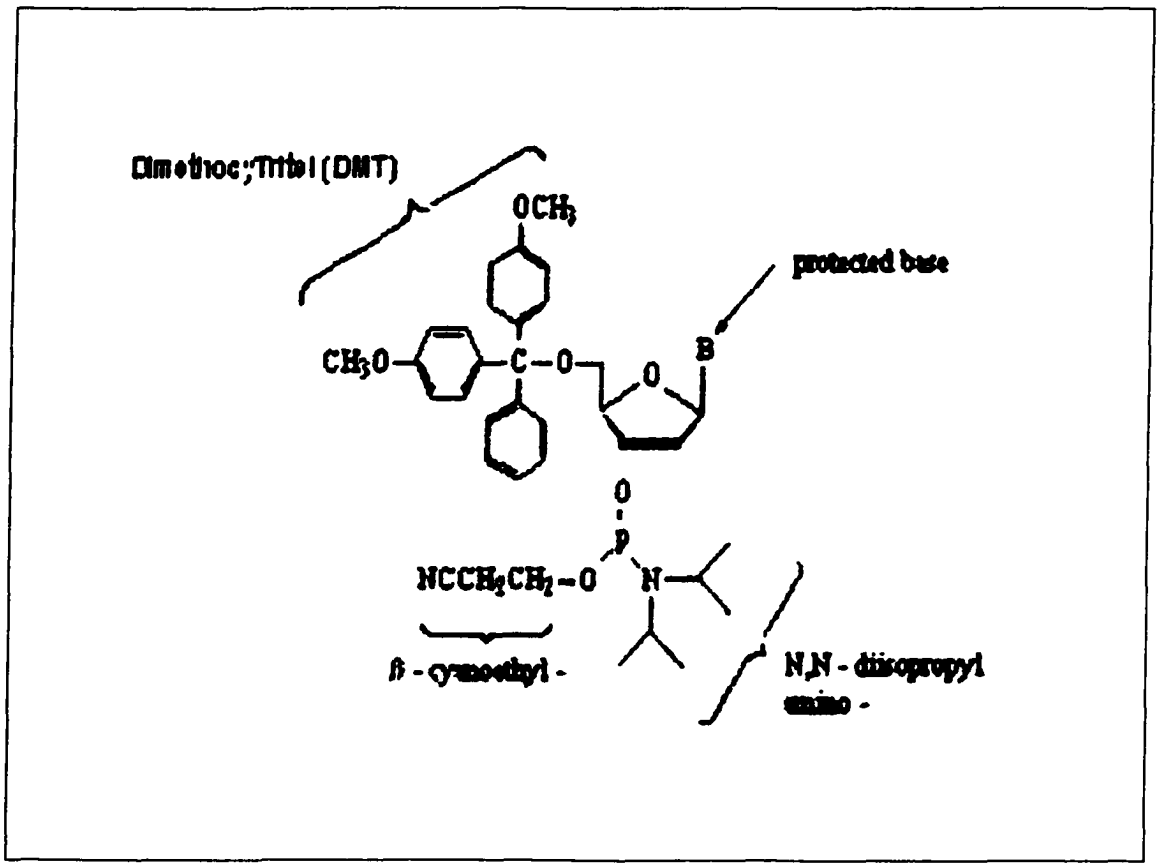


Figure 2.3. An illustration of protected nucleoside-3'-phosphoramidites.

This instrument was pivotal in allowing extensive use of a directed high throughput custom-primer approach applied to DNA sequence closure because of the low cost (~$1/primer), the speed, yield, and reliability. Typically, an average of 2 nmole of primer was synthesized in each well of the 96 well plate on the Mermade.

2.2.2 Multiplex polymerase chain reaction (MPCR)

Although sequencing spanning clones using walking primers was quite successful, some gaps remained which were not spanned by any large-insert sub-clones, or sequencing indicated there were long repeats in these clones that might have

undergone rearrangements changing the original sequence (**Behnke et al. 1979**). The physical gaps often could be spanned using a PCR-based approach as no cloning step is required for producing sequencing templates (**Mullis et al. 1987**). Although the uniplex PCR technique, employing a single set of primer pairs, is very useful if the order of the contigs is known, when the order of the contigs is uncertain, it is more useful to use a multiplex PCR (MPCR) (**Clauster et al. 1989**) approach. In the course of this dissertation research, this method was adapted for use in our high-throughput sequencing environment. This method was developed specifically for use when the order of contigs was unclear. Primers were designed from the ends of all contigs, synthesized, and then combined for MPCR using genomic DNA as template. After sequencing the PCR products generated, the primers that yielded a sequence were subtracted from the pool and the remaining primers were pooled for another round of MPCR.

In the MPCR method, if more than 20 primers were to be used, then 1 µl of each 6.5 µM primer was pooled into a microfuge tube, dried, and dissolved in 10 ul sterile-distilled water. These primers were mixed with 10 µl of 5X MPCR buffer (83 mM $(NH_4)_2SO_4$, 335 mM Tris-HCl (pH 9.0), 33.5 mM $MgCl_2$, 50 mM beta-mercaptoethanol, 850 µg/ml bovine serum albumin, and 34 mM EDTA), about 100 ng of genomic DNA, 12 µl of dNTP mix (25 mM each), 2 units of Taq polymerase XL (Perkin-Elmer# N808-0193), and 2.5 µl DMSO. Sterile distilled deionized water was added to the a final volume of 50 µl. The reaction mix was thermocycled on a Perkin-Elmer GeneAmp® 9700 (# 4314879). The sample was denatured for 6 minutes at 94^0C , and then cycled for 30-40 cycles for 30 seconds of denaturing at 94^0C, annealing for 30 seconds at 55^0 C , and

extension for 4 minutes at 65^{0}C, at which time the samples were held at 4^{0}C until further analysis.

To degrade the excess primers, ten units of shrimp alkaline phosphatase (SAP) and one hundred units of exonuclease I were added to the MPCR reaction products which then were incubated for 30 minutes at 37^{0}C. The temperature was elevated to 80^0 C for 10 minutes to denature the enzymes, and then the reaction was stored at 4^0 C until needed. The resulting MPCR products were evaluated by electrophoresis on 1% agarose gel. The reaction products then were extracted once with an equal volume of phenol/chloroform (1:1), precipitated with 2.5 volumes of ethanol-acetate and washed with 70% ethanol. The samples were dried and resuspend in 50 ul of sterile distilled deionized water.

For sequencing, two microliters of the MPCR products were distributed into each well of a Robbins™ 96-well thermocycle plate. Two microliters of each 6.5 μM primer used in the MPCR reaction were pipetted individually into each well in the 96-thermocylcle plate using the Hydra96 along with 1 μl of 25% DMSO, 1 μl of BigDye mix. Samples then were thermocycled according to the BigDye thermocycling conditions described above. Sequencing reactions were cleaned by filtration through Sephadex G50 columns or by direct ethanol precipitation as described above.

2.3 Computer methods and data analysis

2.3.1 Sequence analysis

When the sequence coverage of *A. actinomycetemcomitans* reached about 6 genome equivalents, the annotation process was begun. There are two reasons for

starting the annotation prior to completing the sequencing. First, at 6 fold coverage, about 95% of the sequence data has been obtained (**Lander et al. 1998**), allowing for a fairly accurate observation of the gene content. Second, data annotation can help closure as discussed earlier by indicating where different contigs might contain a portion of a single gene or by comparative analysis of a related genome of known sequence to align one or more contigs because of conserved gene order. Here, the knowledge of the gene organization in other well-characterized organisms helped to order separate contigs and aid closing sequence gaps in the *A. actinomycetemcomitans* genomic data. For example, if two or more contigs in *A. actinomycetemcomitans* carried different genes that belonged to the same region in another well-characterized organism, it was likely that these contigs also could occur in the same order in *A. actinomycetemcomitans*. This observation easily could be verified by PCR, and often was found to be correct, therefore providing additional sequencing templates that aided the closure and finishing process.

The basic protocol for preliminary annotation of the *A. actinomycetemcomitans* genome is illustrated in figure 1.9. Analysis of the data began by creating a file containing all contigs that were larger than 2 kb in length. This was done by artificially joining the ends of the random contig sequences with a string of 20 Ns representing any type of base, according to the IUB code. The file generated then was input to the Artemis program (**Rutherford et al, 2000**) that allowed viewing the sequence and identifying open reading frames (ORFs) based on the distance between consecutive stop codons above a given sequence length set by user-defined parameters, usually 100 bases (Figure 2.4). The marked ORFs then were analyzed by a series of operations controlled through

Perl scripts written by James White in Dr. Roe's laboratory. The first script performs a BlastP search of the protein sequence for each ORF (**Altschul et al. 1997**). The second script, artemis_blast2EMBL, extracts the best homology information from each Blast output file, defined by the top scores, generated for each ORF, and converts it to EMBL format so the results could be viewed by Artemis (Figure 2.5)

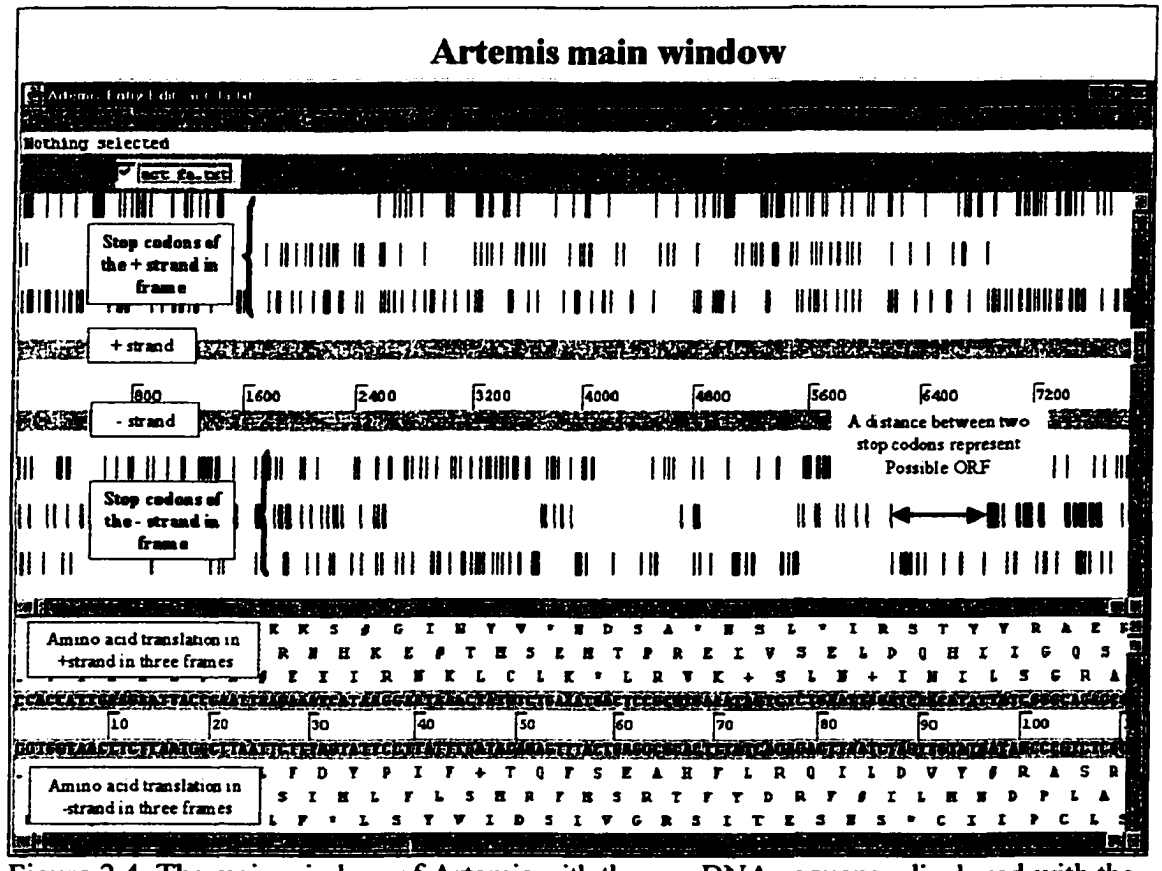


Figure 2.4. The main window of Artemis with the raw DNA sequence displayed with the six frame translations.

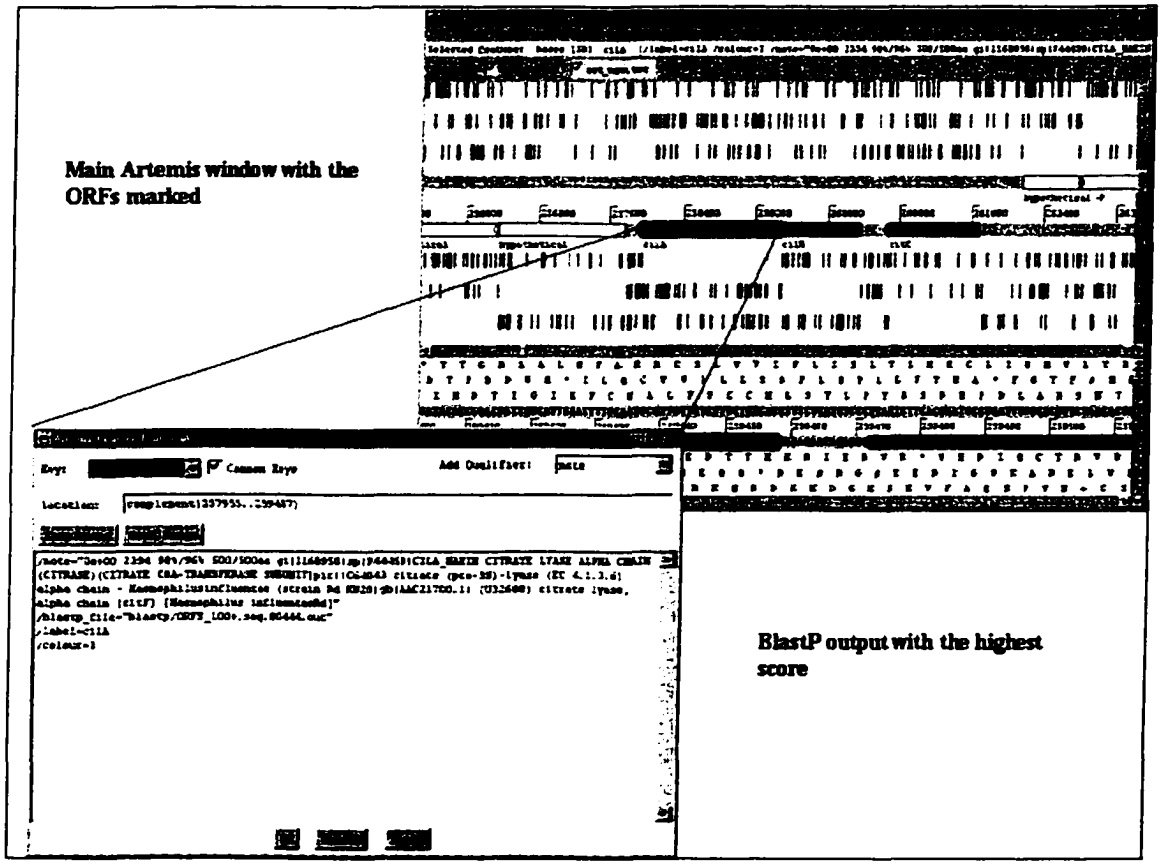


Figure 2.5. Displaying Blast output of marked ORFs in Artemis.

To analyze the BlastP homology results further, a separate list of all known enzymes and proteins involved in the physiology of *E. coli* and other studied prokaryotes modeled after The Reily metabolic schema (**Neidhardt 1999**) was drafted. This list contained all the proteins known in *E. coli* as well as a description of the enzyme, its Enzyme Commission number (EC) when available, and its gene name (figure 2.6). This list of bacterial proteins facilitated categorizing the Blast output into an easy to understand format and also provided information about the metabolic characteristics of *A. actinomycetemcomitans*. The schema was a simple text file that consisted of two columns, the first column was the protein's EC number, if it had a well-studied catalytic

activity, or the most accepted three-letter gene name followed by a number or letter (usually from *E. coli*, and sometimes from the organism in which the gene was first characterized). The second column contained a description of the protein starting with its systematic name and ending with information identifying the gene, and when known, its possible links to different enzymes in metabolic pathways (Figure 2.6). Since several proteins often are involved in more than one pathway, this too was noted to avoid counting them more than once.

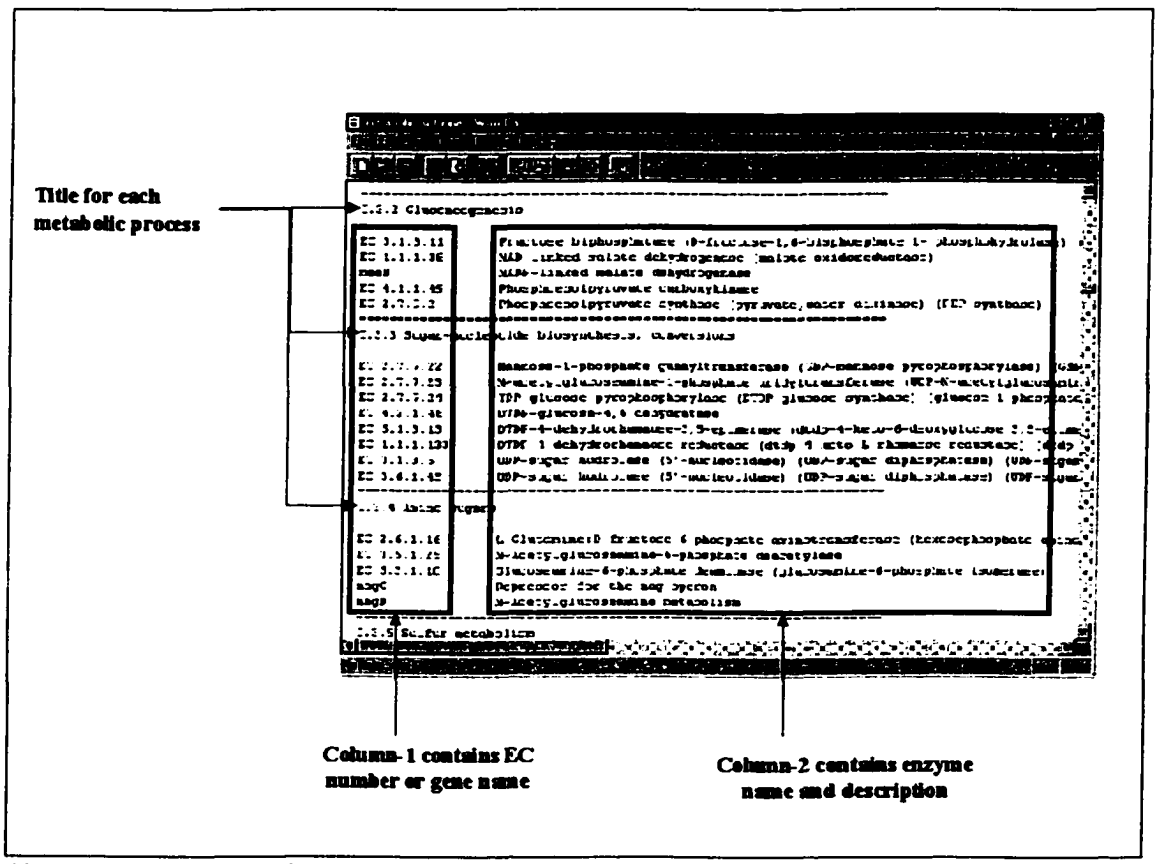


Figure 2.6. Text file of the metabolic schema.

A perl script written and developed by Hongshing Lai in Dr. Roe's laboratory called get_EC that performed a character-search using the first column in the metabolic

schema text file (EC numbers and gene names) as keyword to search the output files generated by BlastP and reported the results in an output file similar to the original metabolic schema text but had a third column with the number of hits and the corresponding BlastP output file that contain the protein of interest (figure 2.7).

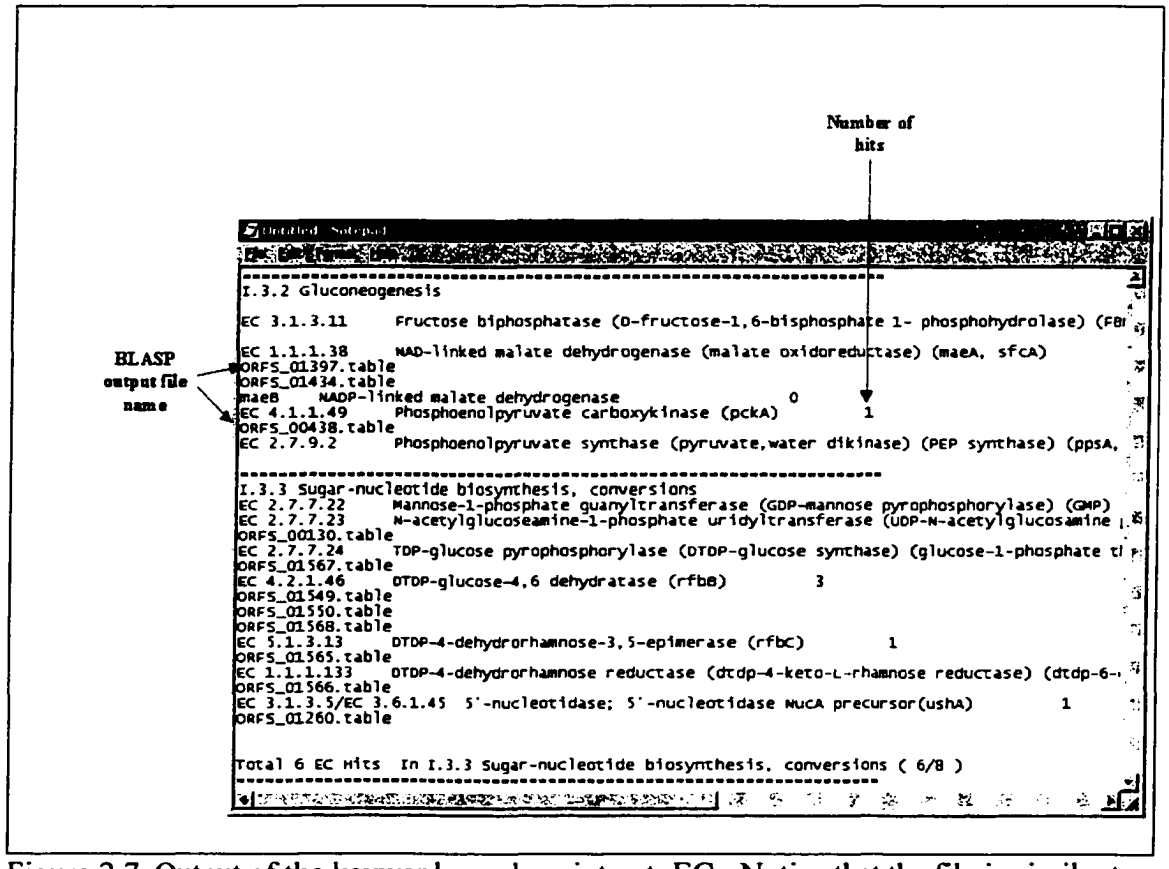


Figure 2.7. Output of the keyword search script get_EC. Notice that the file is similar to the one illustrated in figure 2.6 with the two noted exceptions.

The major problem encountered with the BlastP output was that EC numbers and the short gene abbreviation for enzymes often were not included in the GenBank entry. Therefore, a manual search of the output files was required to insure that no proteins were missed by the automated keyword search. In addition, when a protein had no assigned

EC number, the protein name was used. However, due to variation in spelling protein names causing difficulty in recognition. For example, it was possible to miss the enzyme name if the GenBank entry omitted or replaced a hyphen or a comma as illustrated below:

```
>
575  4e-59      act_1521889:1522500      N-
Acetylglucoseamine-6-phosphate deacetylase (GLCNAC 6-P
DEACETYLASE)  [Haemophilus influenzae]
```

In this example, the enzyme N-Acetylglucoseamine-6-phosphate deacetylase (EC 3.5.1.25) has different names. The name given in the Blast output is N-Acetylglucoseamine-6-phosphate deacetylase and GLCNAC 6-p deacetylase with no EC number or short gene name. Hence, an automated keyword search may miss this enzyme if only the EC number was used in the search. For this reason, an additional manual search had to be performed. Currently, an improved script is being developed by James White that uses both the EC number and multiple versions of a gene name along with the protein name which will collapse all versions to a group of keywords. Thus, the enzyme "**5-Enolpyruvylshikimate-3-phosphate synthetase (3-phosphoshikimate 1-carboxyvinyltransferase) (EPSP synthase)**" would be collapsed to "**shikimate/phosph**" which are the common words found in all the names of this enzyme.

The remaining ORFs with no significant homology in the GenBank database or with homology to other proteins with no known functions then were analyzed separately for known motifs and COG homologies (figure 2.8), to obtain additional information

about the protein's possible biological function. Here, the MOTIF program was used to examine all unknown function ORFs. The resulting output from MOTIF displayed the detected motif signature and the corresponding protein from which the motif was derived (figure 2.8). The data was compared with the COGnitor output that displays the (cluster of orthologous groups (COG) and the pathway or biological function of the protein (figure 2.8). The MOTIF and COG outputs are summarized in appendix A.

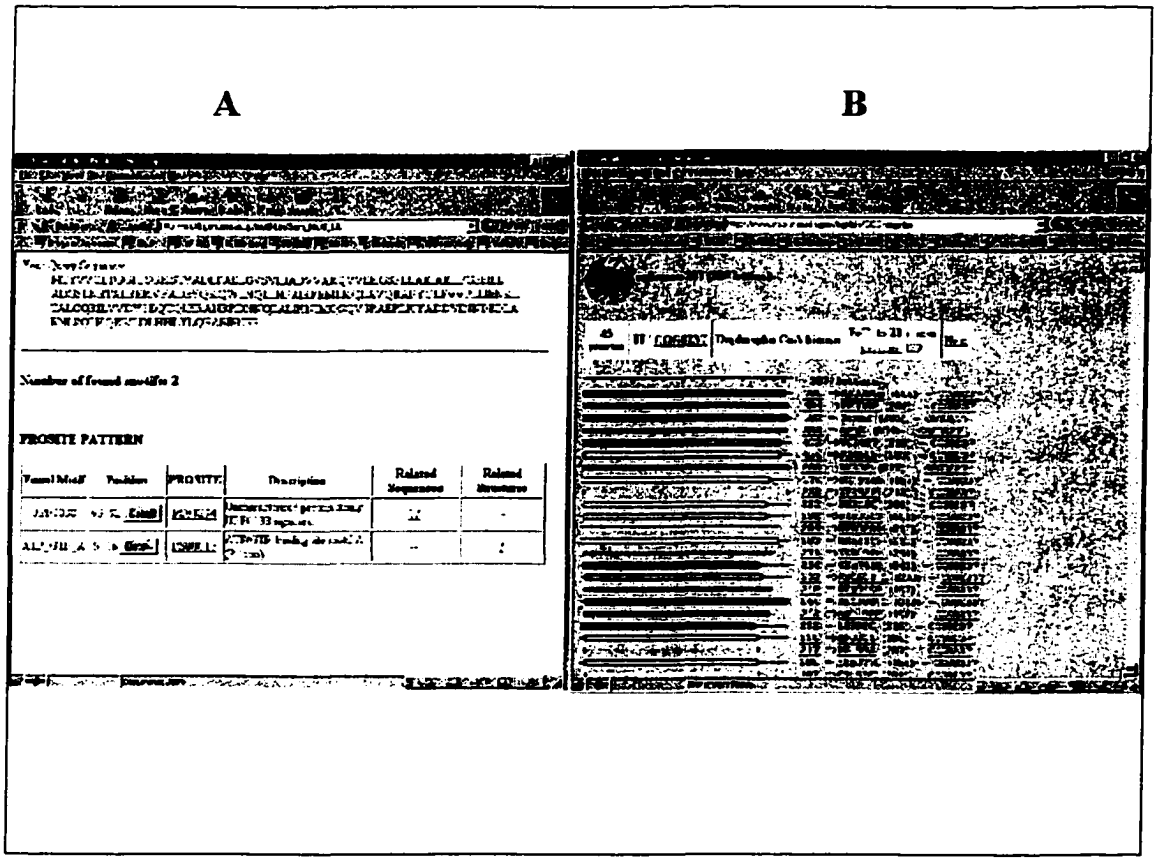


Figure 2.8. Panel A, Output of MOTIF program showing motif sequences found in a given protein. Panel B, output of COGnitor showing the possible function for the given protein with list of homologues in that COG.

After assigning putative functions to each ORF, the data was imported into an Excel spreadsheet as illustrated in figure 2.9. This format contains information similar to the text file mentioned above but it contains additional information such as a link to the peptide sequence of each ORF, the homologies to similar proteins from different organisms and the MOTIF and COGnitor information for each ORF. It is designed to hold all annotation information and a link to that information for future updates. The advantage of using a spreadsheet is the flexibility and ease to add different data and representing it visually. Additionally, it is relatively easy to automate data extraction and change the visualization mode to customize it to the type of data analyzed.

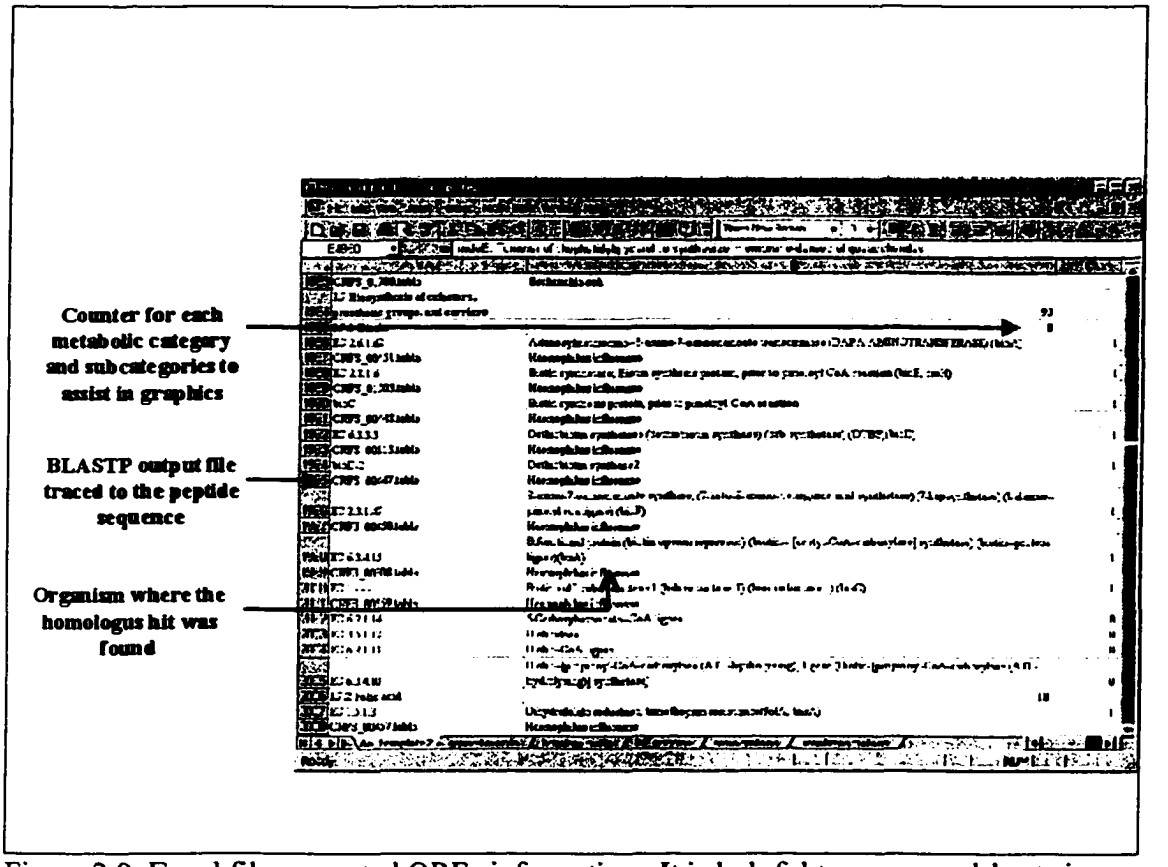


Figure 2.9. Excel file generated ORFs information. It is helpful to use spreadsheets in order to generate graphical representation of the data. Additionally, Excel can be programmed to perform different statistical analysis which, in turn, also, can be represented graphically.

To analyze secretory proteins, the amino acid sequences of all ORFs generated by Artemis were analyzed by the SignalP program (**Nielsen et al. 1997**), which uses a neural network algorithm that is trained to recognize signal peptides and protein cleavage sites based on validated sequences. Usually, signal peptides are characterized by having a positively charged N-terminal region (n-region), followed by a hydrophobic region (h-region) and a C-terminal region (c-region) rich in neutral, polar amino acids (**Von Heijne, 1984**). When the protein sequence is input, the user specifies which model organism characteristics should be used for the signal peptide search. The output of the program includes a graphical illustration representing the score of each of the regions and the cleavage site predicted for the protein (**Nielsen et al. 1997**) as shown in figure 2.10.

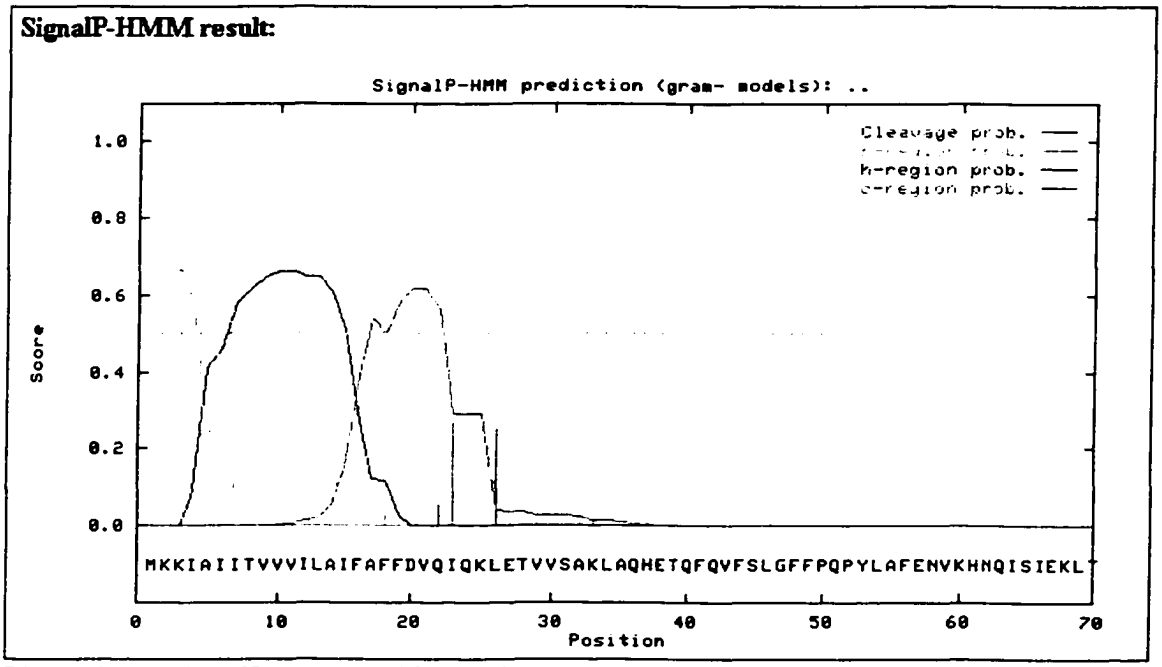


Figure 2.10. SignalP output. This graph represent a typical signal sequence where the n-, h-, and c-regions are distinguished and vividly separated as seen from the three curves. Interestingly, this protein has no known assigned function and moreover, the MOTIF and COG search produced no homology hits. This shows the necessity to use as many analytical tools as possible to elucidate function.

The data generated from SignalP were input in Excel.

Genes that encode for tRNAs were revealed using the tRNAscan-SE. A program developed by Eddy and colleagues (**Eddy et al. 1994**) using the default parameters. A typical output file includes the base coordinates of each tRNA gene, as well as the tRNA type, and anticodon as shown below in figure 2.11.

tRNAscan output

Sequence Name	tRNA #	tRNA Begin	Bounds End	tRNA Type	Anti Codon
Aa	1	15956	16029	Arg	CCG
Aa	2	16064	16136	His	GTG
Aa	3	16142	16215	Pro	TGG
Aa	4	39624	39714	Ser	GCT
Aa	5	188325	188397	Thr	TGT
Aa	6	188431	188512	Tyr	GTA
Aa	8	222817	222890	Pro	TGG
Aa	9	222927	223000	Arg	TCT
Aa	10	590421	590493	Lys	TTT
Aa	11	590523	590595	Lys	TTT
Aa	12	590624	590697	Lys	CTT
Aa	13	655342	655414	Gly	GCC
Aa	14	655426	655496	Cys	GCA
Aa	15	875477	875563	Ser	GGA
Aa	16	968252	968324	Gly	GCC
Aa	17	968329	968413	Leu	TAA
Aa	18	968469	968541	Lys	TTT
Aa	19	1189190	1189262	Glu	TTC
Aa	20	1275241	1275313	Glu	TTC
Aa	21	1639948	1640020	Ala	GGC
Aa	22	1915021	1915103	Leu	CAA
Aa	23	1686402	1686329	Ile	GAT
Aa	24	1686276	1686204	Ala	TGC
Aa	25	1638138	1638066	Val	TAC
Aa	26	1638024	1637952	Val	TAC
Aa	27	1637921	1637849	Val	TAC

Figure 2.11. The typical output file form tRNAscan

The output file is Figure 2.12 summarizes the annotation protocol used for *A. actinomycetemcomitans*.

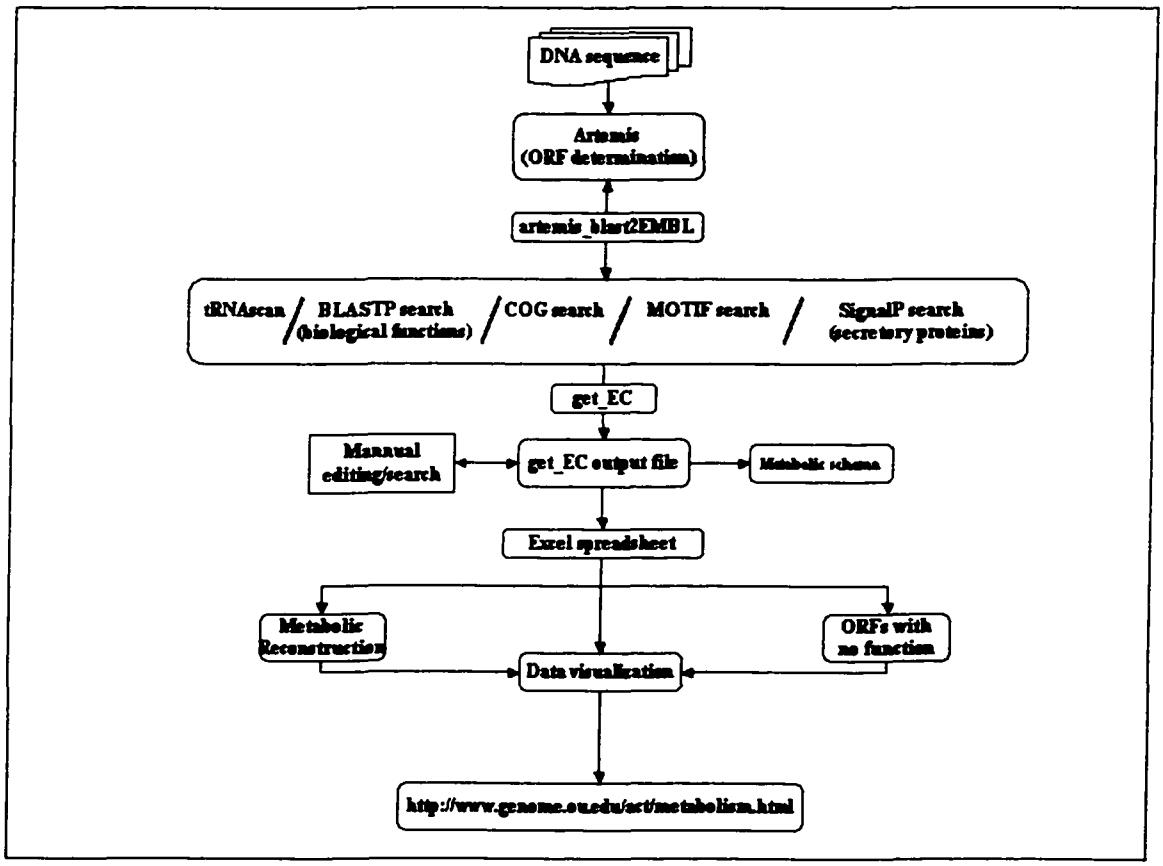


Figure 2.12. Summary of the analysis and annotation process used for *A. actinomycetemcomitans*.

Once the above analyses was completed, the results were entered into a master Excel table and the sequence data was input into KEGG (**Kanehisa et al. 2000**) to aid in constructing putative metabolic pathways. The master Excel table and the metabolic reconstruction were converted to HTML code to display the data interactively on the internet along with the reconstructed metabolic pathways. This data can be found at http://www.genome.ou.edu/act/metabolism.html.

Chapter III

Results and Discussion

3.1 Sequence statistics and quality

Four whole genome shotgun libraries of the *A. actinomycetemcomitans* genomic DNA were constructed as templates for dideoxynucleotide sequencing. The whole genome short-insert library had an average insert size of 1.5 Kb and was sequenced initially to generate approximately 5-6 fold genome coverage. Then, with over 95% of the sequence available, a PCR-based approach, mainly in the form of MPCR, was used to generate templates for sequencing gap and weakly covered genome regions. Since there was no physical map available for *A. actinomycetemcomitans*, and there were no BAC library clones available for this genome, a PCR approach was quite useful to confirm the assembled sequence to improve the sequence quality in weakly covered regions and to close many sequence gaps. The PCR-based approach also was used to span predicted split genes and split operons as described above. Hence it aided the completion or closing of "easy gaps" early in the sequencing project of this dissertation research leaving mainly sequence gaps containing repetitive sequences at the ends of contigs. In an effort to close the remaining gaps, the MPCR-based approach was used to obtain multiple PCR products. This approach was very worthwhile and quite useful to close several gaps. However, when the MPCR products were cloned to generate a second library set, these MPCR-generated inserts were of little use as they often produced a high percentage of chimeric clones, i.e. clones containing two or more inserts that falsely ordered contigs. Therefore, all MPCR-based sub-clone sequences were removed from the database and

only the non-cloned MPCR product sequence results were included in the sequence assembly process.

In an effort to obtain additional cloned sequencing templates, a third library containing larger inserts were generated. This library contained what often are called "walking clones", as they had inserts that were 2-6 Kb. These walking clones were helpful in generating read pairs , i.e. the forward and the reverse sequence reads from the same clone template that were of great use in ordering contigs and in detecting miss-assembled areas of the sequence. In addition, many of these walking clones were sequenced to completion to obtain the accurate sequence of regions containing large repeats. A fourth library containing genomic inserts in a PAC vector was generated for us by Dr. Jon Coren at the Southwestern Oklahoma State University. However, when this library was analyzed by end-sequencing, it was shown to contain mainly *E. coli* inserts, with fewer than 10% inserts from the *A. actinomycetemcomitans* genomic DNA and therefore it was not used furhter. Table 3.1 below summarizes the sequencing statistics for the libraries and PCR-based approach used in this study.

Library type	**Number of reactions**
Short-insert	15686
Large-insert	24161
uniplex PCR	1175
Direct MPCR product sequencing	441
Phage sequence	1937
Total	43400
Confirmed	39092
Success rate	90.1

Table 3.1. Sequencing statistics of *A. actinomycetemcomitans*.

As of this writing, a total of 43676 sequencing reactions have been performed with a 90% success rate. This resulted in 39307 confirmed reads entered into the *A. actinomycetemcomitans* assembly database and a total of 2,024,943 bp of unique genomic sequence representing 99.98% of the genome. Figure 3.1 illustrates the distribution of the resulting contigs based on their size.

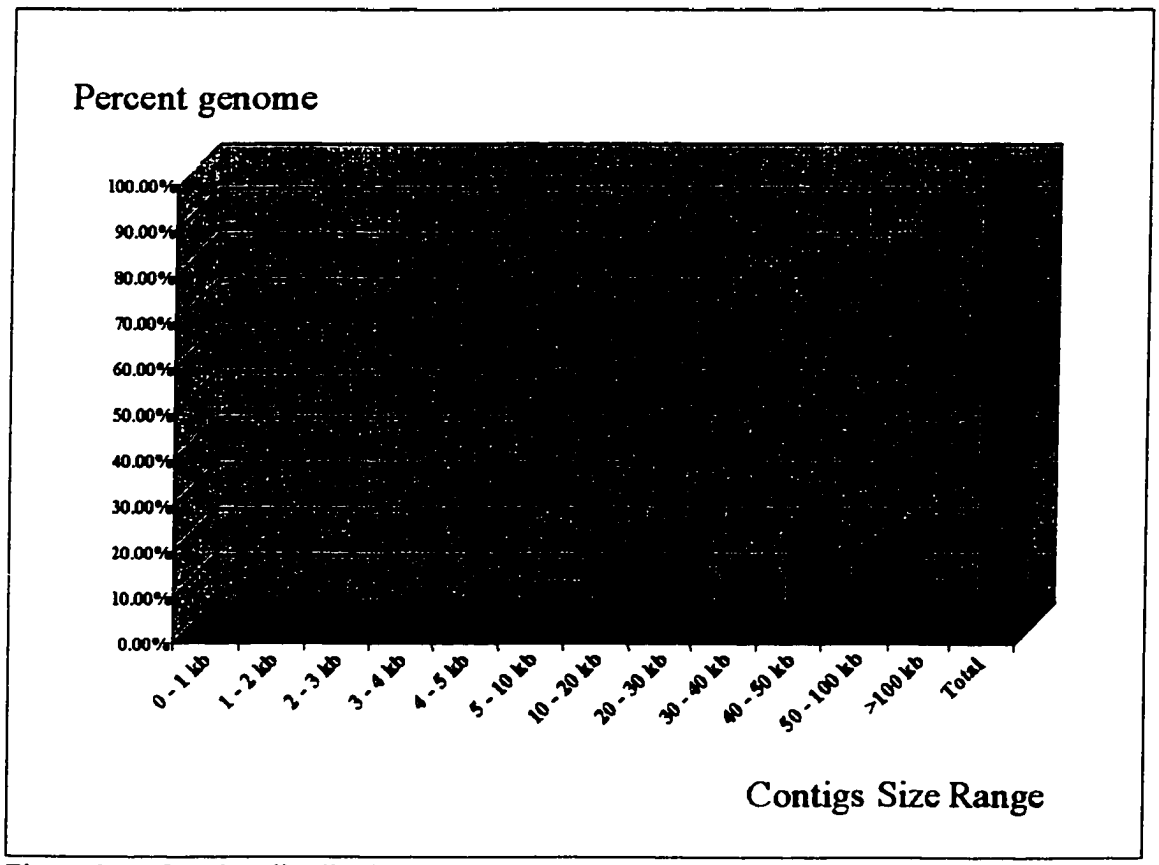


Figure 3.1. Contigs distribution.

The closure strategy for the remaining gaps likely will entail utilizing a PCR-based approach where suitable read pairs may be obtained based on, for example, GC content and trinucleotide frequency.

3.2 Genome overview

Analysis of the *A. actinomycetemcomitans* genome reveals 1877 open reading frames (ORFs) of which approximately 32% code for proteins with no known function. Of these, roughly one-fifth, i.e. 6% of the total, are unique to *A. actinomycetemcomitans* as they have no significant GenBank homologies, while the remaining four-fifths, or 25% of the total, have homology to ORFs with no assigned functions in other organisms. The proteins encoded in the *A. actinomycetemcomitans* ORFs can be grouped into functions categorized as illustrated in figure 3.2.

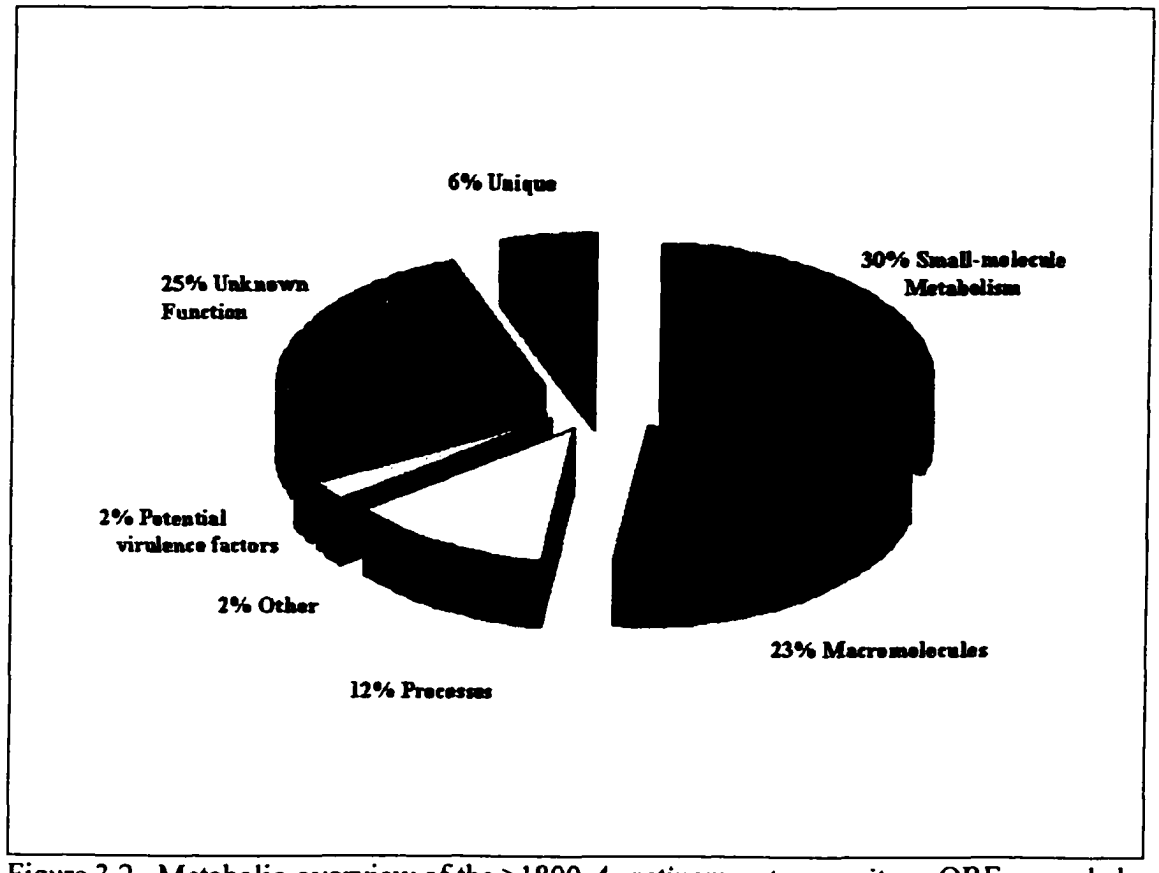


Figure 3.2. Metabolic overview of the >1800 *A. actinomycetemcomitans* ORFs revealed from the sequence. Appendix A Illustrates the details of the metabolic categories.

As seen in figure 3.3, over 60% of the sequence of the *A. actinomycetemcomitans* genome has *H. influenzae* as its closest bacterial homologue, an observation consistent with the recent effort to reclassify *A. actinomycetemcomitans* into the genus *Haemophilus* instead of *Actinobacillus* (**Potts et al., 1985**). The largest gene families observed one those encoding the over 80 putative ABC transporters, 17 putative penicillin-binding protein genes, and at least 16 insertion element regions were the next most prevalent.

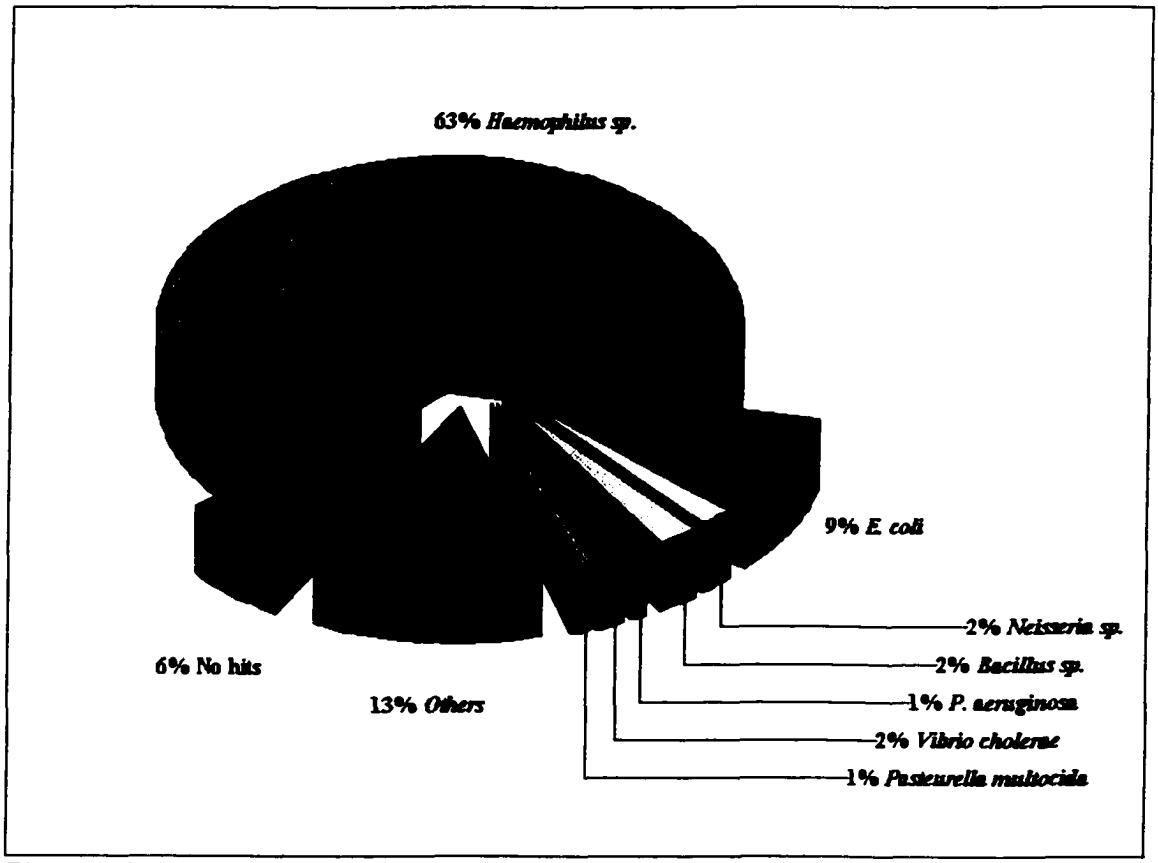


Figure 3.3. Homology profile of *A. actinomycetemcomitans*.

Codon usage, di-, and trinucleotide frequencies were calculated using a Perl script written by Jim White. Output results of the program were tabulated as shown below. In figure 3.4.

A. actinomycetemcomitans codon usage

	T			C			A			G		
T	Phe	15183	58%	Ser	4887	15%	Tyr	12733	68%	Cys	2990	49%	t
	Phe	11039	42%	Ser	8501	26%	Tyr	6118	32%	Cys	3099	51%	c
	Leu	20751	33%	Ser	3461	11%	OCH	1229	ALL	OPA	226	ALL	a
	Leu	17100	27%	Ser	3327	10%	AMB	293	ALL	Trp	6848	ALL	g
C	Leu	7010	11%	Pro	4132	18%	His	7044	56%	Arg	10546	39%	t
	Leu	4592	7%	Pro	2507	11%	His	5468	44%	Arg	9901	37%	c
	Leu	2953	5%	Pro	3147	14%	Gln	19317	69%	Arg	1748	6%	a
	Leu	10211	16%	Pro	12759	57%	Gln	8563	31%	Arg	2390	9%	g
A	Ile	24090	61%	Thr	4721	16%	Asn	15854	58%	Ser	5064	16%	t
	Ile	13040	33%	Thr	14457	49%	Asn	11464	42%	Ser	7386	23%	c
	Ile	2601	7%	Thr	4453	15%	Lys	30213	86%	Arg	1942	7%	a
	Met	13782	ALL	Thr	6173	21%	Lys	4914	14%	Arg	557	2%	g
G	Val	7861	20%	Ala	6143	12%	Asp	19243	66%	Gly	13217	34%	t
	Val	5731	15%	Ala	16117	33%	Asp	9730	34%	Gly	16761	44%	c
	Val	6936	18%	Ala	11140	23%	Glu	28984	83%	Gly	4102	11%	a
	Val	18502	47%	Ala	15765	32%	Glu	6077	17%	Gly	4439	12%	g

Figure 3.4. Codon usage in *A. actinomycetemcomitans*.

Figure 3.5 is a graphical representation which gives a visual assessment of the codon usage in comparison with that of *H. influenzae*. Interestingly, the general codon usage features both organisms appear to be very similar, i.e. relative to each genome, codons such as "aaa" of lysine, "gaa" of glutamate, and "gat" of aspartate are the most abundant. However, as can be seen in figure 3.5, the unique codon usage pattern in each organism that gives it a slightly different and distinct signature. For example, there are obvious differences in the usage for alanine, and threonine. Also, certain codons such as "cta" of leucine, and "ccg" of proline are used with different frequencies in each organism.

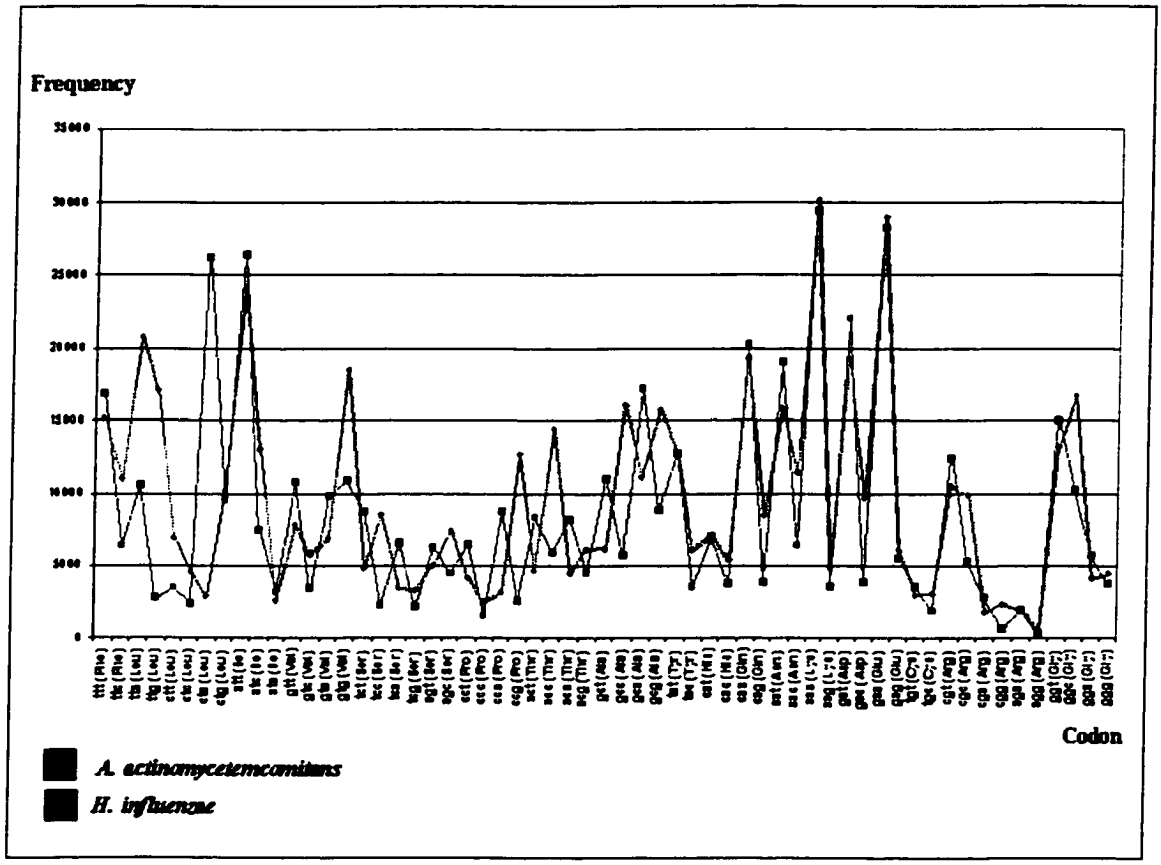


Figure 3.5. Codon usage of both *A. actinomycetemcomitans* and *H. influenzae*.

It has been reported previously that there is a correlation between the number of tRNA genes and codon usage (**Duret et al., 2000**). A similar pattern was observed in *A. actinomycetemcomitans* although it was not as dramatic as observed in *E. coli*. Figures 3.6 to 3.8 illustrate the correlation between the tRNA gene number and the codon usage in *A. actinomycetemcomitans*, *H. influenzae*, and *E. coli* respectively. The correlation appears stronger in *E. coli* than it does in either *A. actinomycetemcomitans* or *H. influenzae*.

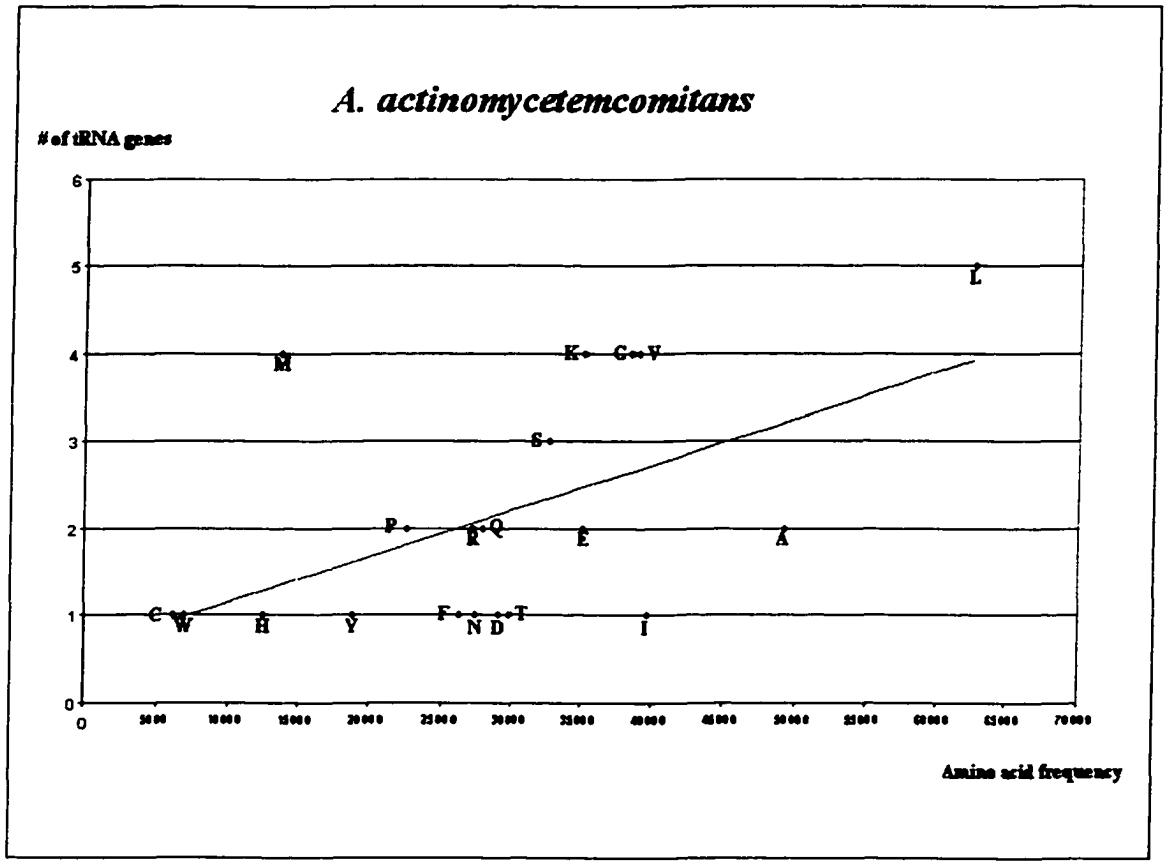


Figure 3.6. Correlation between the number of tRNA genes and cognate amino acid frequencies in *A. actinomycetemcomitans*.

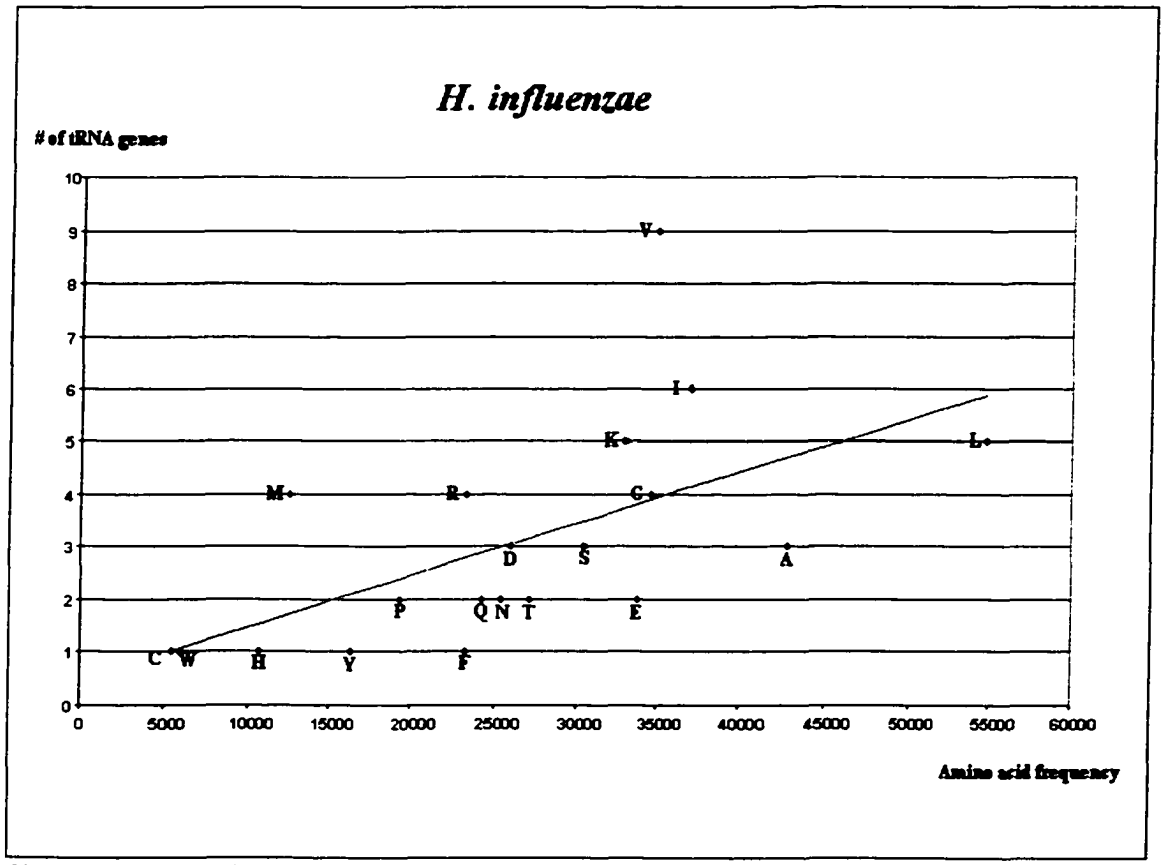


Figure 3.7. Correlation between the number of tRNA genes and cognate amino acid frequencies in *H. influenzae*.

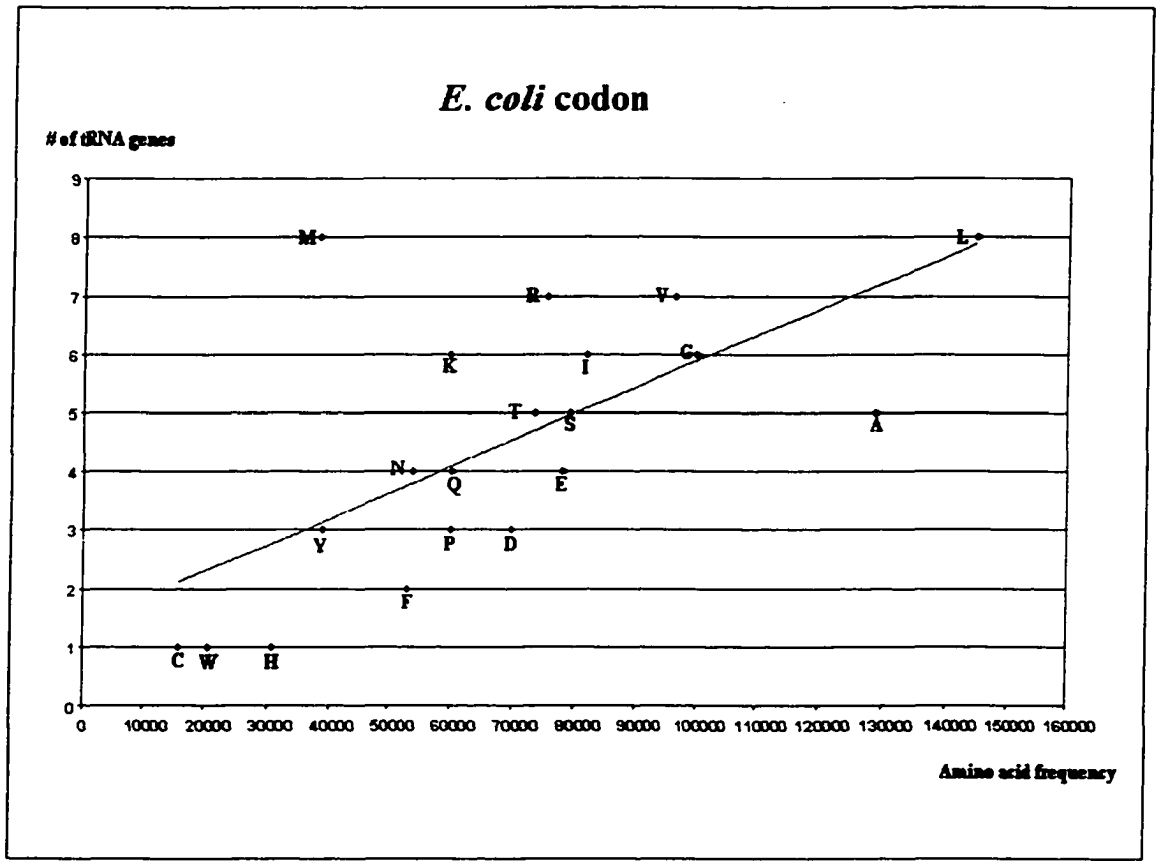


Figure 3.8. Correlation between the number of tRNA genes and cognate amino acid frequencies in *E. coli*.

The di- and trinucleotide frequencies of the *A. actinomycetemcomitans* genome and for several Gram-positive and Gram-negative organisms were calculated and are shown in figures 3.9 through 3.12. Although these frequencies are quite similar, *A. actinomycetemcomitans* seems to have a slightly higher GG (or CC), and GC (or CG) dinucleotide frequency and slightly lower AA (or TT), and AT (or TA) dinucleotide frequency than observed in *H. influenzae*, *H. ducreyi*, and *H. pylori*. In contrast, the GG (or CC), and GC (or CG) contents of *A. actinomycetemcomitans* are lower than observed in *E. coli*, and *N. meningitidis* whereas the AA (or TT), and AT (or TA) contents are higher than those observed in *E. coli*, and *N. meningitidis*.

Although similar observations were observed for the trinucleotide frequencies, the rationale for the correlation between these observations and the environmental niche or metabolic life style of these organisms is unknown.

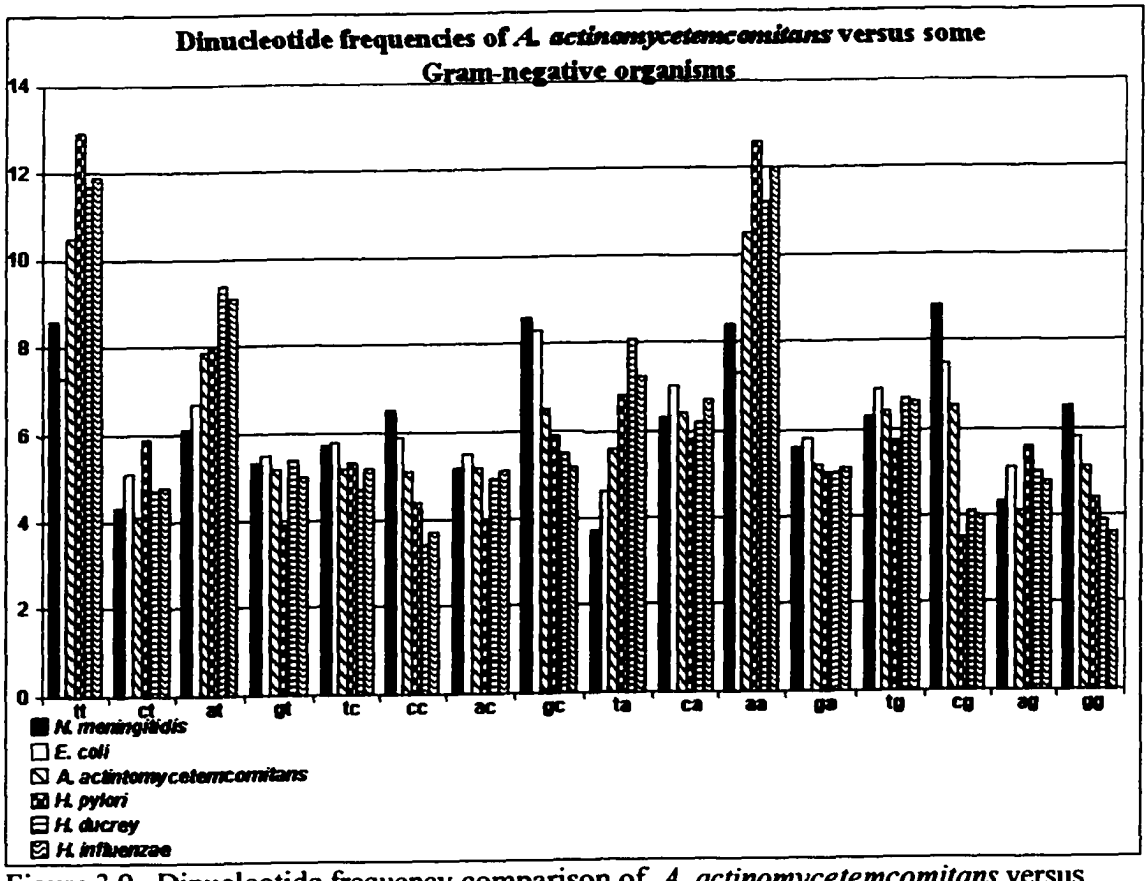


Figure 3.9. Dinucleotide frequency comparison of *A. actinomycetemcomitans* versus other Gram-negative bacteria.

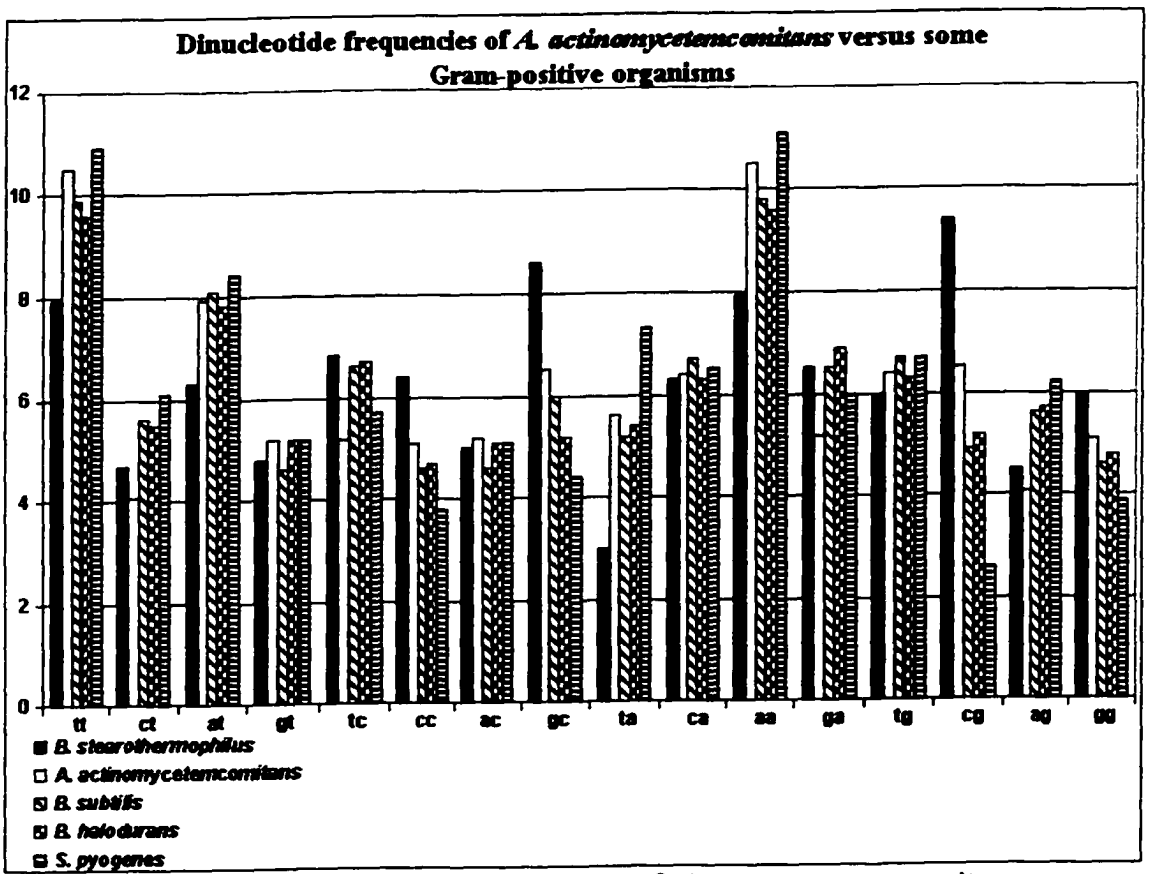


Figure 3.10. Dinucleotide frequency comparison of *A. actinomycetemcomitans* versus other Gram-positive bacteria.

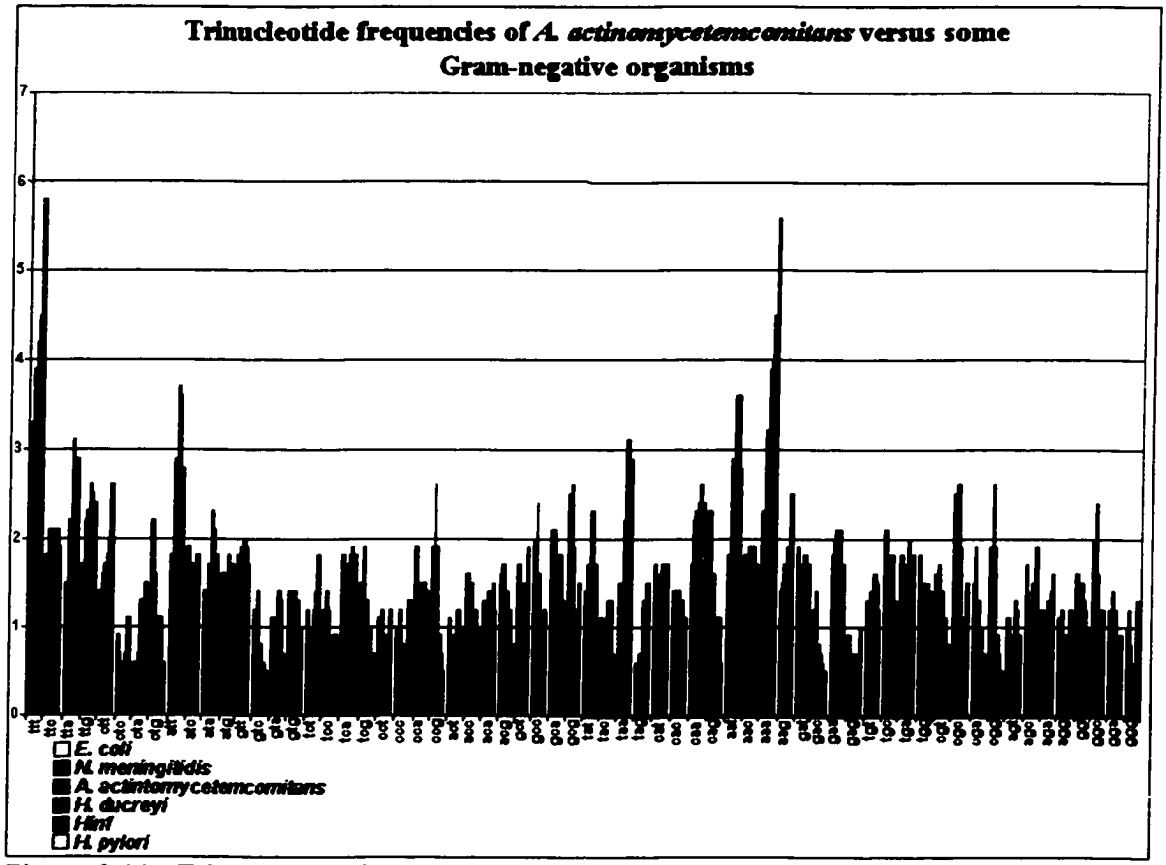


Figure 3.11. Trinucleotide frequency comparison across both of *A. actinomycetemcomitans* versus other Gram-negative bacteria.

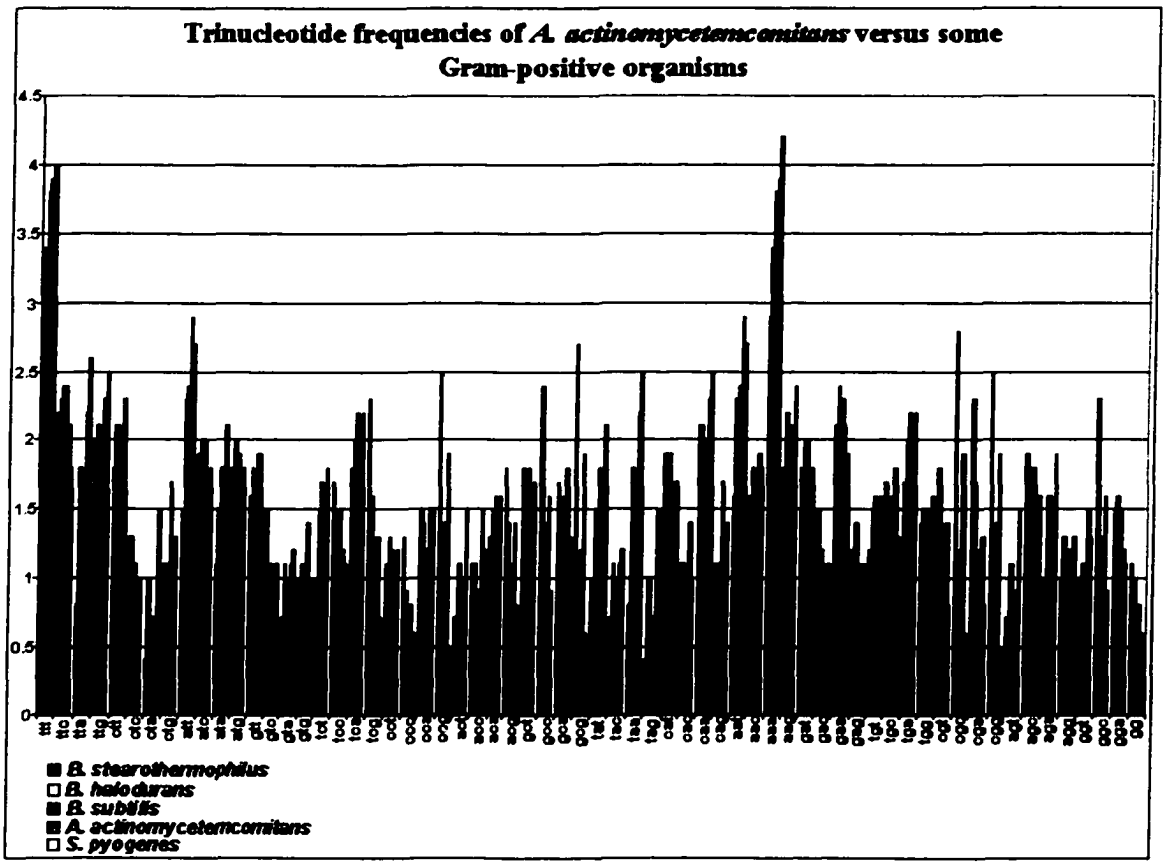


Figure 3.12. Trinucleotide frequency comparison across both of *A. actinomycetemcomitans* versus other Gram-positive bacteria.

3.3 *Actinobacillus actinomycetemcomitans* metabolism

Analysis of the sequence of *A. actinomycetemcomitans* genome reveals that glycolysis is the major pathway both for energy production and as a source of precursors for numerous biosynthetic pathways. Although *A. actinomycetemcomitans* lacks a complete Entner-Duodorof pathway, it does possess the complete pentose phosphate shunt which can produces NADPH and pentose sugars. The genes for the first three enzyme of the citric acid cycle also are missing from the *A. actinomycetemcomitans* genome, but the genes encoding the enzymes needed for the synthesis of a variety of amino acids, fatty acids and phospholipids were observed. In addition, the lipoprotein

and peptidoglycan biosynthetic pathways appear to be complete, as are pathways for vitamin and cofactor synthesis including thiamin, biotin, pantothenate, NAD(P), molybdenum cofactor, and riboflavin. *A. actinomycetemcomitans* has replication, transcription, and translation machineries similar to those in *E. coli*, and *H. influenzae* as well as the genes coding enzymatic pathways for DNA repair, restriction, and those encoding proteins involved in DNA recombination, and modification systems.

3.3.1 Energy

Analysis of the *A. actinomycetemcomitans* genome reveals the full complement of genes which encode for the proteins involved in small carbohydrate, fatty acid and amino acid metabolism. These enzymes are members of the central metabolic pathways where the precursors of the anabolic pathways are produced. A detailed description of *A. actinomycetemcomitans* genes predicted to encode proteins with energy-related functions including glycolysis, the pentose phosphate pathway, pyruvate metabolism, the TCA cycle, fermentation, respiration, and electron transport is described below.

3.3.1.1 Glycolysis

Glycolysis not only serves as an oxidative pathway for carbohydrates, but also provides precursors to many other pathways. As in most members of the family *Pasteurellaceae*, *A. actinomycetemcomitans* encodes the genes needed to produce a phosphotransferase system (PTS) (Section 3.3.8.3) which is dedicated to transport of variety of different carbohydrates including glucose, all of which can serve as substrates

for conversion to glucose and enter glycolysis. Figure 3.13 summarizes glycolysis with the enzymes involved listed in table 3.2

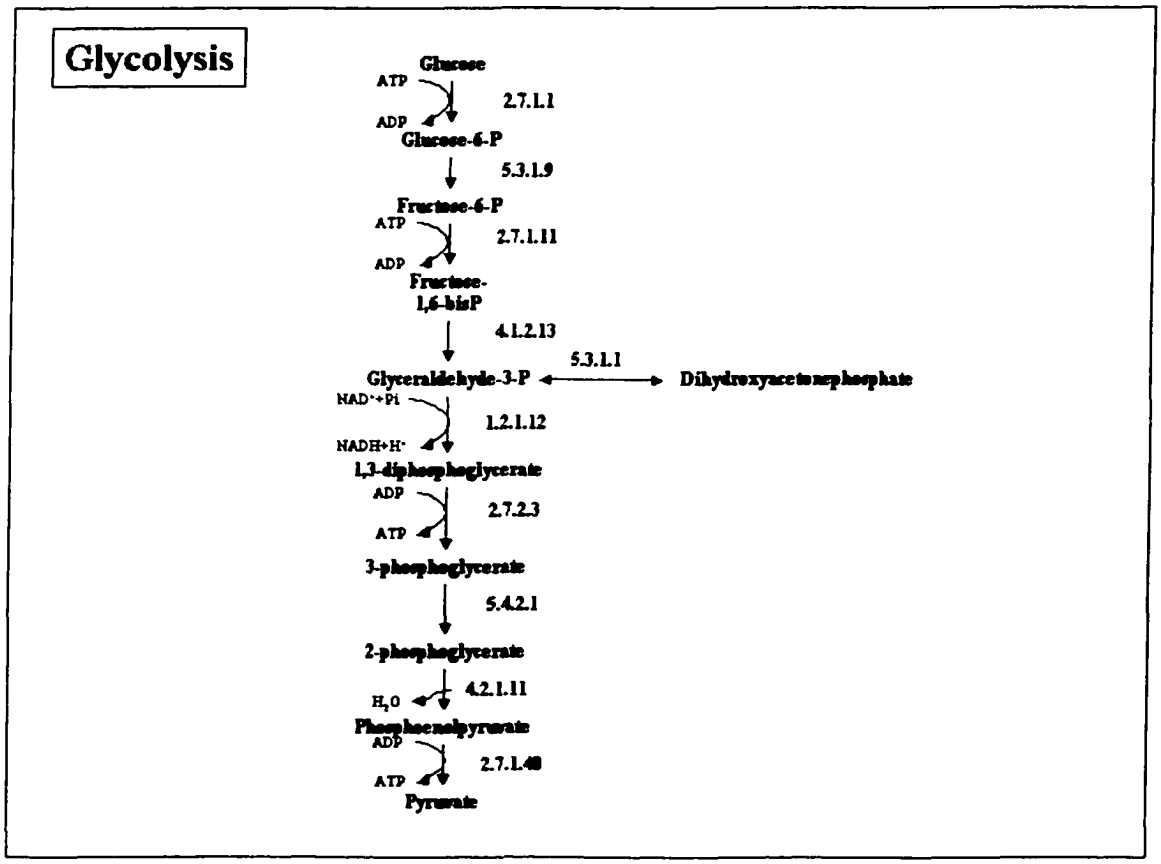


Figure 3.13. The predicted glycolytic pathway in *A. actinomycetemcomitans*

EC#	Description	Genes in *A. actinomycetemcomitans*
2.7.1.1	Hexokinase	Yes
5.3.1.9	Glucosephosphate isomerase	Yes
2.7.1.11	6-Phosphofructokinase	Yes
4.1.2.13	Fructose-biphsphate aldolase	Yes
5.3.1.1	Triosphosphate isomerase	Yes
1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase	Yes
2.7.2.3	Phosphoglycerate kinase	Yes
5.4.2.1	Phosphoglyceromutase	Yes
4.2.1.11	Enolase	Yes
2.7.1.40	Pyruvate kinase	Yes

Table 3.2. *A. actinomycetemcomitans* encoded enzymes involved in the glycolytic pathway, their EC numbers, a description, and an indication of the presence in *A. actinomycetemcomitans*.

3.3.1.2 Pentose Phosphate Pathway (PPP)

The Pentose Phosphate pathway provides pentose sugars as precursors for nucleotides and as precursors for aromatic amino acids synthesis, erythrose 4-phosphate; and NADPH. Analysis of the *A. actinomycetemcomitans* genomic sequence reveals that pentose phosphate pathway is complete. This pathway and its encoded enzymes are summarized in figure 3.14 and table3.3.

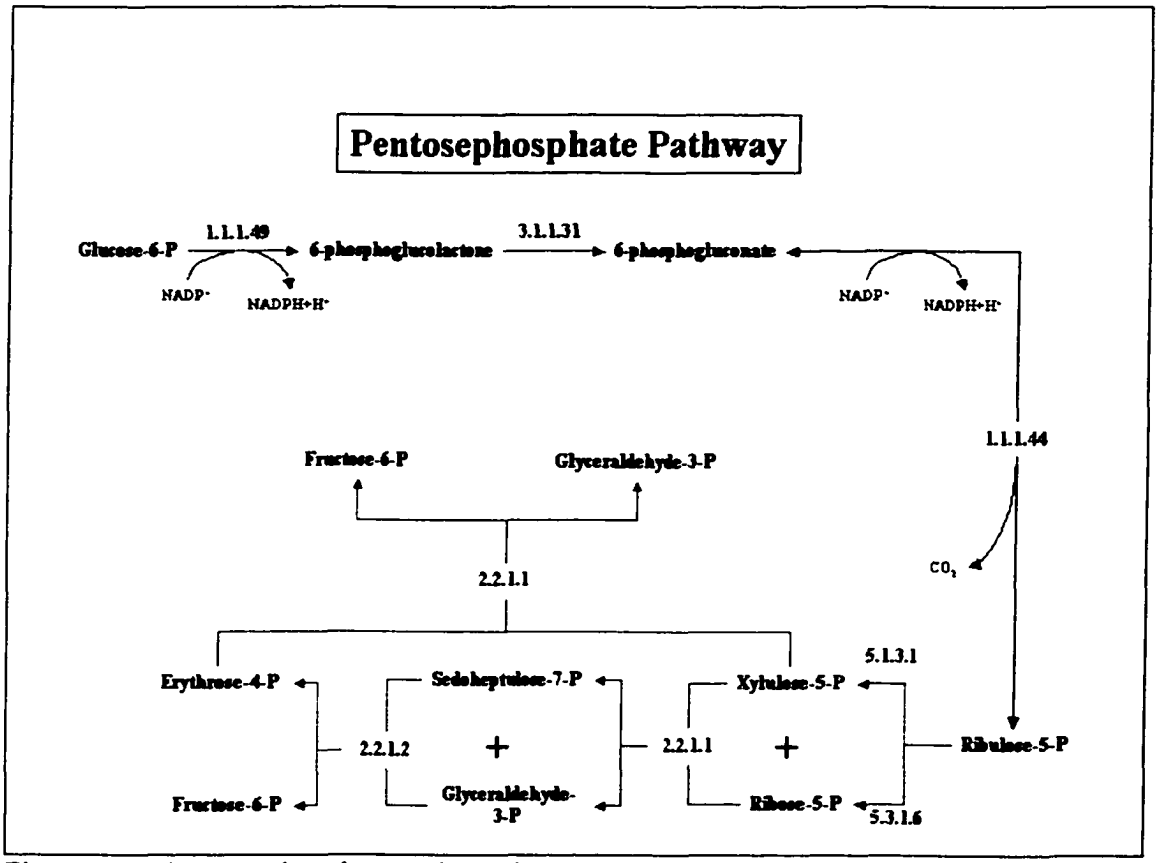


Figure 3.14. Pentosephosphate pathway in *A. actinomycetemcomitans*.

EC#	Description	Genes in *A. actinomycetemcomitans*
1.1.1.49	Glucose-6-phosphate dehydrogenase	Yes
3.1.1.31	6-Phosphogluconolactonase	Yes
1.1.1.44	Gluconate-6-phosphate dehydrogenase	Yes
5.1.3.1	Ribulose-phosphate 3-epimerase	Yes
5.3.1.6	Ribosephosphate isomerase	Yes
2.2.1.1	Transketolase	Yes
2.2.1.2	Transaldolase B	Yes

Table 3.3. Pentosephosphate pathway enzymes in *A. actinomycetemcomitans*.

3.3.1.4 Alternative pathways for pyruvate metabolism

3.3.1.4.1 Anaerobic – Pyruvate-formate lyase

The enzyme pyruvate formate-lyase (EC 2.3.1.54) oxidatively decarboxylates pyruvate to acetyl-CoA and formate. Acetyl-CoA is converted to acetyl-phosphate by phosphotransacetylase (EC 2.3.1.8) and then acetatephosphate donates its high-energy phosphate to ADP to produce ATP via acetate kinase (EC 2.7.2.1). Although discussed in more detail in the Fermentation section 3.3.1.6, *A. actinomycetemcomitans* contains the genes for the three enzymes needed for this process and listed here in table 3.4

EC#	Description	Genes in *A. actinomycetemcomitans*
2.3.1.54	Pyruvate formate lyase	Yes
2.3.1.8	Phosphotransacetylase	Yes
2.7.2.1	Acetate kinase	Yes

Table 3.4. Enzymes involved in anaerobic pyruvate metabolism in *A. actinomycetemcomitans*.

3.3.1.4.2 Aerobic – Pyruvate Dehydrogenase Reaction

Pyruvate also can be oxidized to acetyl-CoA via pyruvate dehydrogenase complex enzyme. This complex, which is about twice the size of the ribosome, contains three enzymes whose genes are present in *A. actinomycetemcomitans* and listed in Table 3.5

EC#	Description	Genes in *A. actinomycetemcomitans*
1.2.4.1	Pyruvate decarboxylase	Yes
2.3.1.12	Dihydrolipoate transacetylase	Yes
1.8.1.4	Dihydrolipoate dehydrogenase	Yes

Table 3.5. Enzymes involved in aerobic pyruvate metabolism in *A. actinomycetemcomitans*.

3.3.1.5 Tricarboxylic Acid (TCA) cycle

The TCA cycle produces reducing equivalents (NADH and $FADH_2$) for the electron transport chain and provides anabolic precursors to different amino acid synthetic pathways. The genes coding for only five of the TCA cycle enzymes were observed in *A. actinomycetemcomitans* (Table 3.6 and figure 3.15).

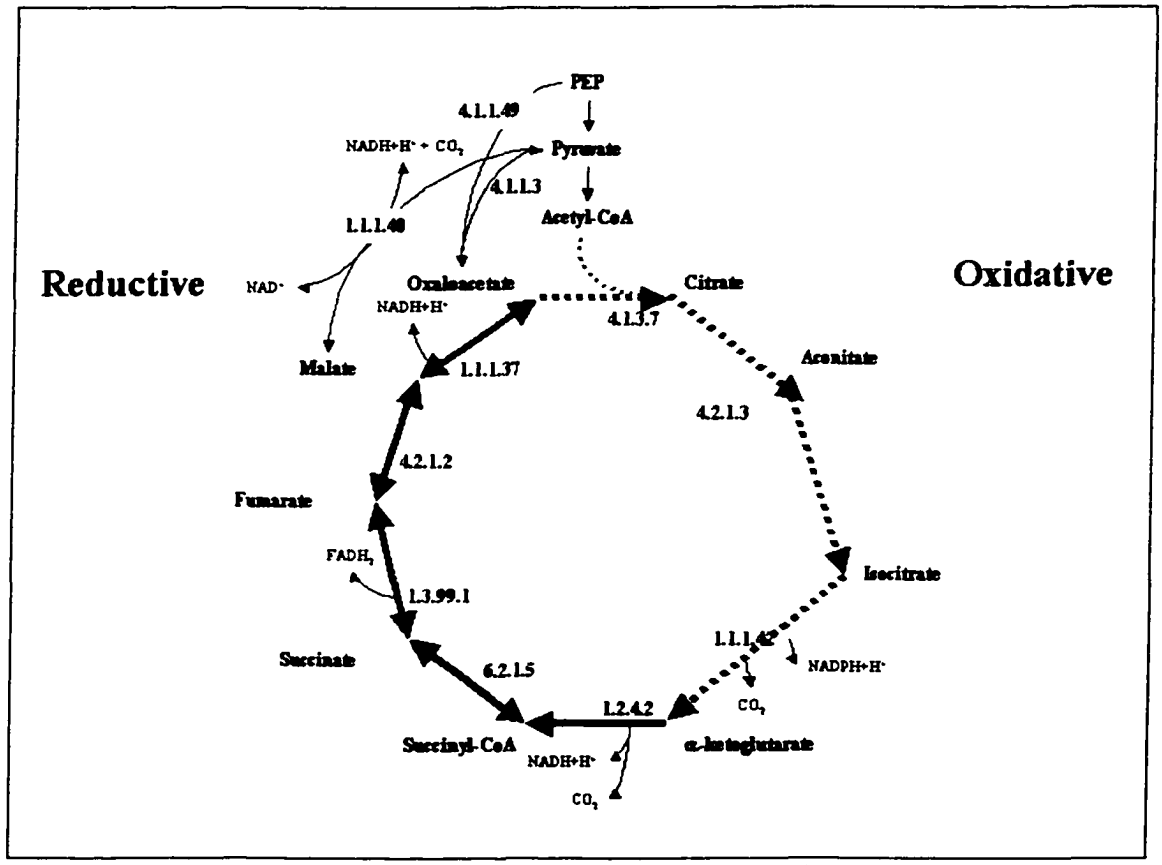


Figure 3.15. TCA cycle in *A. actinomycetemcomitans*, where the enzymes missing are indicated in dotted arrows and those present by the solid arrow.

EC#	Description	Genes in *A. actinomycetemcomitans*
4.1.3.7	Citrate Synthase	No
4.2.1.3	Aconitase	No
1.1.1.42	Isocitrate dehydrogenase	No
1.2.4.2	α-ketoglutarate dehydrogenase (2-oxoglutarate synthase)	Yes
6.2.1.5	Succinate thiokinase	Yes
1.3.99.1	Succinate dehydrogenase	Yes
4.2.1.2	Fumarase (fumarate hydratase)	Yes
1.1.1.37	Malate dehydrogenase	Yes
4.1.1.3	Oxaloacetate decarboxylase (Oxaloacetate beta-decarboxylase)	Yes
4.1.1.49	Phosphoenolpyruvate carboxykinase	Yes
1.1.1.40	Malate dehydrogenase (malic enzyme)	Yes

Table 3.6. Enzymes involved in citric acid cycle and the conversion of PEP and pyruvate to oxaloacetate and malate.

Interestingly, only a few of the sequenced bacterial genomes contain all the genes coding for the complete set of TCA enzymes. These organisms are *E. coli*, *Bacillus subtilis*, *Mycobacterium tuberculosis*, and *Rickettsia prowazekkii* (**Huynen et al., 1999**). All other bacteria sequenced to date have an incomplete TCA cycle. Therefore, it has been suggested that the TCA cycle evolved from two separate pathways branching from pyruvate (**Huynen et al., 1999**). One pathway is the oxidative pathway that oxidizes acetyl-CoA to 2-ketoglutarate, and the other pathway reduces pyruvate to succinate (**Huynen et al., 1999**). Hence, the TCA cycle can be seen as two halves, the reductive half (right half) and the oxidative half (left half) (Figure 3.15). The right or reductive half which starts with the synthesis of citrate from oxaloacetate and acetyl-CoA and ends with the oxidation of isocitrate to 2-ketoglutarate is missing in *A. actinomycetemcomitans* (Figure 3.15), and the left or oxidative half which starts with the reduction of oxaloacetate to fumarate and ends with the production of succinyl-CoA is present. This insures the production of all needed precursors for amino acid synthesis and the reoxidization of NADH back to NAD^+ (**Huynen et al., 1999**). In addition, analysis of the data reveals the presence of 2-ketoglutarate dehydrogenase enzyme that catalyzes the oxidative decarboxylation of 2-ketoglutarate to succinyl-CoA. This suggests that the 2-ketoglutarate might be imported from the environment since the data suggests that the genome encodes at least one dicarboxylate transporter.

Alternatively, oxaloacetate can be produced from pyruvate directly utilizing pyruvate carboxylase (EC 4.1.1.3), or from phosphoenol pyruvate utilizing phosphoenolpyruvate carboxykinase (EC 4.1.1.49) and malate can be oxidized to pyruvate through malate dehydrogenase (EC 1.1.1.40). Since these enzymes all are

encoded in the *A. actinomycetemcomitans* genome, it is likely that this organism can utilize malate, oxaloacetate, succinate, or 2-ketoglutarate as alternative carbon sources.

3.3.1.6 Fermentation

Fermentations are cytosolic pathways in which the NADH generated from central metabolism is reoxidized and ATP is produced through substrate-level phosphorylation. Analysis of the *A. actinomycetemcomitans* genomic sequence reveals a possible mixed-acid fermentation pathway whose end products are lactate, formate, acetate, succinate, and ethanol. This is consistent with the earlier studies which showed that *A. actinomycetemcomitans* is capable of fermenting mannose and fructose (**Ohta et al., 1989**) and the biotyping of *A. actinomycetemcomitans* which is based on its fermentation capabilities (**van Steenbergen et al., 1994**). Figure 3.16 below shows the putative mixed-acid fermentation pathway reconstructed from the sequencing data. Table 3.7 summarizes the *A. actinomycetemcomitans* enzymes involved.

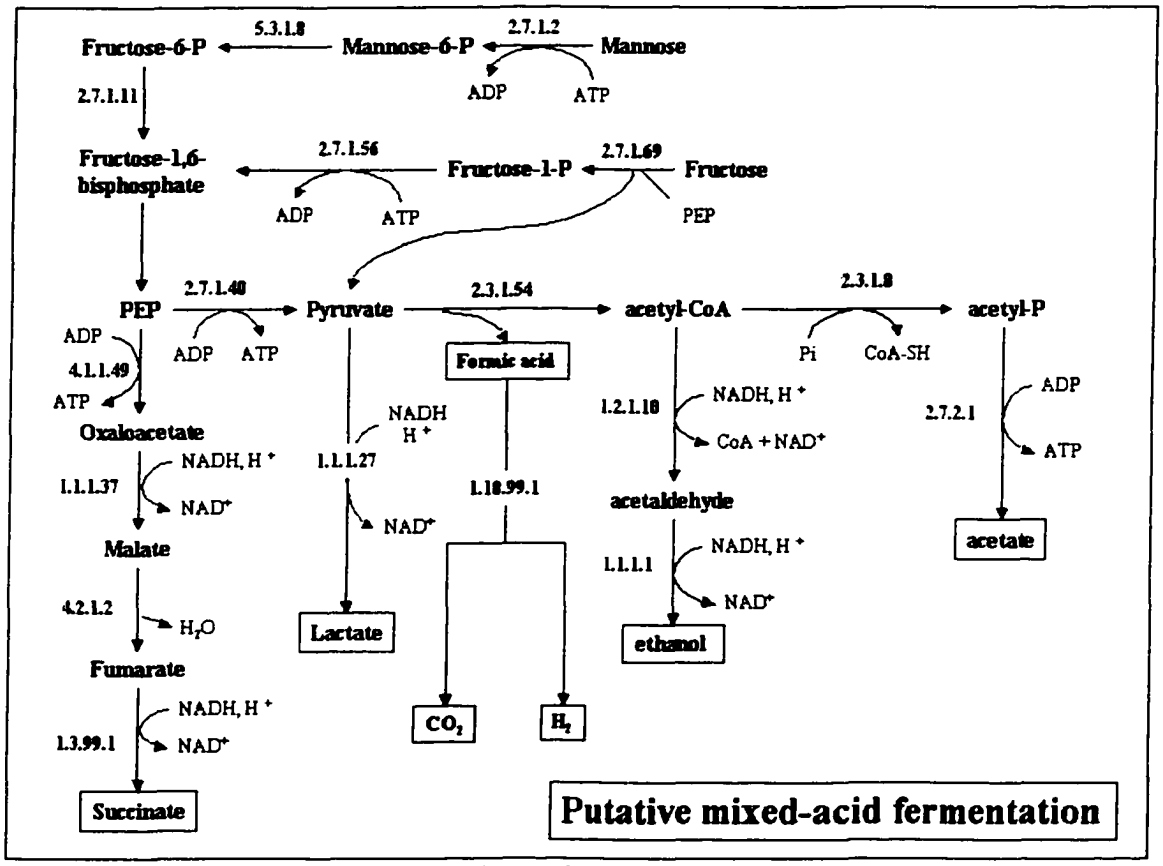


Figure 3.16. The predicted fermentative pathway for *A. actinomycetemcomitans*.

EC#	Description	Genes in *A. actinomycetemcomitans*
2.7.1.2	Glucokinase	Yes
5.3.1.8	Mannose-6-phosphate isomerase	Yes
2.7.1.11	6-Phosphofructokinase I	Yes
2.7.1.69	sugar phosphotransferase.	Yes
2.7.1.56	Fructose-1-phosphate kinase	Yes
2.3.1.8	phosphate acetyltransferase	Yes
2.3.1.54	Pyruvate formate-lyase	Yes
2.7.1.40	Pyruvate kinase	Yes
2.7.2.1	Acetate kinase	Yes
1.2.1.10	Acetaldehyde dehydrogenase	Yes
1.1.1.1	Alcohol dehydrogenase	Yes
1.18.99.1	Formate hydrogenlyase	Yes
4.1.1.49	Phosphoenolpyruvate carboxykinase	Yes
1.1.1.37	Malate dehydrogenase	Yes
4.2.1.2	Fumarase	Yes
1.1.1.27	Lactate dehydrogenase	Yes
1.3.99.1	Fumarate reductase	Yes

Table 3.7. Enzymes encoded in *A. actinomycetemcomitans* involved in fermentation.

3.3.1.7 Respiration and Electron Transport

Generally, in bacteria the electron transport chain (ETC) in which electrons flow from a primary electron donor to a terminal electron acceptor through a series of electron-carrier proteins, includes components equivalent to those of the eukaryotic mitochondrial electron transport chain. The flow of the electrons through the bacterial cytoplasmic membrane to generate potential is referred to as respiration. If the terminal acceptor is oxygen, then the respiration is said to be aerobic (see Gennis et al., 1996 for a review of electron transport). If, on the other hand, the terminal acceptor is not oxygen (e.g. nitrate or sulfate), then the respiration is anaerobic. Several classes of electron carriers are involved in the electron transport chain such as flavoproteins, iron-sulfur proteins (FeS), and cytochromes. All the genes for the proteins in bacterial respiration, i.e. the electron transport chain, are present in the *A. actinomycetemcomitans* genome.

In *E. coli*, the respiration system is highly modular and the scheme of the electron flow is similar between different modules. Basically, the cell needs a substrate-specific dehydrogenase as the electron carrier (in the form of quinone derivative), and a terminal oxidoreductase that reduces the final acceptor. Several putative respiration pathways can be constructed for *A. actinomycetemcomitans* (Figure 3.17) based on the genomic sequence of possible donor and acceptor encoding genes. The presence of formate dehydrogenase-N suggests that formate oxidation to carbon dioxide can be coupled with nitrate reduction forming an anaerobic respiratory chain. In *E. coli*, this protein is expressed mostly under anaerobic conditions and is a proton pump. Glycerol 3-phosphate, a product from phospholipid degradation, can donate electrons through glycerol-3-phosphate dehydrogenase, the electron acceptor can be nitrate. Lactate,

another product from the anaerobic cleavage of pyruvate, donates electrons through lactate dehydrogenase to quinone. NADH generated from different metabolic pathways donates its electrons to NADH dehydrogenase which reduces quinone that passes the electron to succinate:ubiquinone oxidoreductase that, in turn passes the electron to the terminal electron acceptor.

As illustrated in figure 3.17, the data suggests that *A. actinomycetemcomitans* likely utilizes nitrite, DMSO, fumarate, and oxygen as terminal electron acceptors, since the genes for all the enzymes given in figure 3.17 are encoded in the *A. actinomycetemcomitans* genome.

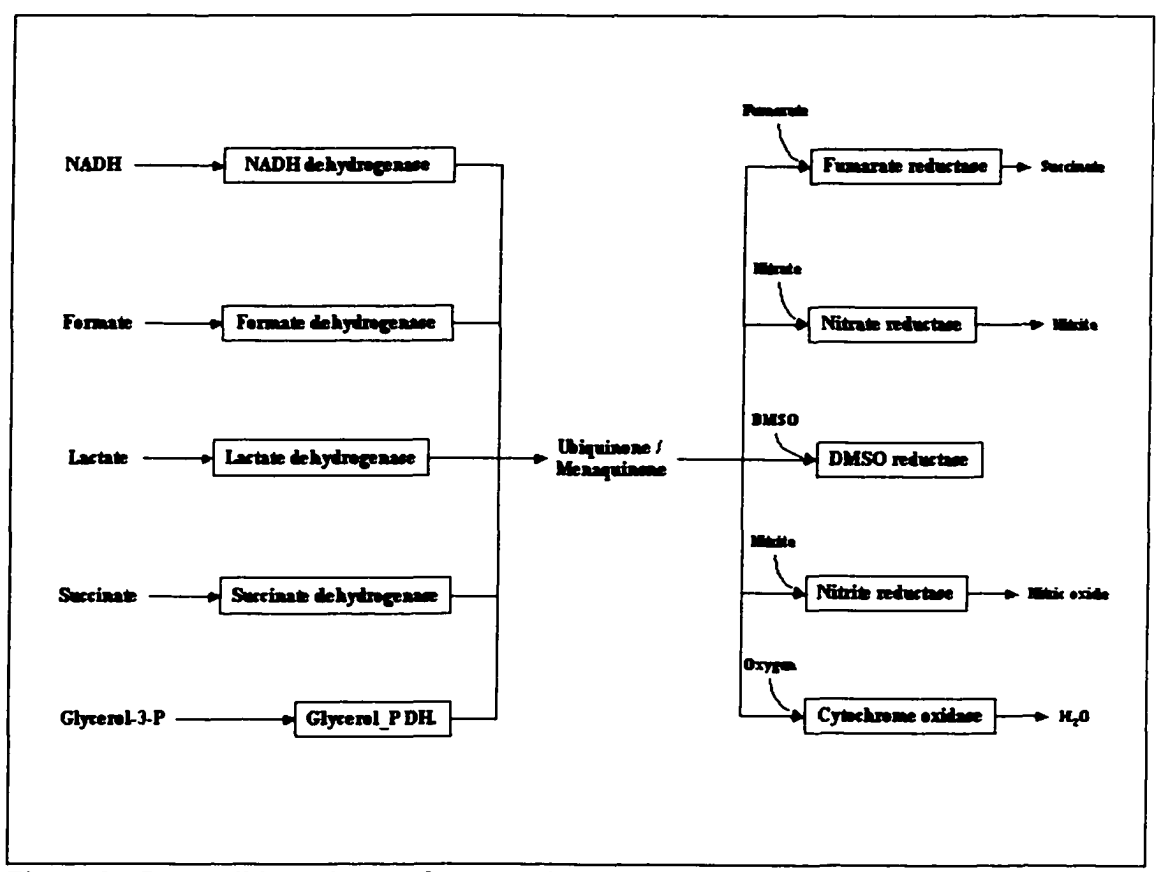


Figure 3.17. Possible pathways for anaerobic and aerobic respiration in *A. actinomycetemcomitans* as reconstructed from the genes encoded in this bacterium.

3.3.2 Metabolism of Lipids

Because of their large hydrophobic regions, fatty acids rarely are free molecules inside the cell, Instead, they usually are esterified to glycerol (phosphoglycerides), to carbohydrate moieties such as glucosamine in the lipid A portion of lipopolysaccharides (LPS), or to proteins bound to the peptidoglycan (See White, 2000 for an informative review). They also are an intergral part of the composition of membranes. In this following section, the synthesis of fatty acids, and phospholipids in *A. actinomycetemcomitans* will be discussed.

3.3.2.1 Fatty acids synthesis

Although the preliminary analysis of the *A. actinomycetemcomitans* genome indicates that it can not metabolize fatty acids through the β-oxidation pathway, *A. actinomycetemcomitans* does seem to contain the genes which encodes all the enzymes needed for fatty acid synthesis. Figure 3.18 illustrates the fatty acid biosynthesis pathway and the enzyme description is provided in table 3.8. The absence of the genes for β-oxidation enzymes in *A. actinomycetemcomitans* suggests that it can not utilize fatty acids as a carbon source. The absence of this pathway interestingly also has been reported in *H. influenzae* (**Sutton et al., 1995, and KEGG**), a close relative of *A. actinomycetemcomitans*.

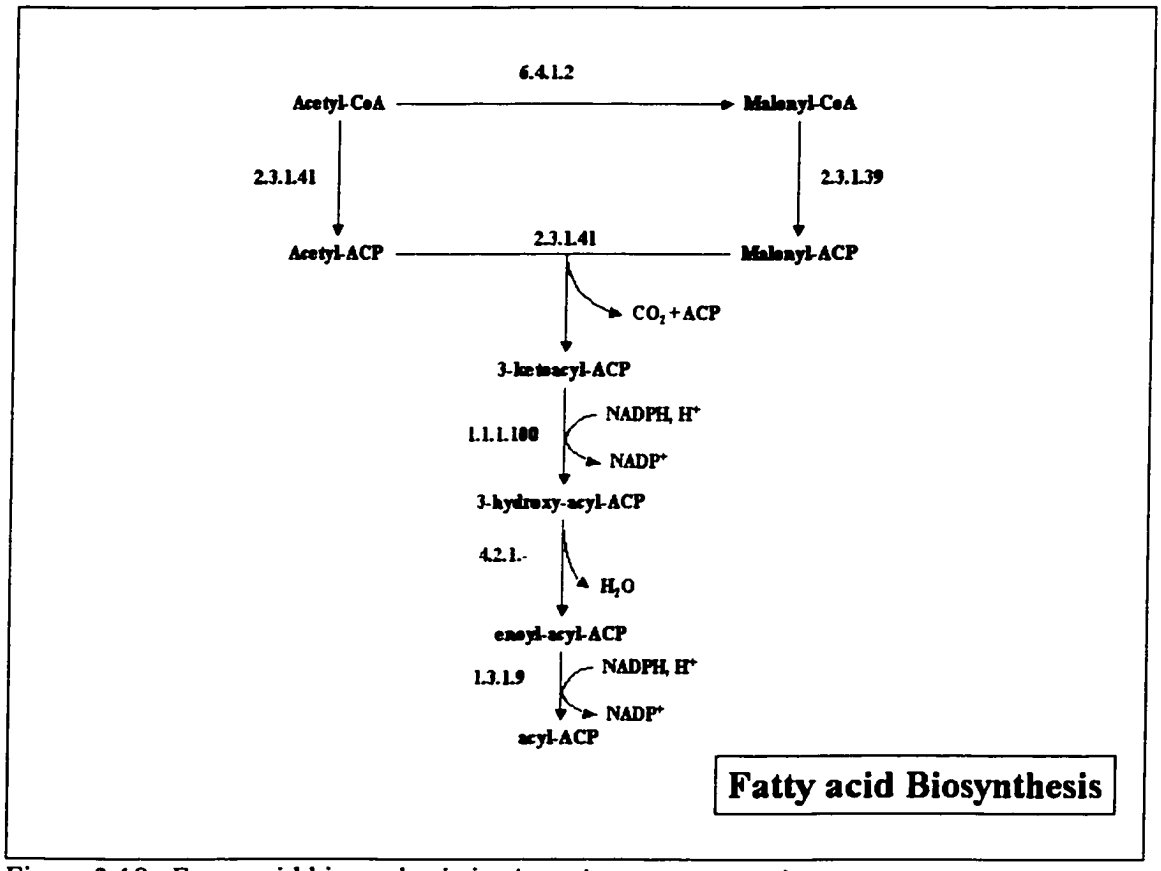


Figure 3.18. Fatty acid biosynthesis in *A. actinomycetemcomitans*.

EC#/Gene name	Description	Genes in *A. actinomycetemcomitans*
6.4.1.2	acetyl-CoA carboxylase carboxyl transferase subunit alpha(accA)	Yes
6.4.1.2	acetyl-CoA carboxylase biotin carboxyl carrier protein (accB, fabE)	Yes
6.3.4.14	biotin carboxylase (A subunit of acetyl-CoA carboxylase) (accC)	Yes
6.4.1.2	acetyl-CoA carboxylase (EC 6.4.1.2) carboxyltransferase beta chain (accD)	Yes
4.2.1.-	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase (fabZ, sefA)	Yes
4.2.1.60	3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase (beta-hydroxydecanoyl thioester dehydrase) (fabA)	Yes
2.3.1.41	3-oxoacyl-[acyl-carrier-protein] synthase (fabB)	Yes
2.3.1.39	malonyl CoA-acyl carrier protein transacylase (MCT) (fabD)	Yes
2.3.1.41	3-oxoacyl-[acyl-carrier-protein] synthase II (fabF)	Yes
1.1.1.100	3-oxoacyl-[acyl-carrier protein] reductase (3-ketoacyl-acyl carrier protein reductase) (fabG)	Yes
2.3.1.41	3-oxoacyl-[acyl-carrier-protein] synthase III (fabH)	Yes
1.3.1.9	enoyl-(acyl-carrier-protein) reductase (NADH) (fabI)	Yes
LktC	Acyl carrier protein for processing of prohemolysin (Toxin activation protein C)	Yes
3.1.2.-	acyl-CoA thioesterase II (tesB)	Yes

Table 3.8. Enzymes of the proposed fatty acid biosynthesis pathway present in *A. actinomycetemcomitans*.

3.3.2.2 Synthesis of phospholipids

Phospholipids, important constituent of the cell membrane, are fatty acids esterified to glycerol and other molecules such as amino acids, or ethanol amine. Figure 3.19 shows the pathway responsible for the biosynthesis of the major phospholipids. Analysis of the *A. actinomycetemcomitans* genomic sequence reveals that all of the genes are present that could encode the enzymes associated with phospholipid biosynthesis as shown in figure 3.19 and listed in table 3.9 Additionally, two lipase genes which encode proteins with broad specificity for phospholipid degradation also are observed in the *A. actinomycetemcomitans* genome.

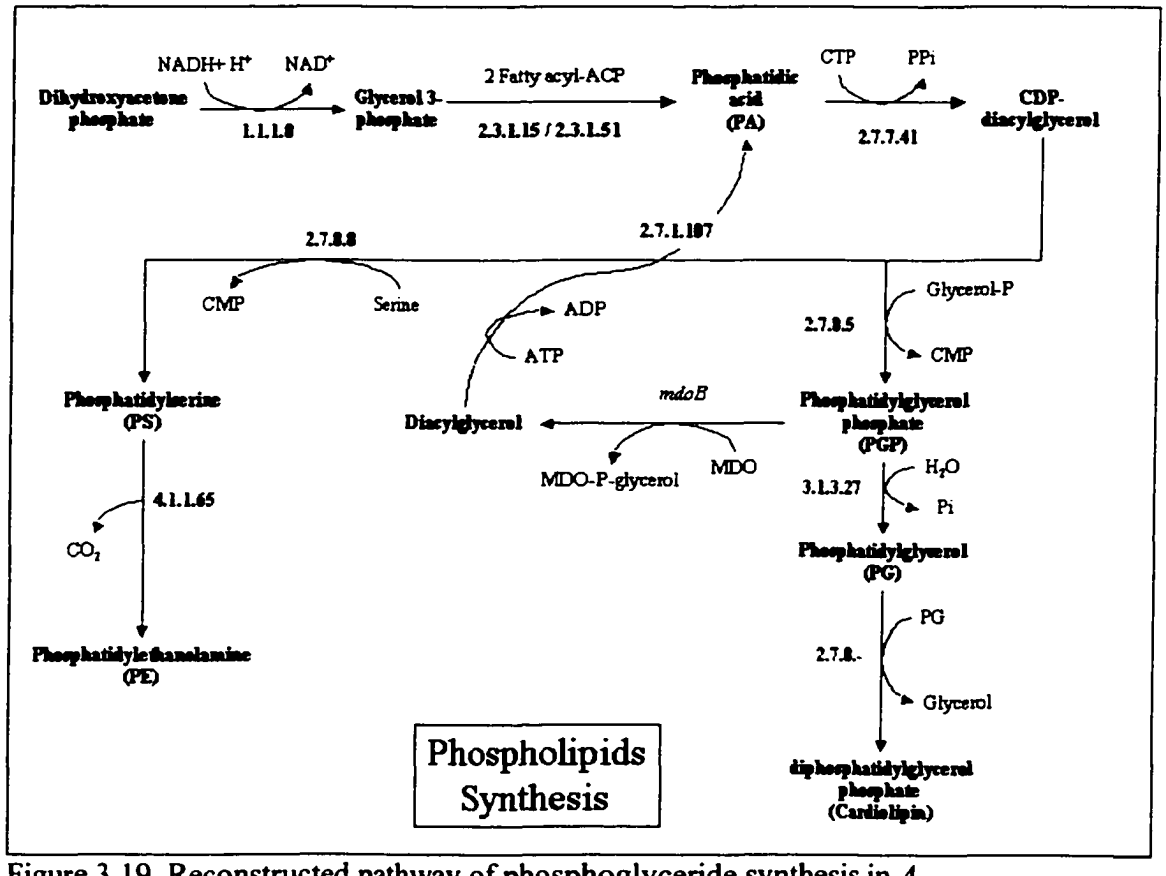


Figure 3.19. Reconstructed pathway of phosphoglyceride synthesis in *A. actinomycetemcomitans*.

EC#/Gene name	Description	Genes in *A. actinomycetemcomitans*
1.1.1.8	Glycerol-3-phosphate dehydrogenase	Yes
2.3.1.15	Glycerol-3-phosphate acyltransferase (plsB)	Yes
2.3.1.51	1-acyl-sn-glycerol-3-phosphate acyltransferase	Yes
2.7.7.41	phosphatidate cytidylyltransferase (CDP-diglyceride synthetase) (cdsA)	Yes
2.7.8.5	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase (pgsA)	Yes
3.1.3.27	Phosphatidylglycerophosphatase A, membrane bound (pgpA)	Yes
2.7.8.-	Cardiolipin synthetase (cls)	Yes
2.7.8.8	phosphatidylserine synthase (pssA)	Yes
4.1.1.65	Phosphatidylserine decarboxylase proenzyme	Yes
mdoB	Membrane-derives oligosaccharides; Phosphoglycerol transferase I activity	yes
2.7.1.107	diacylglycerol kinase (dgkA)	Yes

Table 3.9. Enzymes in *A. actinomycetemcomitans* phospholipid biosynthesis.

3.3.3 Amino acid Biosynthesis

The three central pathways (glycolysis, TCA cycle and pentose phosphate pathway) provide the precursors needed to synthesize all 20 amino acids. However, *A. actinomycetemcomitans* does not seem to be capable of synthesizing all 20 amino acids as it lacks the genes for the enzymes required for the biosynthesis of leucine, isoleucine, valine, methionine, lysine, arginine, and histidine (Table 3.10). Considering the environment in which it lives, it is not surprising that *A. actinomycetemcomitans* does not need to synthesize all the amino acids. Nutrients in the oral cavity usually are abundant. Table 3.10 below shows the precursor of each of the amino acids and the presence or absence of each biosynthetic pathway from the genome sequence. Since *A. actinomycetemcomitans* possess a variety of transport systems that are specialized for uptake of amino acids as discussed in the ABC transport section, the presence of these transporters may compensate for the lack of some of the biosynthetic pathways.

Additionally, *A. actinomycetemcomitans* possesses an oligopeptide uptake system, an *opp* homologue, that imports short peptides as a source for amino acids (**Naider et al., 1975**).

Precursor	Amino Acid	Complete biosynthetic pathway present in *A. actinomycetemcomitans*
Pyruvate	Ala	Yes
	Val	No
	Leu	No
Oxaloacetate	Asp	Yes
	Asn	Yes
	Met	Yes
	Lys	No
	Thr	Yes
	Ile	No
α-ketoglutarate	Glu	Yes
	Gln	Yes
	Arg	No
	Pro	Yes
3-phosphoglyceraldehyde	Ser	Yes
	Gly	Yes
	Cys	Yes
Phosphoenolpyruvate & erythrose 4-phosphate	Phe	Yes
	Tyr	Yes
	Trp	Yes
5-Phosphoribosyl-1-pyrophosphate	His	No

Table 3.10. Amino acid precursor and the genomic encoded biosynthesis capabilities in *A. actinomycetemcomitans*.

3.3.3.1 Glutamate and glutamine

Glutamate synthesis is one of the extremely important reactions in bacteria since it is important for incorporating inorganic nitrogen into the different cellular processes (**White, 2000**). This reaction is part of the minimal sets of genes speculated to be sufficient to support bacterial growth (**Mushegian, 1996**) as illustrated below in figure 3.20. In *A. actinomycetemcomitans*, *H. influenzae*, *E. coli*, and *Helicobacter pylori*, glutamate synthesis is catalyzed by glutamate dehydrogenase (**White, 2000**), while glutamine is synthesized via glutamine synthetase (Figure 3.20).

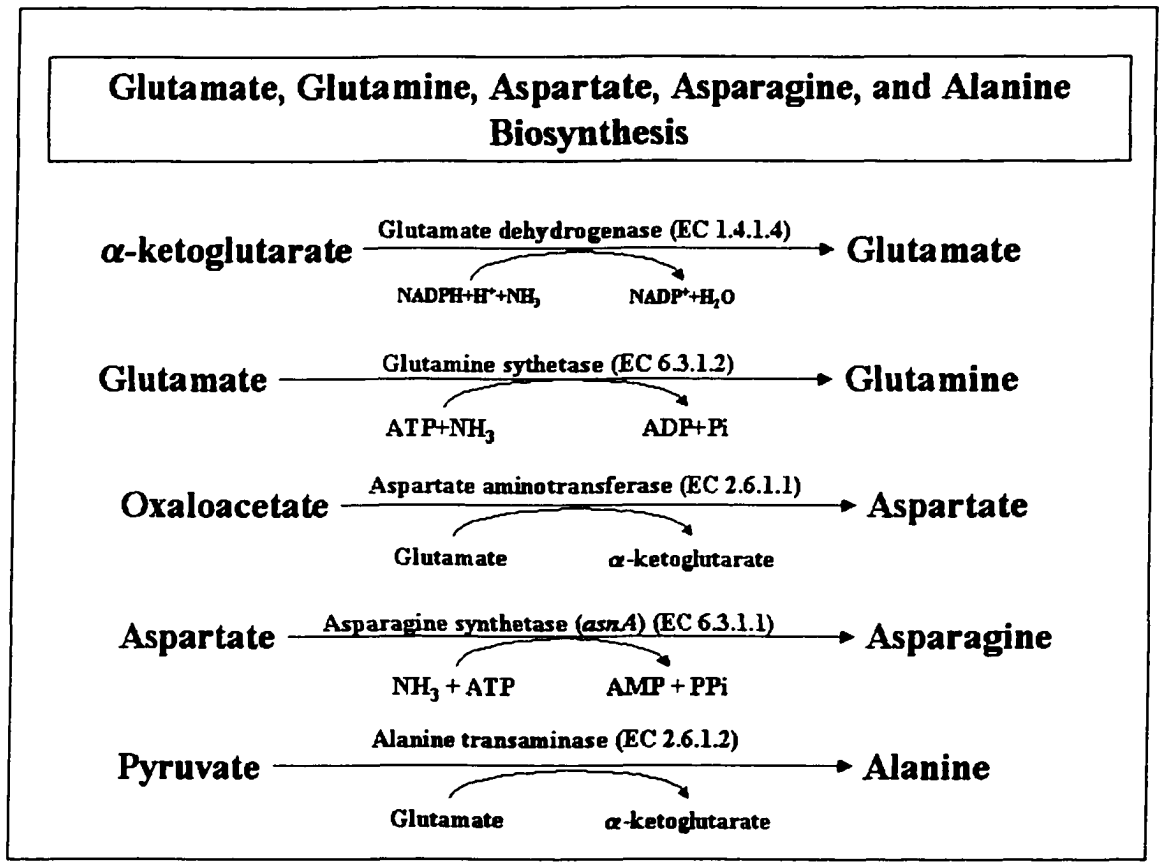


Figure 3.20. Synthesis of glutamate, glutamine, aspartate, asparagine, and alanine.

3.3.3.2 Aspartate, Asparagine and Alanine

As shown in figure 3.20, of the *A. actinomycetemcomitans* genomic sequence contains the genes needed for the synthesis of aspartate and alanine by transamination of their carboxylic acid counterparts via aspartate transaminase (EC 2.6.1.1) and alanine transaminase (EC 2.6.1.2) respectively **(White, 2000)**.

3.3.3.3 Branched-Chain Amino Acids

Analysis of *A. actinomycetemcomitans* genomic sequence data reveals that *A. actinomycetemcomitans* can not synthesize any of the branched-chain amino acids

because the genes for these enzymes are not present in *A. actinomycetemcomitans* (Figure 3.21 and table 3.11). It is interesting to note that for valine and leucine, the missing enzyme is a dihydroxy acid dehydratase. In an effort to locate a homologue for this enzyme, a group of orthologus dihydroxy acid dehydratase genes were downloaded from GeneBank and blasted aginst *A. actinomycetemcomitans* sequence. No homologue was detected using this approach.

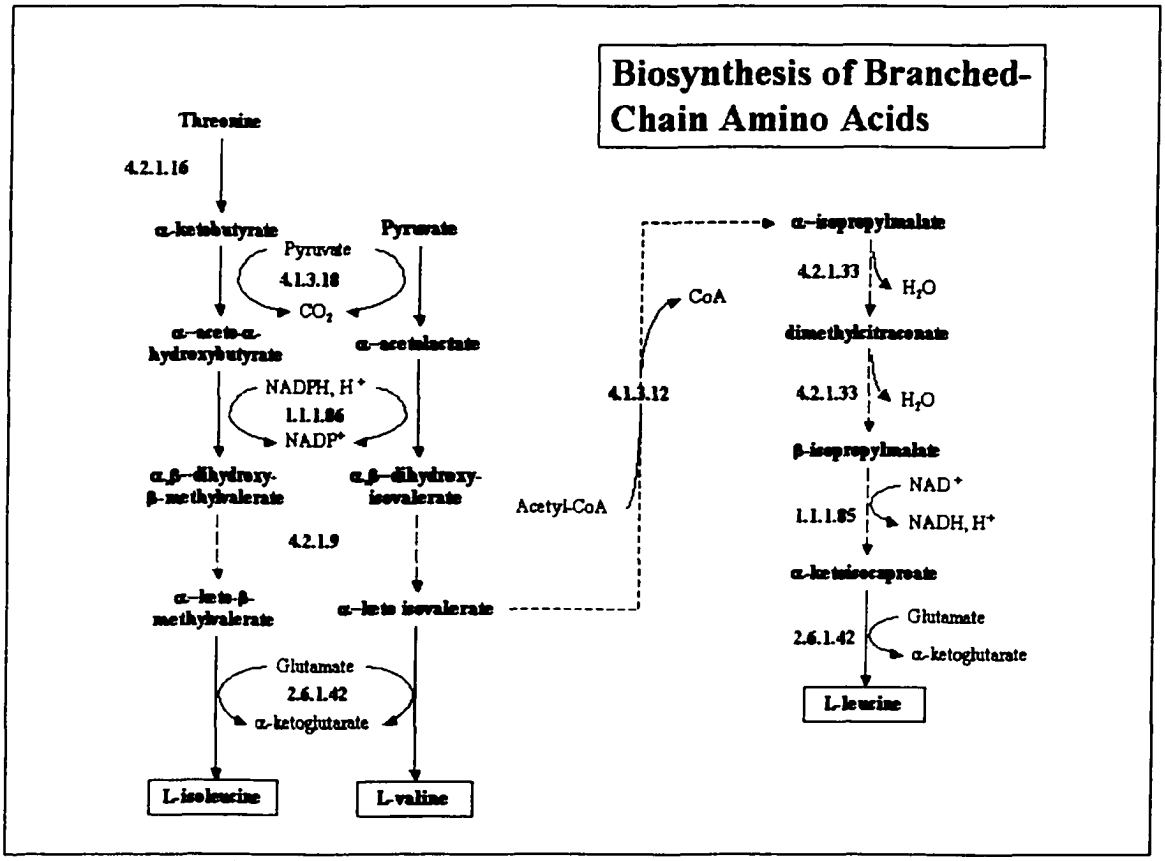


Figure 3.21. Synthesis of branched-chain amino acids. Dotted arrows represent reactions catalyzed by enzymes not found in *A. ctinomycetemcomitans* sequencing data.

EC#	Description	Genes in *A. actinomycetemcomitans*
4.2.1.16	Threonine dehydratase	Yes
4.1.3.18	Acetolactate synthase	Yes
1.1.1.86	ketol-acid reductoisomerase	Yes
4.2.1.9	Dihydroxy-acid dehydratase.	No
2.6.1.42	Branched-chain amino acid aminotransferase	Yes
4.1.3.12	2-isopropylmalate synthase (leuA)	No
4.2.1.33	3-isopropylmalate dehydratase (leuC)	No
1.1.1.85	3-isopropylmalate dehydrogenase (leuB)	No
2.6.1.42	Branched-chain amino acid aminotransferase	Yes

Table 3.11. Enzymes involved in the biosynthesis of branched-chain amino acids.

Interestingly, the gene for the branched-chain transporter, BrnQ, homologue is present in the *A. actinomycetemcomitans* genome suggesting that *A. actinomycetemcomitans* can take up these amino acids from the environment. BrnQ uptake of branched-chain amino acids seems to be energized by a proton motif force (**Umbarger, 1996**). It is interesting that one enzyme from the valine and isoleucine biosynthetic pathways is missing while all other enzymes are present. It is always possible that a substitute dehydratase enzyme can perform this reaction.

3.3.3.4 Lysine, Methionine, and Threonine

Analysis of the *A. actinomycetemcomitans* genomic sequence reveals that the genes encoding the enzymes required for the biosynthesis of threonine all are present, while the genes for the enzymes required for biosynthesis of methionine all are missing. However, as the gap closure of the *A. actinomycetemcomitans* progressed, the enzyme cystathionine gamma-synthase was revealed. Using KEGG alternative pathways search, a putative pathway to the synthesis of methionine was revealed from homoserine utilizing homoserine O-acetyltransferase to generate acetyl-homoserine that then is converted to homocystein. Homocystein then is converted to metionine using another enzyme called

Tetrahydropteroyltriglutamate methyltransferase. The genes encoding all of the enzymes needed for lysine biosynthesis except that for succinyldiaminopimelate transaminase (EC 2.6.1.17) (Figure 3.22 and table 3.12) are present, similar to that observed for *H. influenzae* (**Fleischmann et al., 1995**). Thus, it is likely that both *A. actinomycetemcomitans* and *H. influenzae* rely on specific transporters to import lysine. On the other hand, as listed in appendix-A, there are unassigned transaminases that were detected using motif analysis that may fill in the missing function in the lysine biosynthetic pathway. As will be discussed below, the *A. acinomycetemcomitans* genome encodes many different amino acid transporters that likely are responsible for importing lysine and methionine as well as other metabolites from the environment.

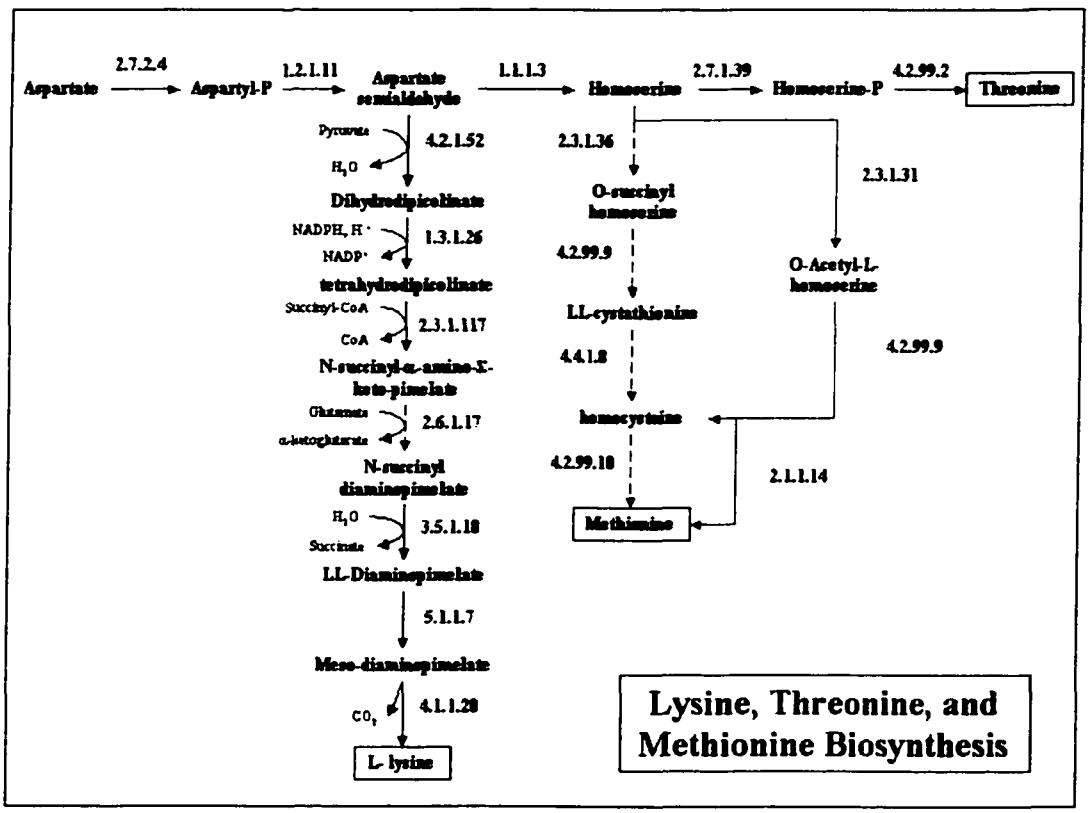


Figure 3.22. Synthesis of lysine, threonine, and methionine. Dotted arows represent reactions carried out by enzymes that were not found in *A. actinomycetecomitans*. Solid arrowsrepresent enzymes that were found in the sequencing data.

EC#	Description	Genes in *A. actinomycetemcomitans*
2.7.2.4	Aspartokinase I-homoserine dehydrogenase	Yes
1.2.1.11	aspartate semialdehydedehydrogenase	Yes
1.1.1.3	Homoserine dehydrogenas	Yes
2.7.1.39	Homoserine kinase (thrB)	Yes
4.2.1.52	Dihydrodipicolinate synthase (dapA)	Yes
1.3.1.26	Dihydrodipicolinate reductase (dapB)	Yes
2.3.1.117	Tetrahydrapicolinate N-succinyltransferase	Yes
2.6.1.17	Succinyldiaminopimelate transaminase	No
3.5.1.18	Succinyl-diaminopimelate desuccinylase	Yes
5.1.1.7	Diaminopimelate epimerase (dapF)	Yes
4.1.1.20	Diaminopimelate decarboxylase (lysA)	Yes
2.3.1.36	D-Amino-acid N-acetyltransferase	No
4.2.99.9	Cystathionine gamma-synthase (metB)	No
4.4.1.8	Cystathionine beta-lyase (metC)	No
2.3.1.31	Homoserine O-acetyltransferase	Yes
2.1.1.14	Tetrahydropteroyltriglutamate methyltransferase	Yes
4.2.99.9	Cystathionine gamma-synthase	Yes
4.2.99.10	O-Acetylhomoserine (thiol)-lyase	No

Table 3.12. Enzymes involved in the biosynthesis of lysine, threonine, and methionine.

3.3.3.5 Serine, glycine, and cycteine

As illustrated in figure 3.23 and table 3.13, *A. actinomycetemcomitans* appears to encode the enzymes for the complete biosynthetic pathways for serine, glycine and cysteine. It is interesting that the last reaction for cycteine biosynthesis pathway (catalyzed by 4.2.99.2: Threonine synthase) utilizes hydrogen sulfide which the analysis of the *A. actinomycetemcomitans* data suggests that it can not be produced through the sulfate reduction pathway found in *E. coli*. Even though a putative sulfate transporter was revealed from the analysis, it is still unclear how it is reduced to hydrogen sulfide. It is possible that a different pathway might exist to utilize sulfate or that *A. actinomycetemcomitans* rely only on external cysteine through the different amino acid and peptide transport systems that the analysis of the genome revealed. This observation was found to be true in *H. influenzae* as well.

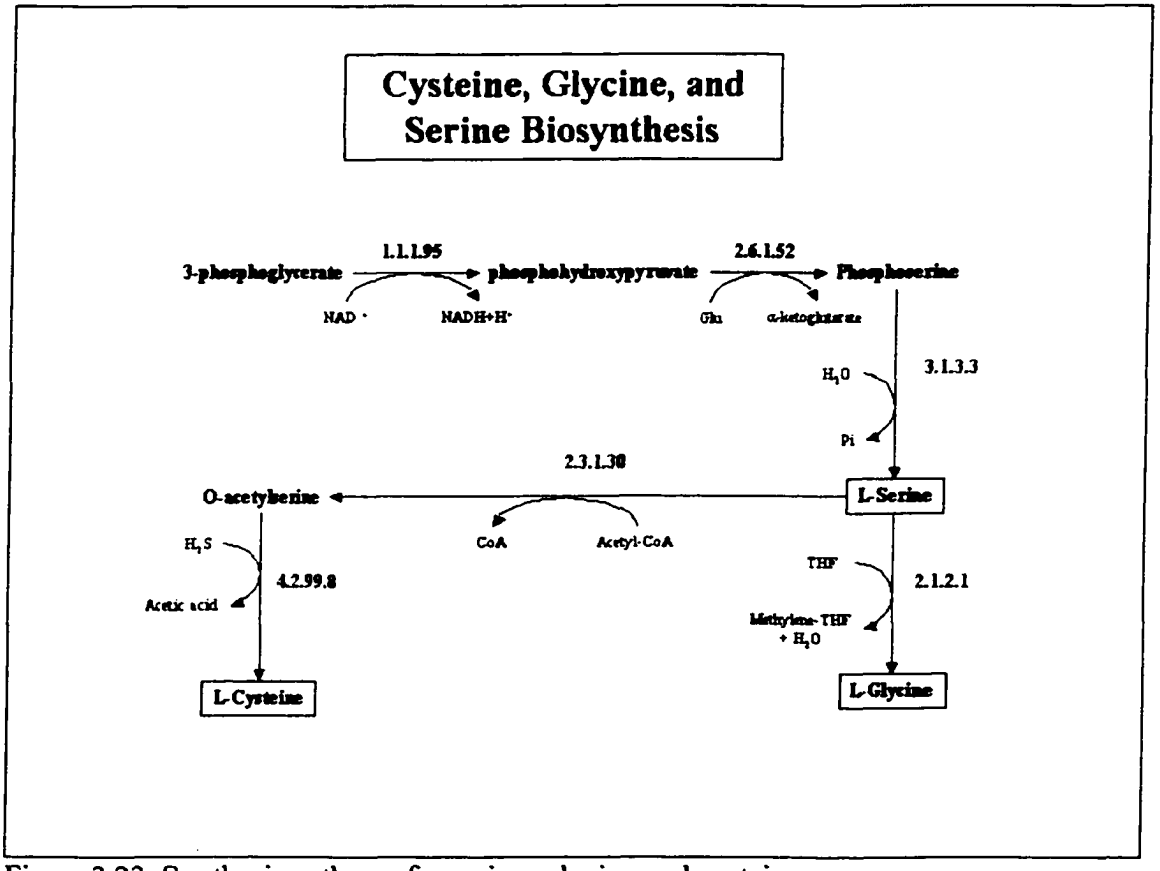


Figure 3.23. Synthesis pathway for serine, glycine and cysteine.

EC#	Description	Genes in *A. actinomycetemcomitans*
1.1.1.95	Phosphoglycerate dehydrogenase	Yes
2.6.1.52	3-Phosphoserine aminotransferase(serC)	Yes
3.1.3.3	3-Phosphoserine phosphatase	Yes
2.3.1.30	Serine tranacetylase	Yes
4.2.99.8	O-acetylserine sulfhydrolase	Yes
2.1.2.1	Serine hydroxymethyltransferase	Yes

Table 3.13. *A. actinomycetemcomitans* enzymes involved in the synthesis of serine, glycine and cysteine.

Surprisingly, the *A. actinomycetemcomitans* genome also encodes for the serine-specific transporter SdaC (**Shao et al., 1994**) even though the enzymes for serine biosynthesis from phosphohydroxypyruvate are present.

3.3.3.6 Aromatic amino acids

A. actinomycetemcomitans is capable of synthesizing all three aromatic amino acids since all of the genes needed for the enzymes involved in the aromatic amino acid biosynthetic pathways are present Figure 3.24 and table 3.14).

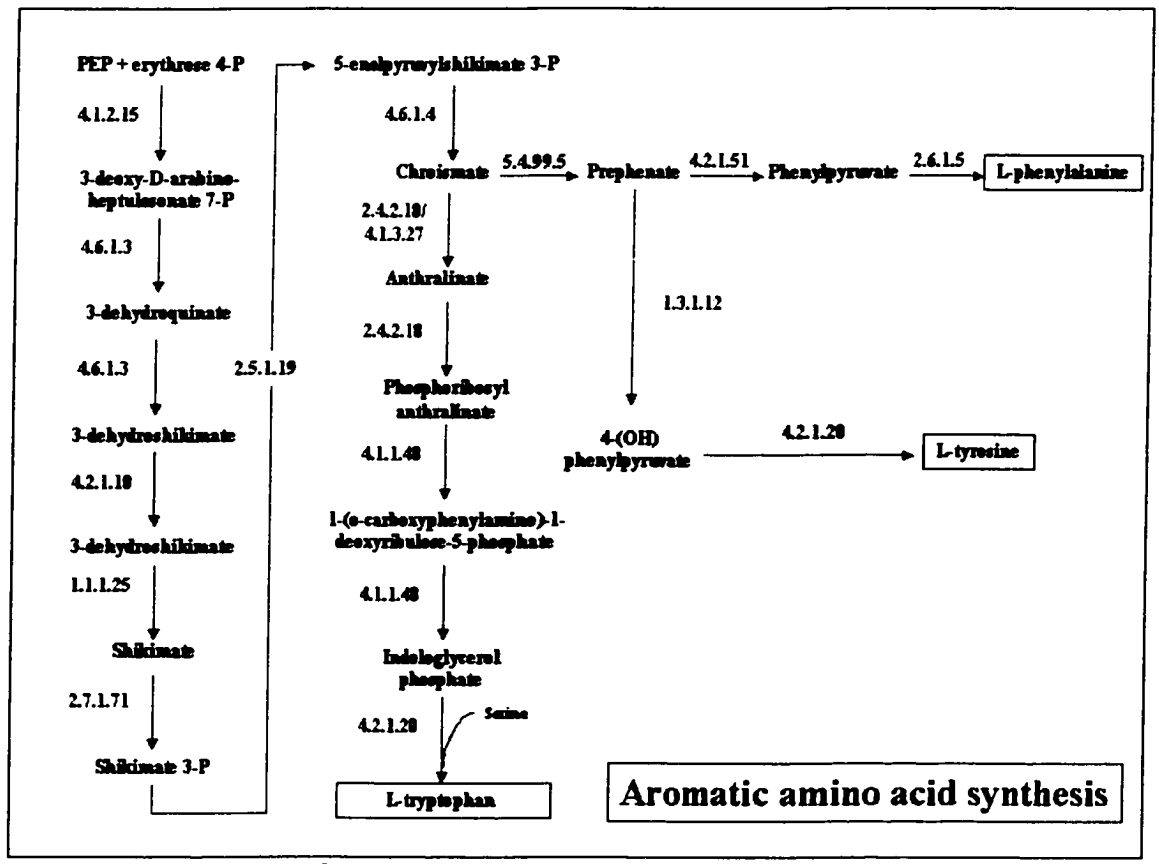


Figure 3.24. Synthesis of aromatic amino acids.

EC#	Description	Genes in *A. actinomycetemcomitans*
2.5.1.19	5-Enolpyruvylshikimate-3-phosphate synthetase (3-phosphoshikimate 1-carboxyvinyltransferase) (EPSP synthase)(aroA)	Yes
4.6.1.3	Dehydroquinate synthase(aroB)	Yes
4.6.1.4	Chorismate synthase (5-enolpyruvylshikimate-3-phosphate phospholyase)(aroC)	Yes
4.2.1.10	5-Dehydroquinate dehydratase (3-dehydroquinase) (aroD)	Yes
1.1.1.25	Dehydroshikimate reductase, shikimate 5-dehydrogenase(aroE)	Yes
4.1.2.15	Phospho-2-dehydro-3-deoxyheptonate aldolase (DAHP synthetase, phenylalanine repressible)(aroG)	Yes
2.7.1.71	Shikimate kinase I (aroK)	Yes
4.2.1.51	Chorismate mutase	Yes
5.4.99.5	Prephenate dehydratase	Yes
1.3.1.12	Chorismate mutase T and prephenate dehydrogenase	Yes
4.2.1.20	Tryptophan synthase	Yes
4.1.1.48	N-(5-Phosphoribosyl)anthranilate isomerase and indole-3-glycerolphosphate synthetase(trpC)	Yes
2.4.2.18	Anthranilate phosphoribosyltransferase (Phosphoribosyl-anthranilate pyrophosphorylase)	Yes
2.6.1.5	Tyrosine aminotransferase, tyrosine repressible	Yes
2.4.2.18/ 4.1.3.27	Anthranilate phosphoribosyltransferase/ Glutamine amidotransferase	Yes

Table 3.14. Enzymes involved in aromatic amino acid biosynthesis as found from *A. actinomycetemcomitans* genome.

3.3.3.7 Arginine and proline

Analysis of the sequence of the *A. actinomycetemcomitans* genome indicates that it contains the genes encoding for the proteins which synthesize proline. However, as seen in figure 3.25 and table 3.15, several of the genes encoding for arginine biosynthesis enzymes are missing from the genome. However, the *A. actinomycetemcomitans* genome encodes for the arginine transport system, *art* (**Wissenbach et al., 1995**). This system is composed of four genes (*artPIQM*) where ArtI is a perimplasmic protein that binds arginine with high affinity (**Wissenbach et al., 1995**). ArtM, Q, and P are permeases that seem to facilitate arginine entry into the cell and ArtI also binds other basic amino acids with a lower affinity (**Wissenbach et al., 1995**). In addition, analysis of the A. actinomycetemcomitans data revealed another putative pathway for arginine synthesis

which requires the reduction og glutamate to glutamate semialdehyde via 1-Pyrroline-5-carboxylate dehydrogenase and then transaminating the semialdehyde to ornithine (possibly by any transaminase) which then follows the path as seen ion figure 3.25.

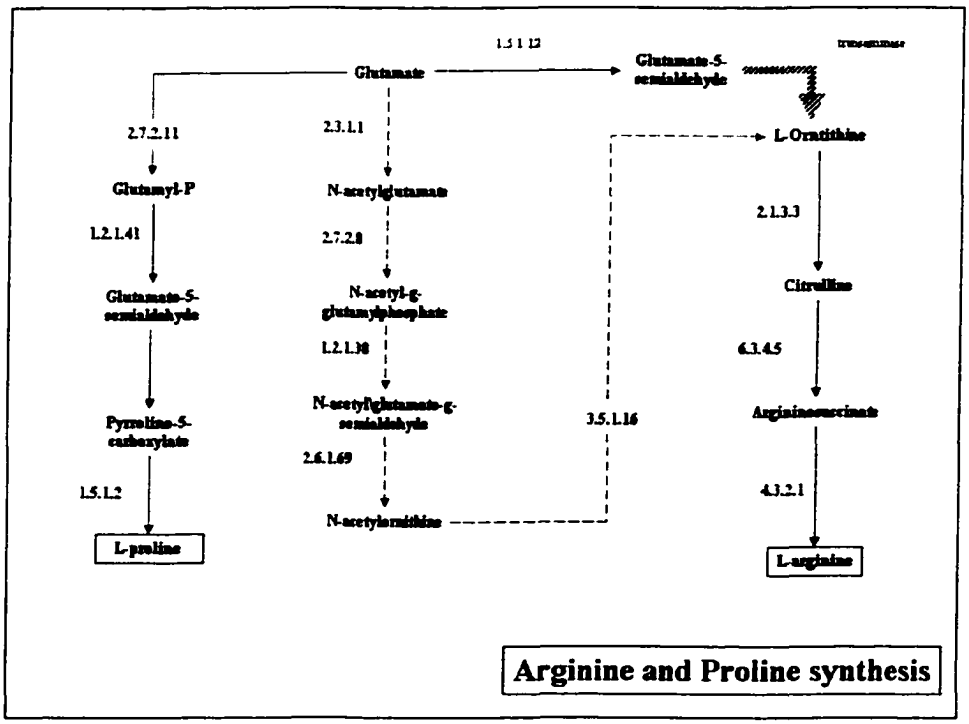


Figure 3.25. Biosynthesis of arginine and proline. The dotted arrows represent reactions by enzymes not found in *A. actinomycetemcomitans*, while the solid arrows are reactions by enzymes that were encoded in the genome.

EC#	Description	Genes in *A. actinomycetemcomitans*
1.4.1.4	NADP-specific glutamate dehydrogenase (NADP-gdh)	Yes
2.3.1.1	Amino acid acetyltransferase; N-acetylglutamate synthase	No
2.7.2.8	Acetylglutamate kinase	No
1.2.1.38	N-acetyl-g-glutamylphosphate reductase	No
2.6.1.69	N_2-Acetylornithine 5-transaminase	No
1.5.1.2	Pyrroline-5-carboxylate reductase (P5CR) (p5c reductase)	Yes
3.5.1.16	Acetylornithine deacetylase (acetylornithinase)	No
2.1.3.3	Ornithine carbamoyltransferase chain I (otcase-1)	Yes
6.3.4.5	Argininosuccinate synthetase (citrulline--aspartate ligase)	Yes
4.3.2.1	Argininosuccinate lyase (argH)	Yes
2.7.2.11	g-Glutamylphosphate kinase glutamate 5-kinase (GK)	Yes
1.2.1.41	glutamate-5-semialdehyde dehydrogenase	Yes
1.5.1.12	1-Pyrroline-5-carboxylate dehydrogenase	Yes
2.1.3.3	Ornihine transcarbamoylase	Yes

1.5.1.2	Pyrroline-5-carboxylate reductase (P5CR) (p5c reductase)	Yes

Table 3.15. Enzymes involved in the biosynthesis of arginine and proline.

3.3.3.8 Histidine

As was observed for arginine biosynthsis, the genes that are needed to encode many of the histidine biosynthesis pathway enzymes are missing in *A. actinomycetemcomitans* as shown by figure 3.26 and table 3.16. This is a surprising observation considering that its close relative, *H. influenzae*, is capable of synthesizing histidine encodes all of the histidine biosynthetic enzymes. Interestingly, *H. pylori*, which populates the gut, lacks histidine biosynthetic enzymes. This lack or presence of histidine biosynthetic enzymes may reflect the difference in their respective ecological niche, as *H. influenzae* populates in the respiratory tract while *A. actinomycetemcomitans* populates the nutrient-rich oral cavity. However, as discussed above, *A. actinomycetemcomitans* has several general amino acid transport systems in addition to the *art* system involved in arginine transport, and although not discovered yet, it is likely that the *A. actinomycetemcomitans* genome also encodes the genes needed for histidine transport.

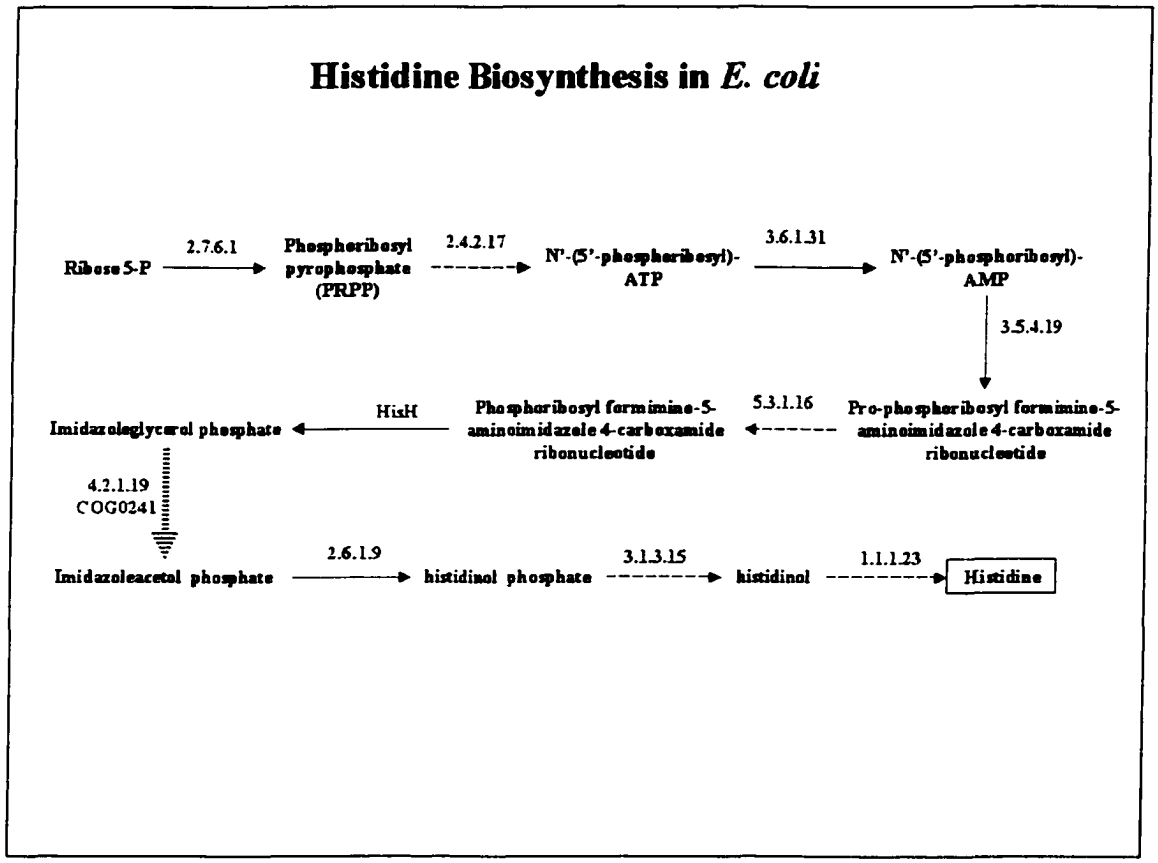


Figure 3.26. Histidine biosynthesis. Dotted arrows represent enzymes missing from *A. actinomycetemcomitans*. The thick arrow represent a possible functional hit from the COGs database.

EC#	Description	*A. actinomycetemcomitans*
2.7.6.1	Ribose-phosphate pyrophosphokinase (phosphoribosyl pyrophosphate synthetase)(prs, prsA)	Yes
2.4.2.17	ATP phosphoribosyltransferase (hisG)	No
3.6.1.31	phosphoribosyl-ATP pyrophosphohydrolase (hisIE, hisI)	Yes
3.5.4.19	Phosphoribosyl-AMP cyclohydrolase	Yes
5.3.1.16	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase(hisA)	No
2.4.2.-	hisF protein (CYCLase)(hisF)	No
2.4.2.-	Amidotransferase (hisH)	Yes
4.2.1.19	histidinol- phosphatase (hisB)	No
2.6.1.9	Histidinol-phosphate transaminase (hisC)	Yes
3.1.3.15	Imidazoleglycerol-phosphate dehydratase (IGPD)	No
1.1.1.23	Histidinol dehydrogenase (HDH) (hisD)	No

Table 3.16. Histidine biosynthetic enzymes.

3.3.4 Nucleotide Metabolism

3.3.4.1 Pyrimidine nucleotides

Analysis of the *A. actinomycetemcomitans* genomic sequence indicates that this organism is not capable of *de novo* synthesis of pyrimidines (Figure 3.27 and Table 3.17). Interestingly, *H. influenzae* also is missing the genes for these enzymes.

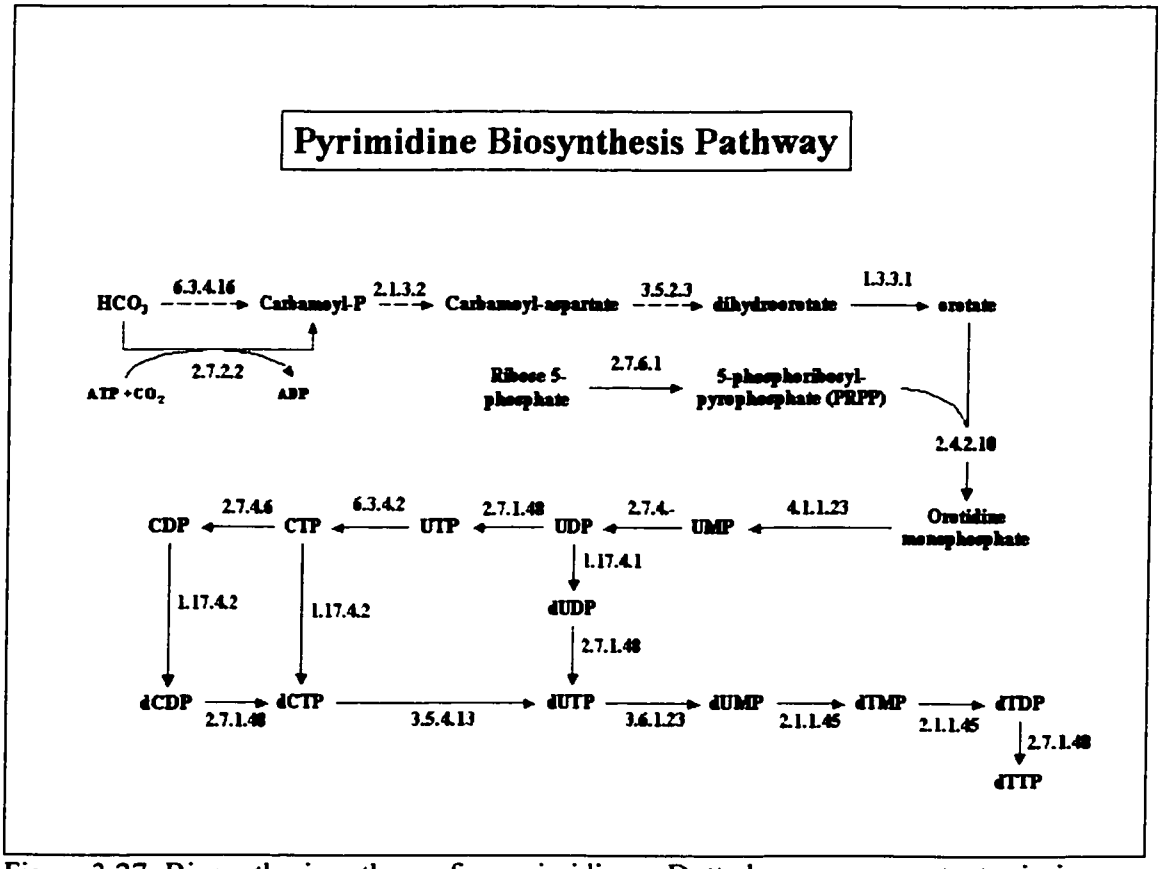


Figure 3.27. Biosynthesis pathway for pyrimidines. Dotted arrows represtent missing enzymes and solid arrows represents enzymes found in *A. actinomycetemcomitans*.

However, *A. actinomycetemcomitans* appears to be capable of salvage synthesis of pyrimidines as shown in figure 3.28 and listed in table 3.17. The pyrimidine salvage

pathway provides the cell with bases from exogenous nucleosides. Metabolizing exogenous nucleosides to free bases and pentose moieties also may provide the cell with a needed carbon source which can be metabolized for energy production (**Neuhard et al., 1996**). The gene for two nucleoside transport proteins also are present in the *A. actinomycetemcomitans* genome, and thus the purine and the pyrimidine substrates for the salvage pathway readily can enter the cell. Both NupC, which can import all nucleosides except for guanosine (**Craig et al., 1994**), and UraA, a uracil-specific transporter, previously were observed in the *A. actinomycetemcomitans* genome (**Andersen et al., 1995**). The UraA transporter is encoded in an operon with the *upp* gene that catalyzes the phosphoribosylation of uracil to UMP (**Andersen et al., 1995**), a gene organization which is conserved in other prokaryotes as seen in figure 3.29.

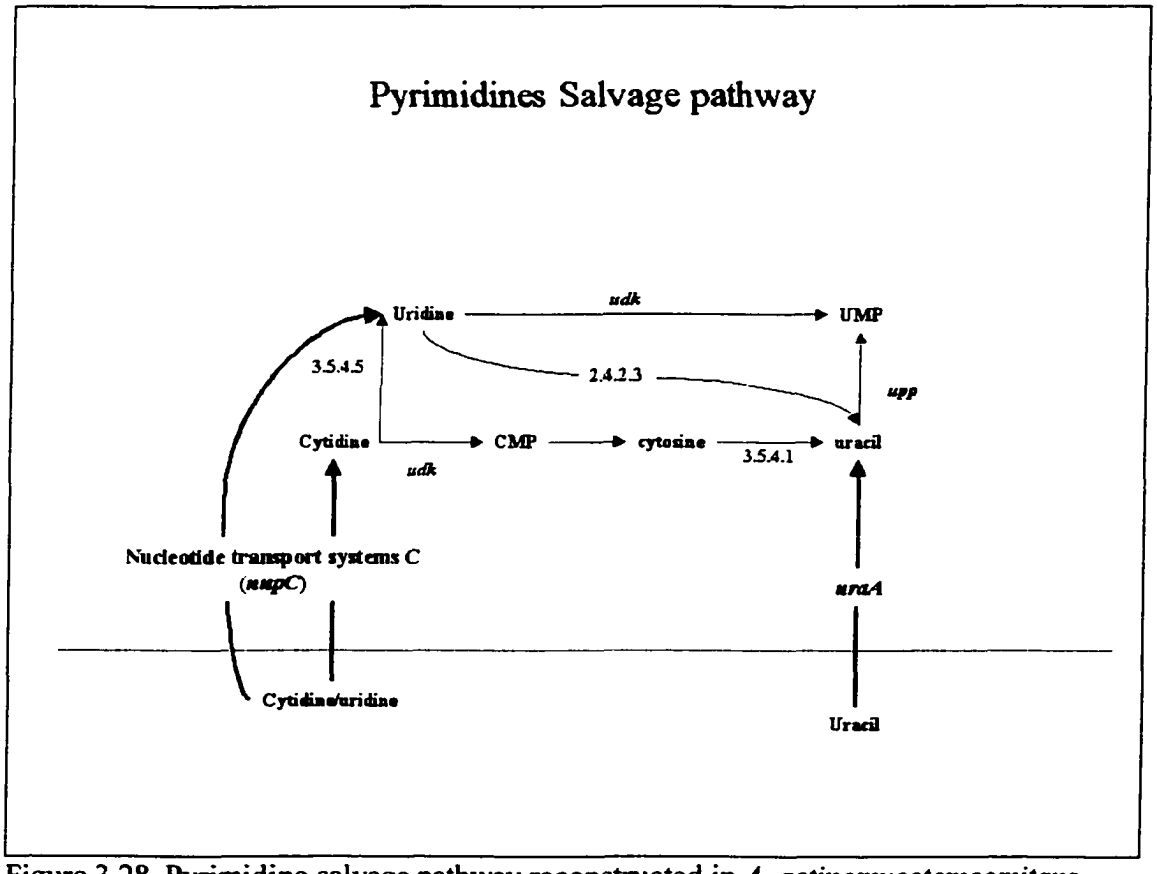


Figure 3.28. Pyrimidine salvage pathway reconstructed in *A. actinomycetemcomitans*.

EC#/Gene name	Description	Genes in *A. actinomycetemcomitans*
6.3.4.16	Carbamoylphosphate synthetase	No
2.7.2.2	Carbamate kinase	Yes
2.1.3.3	Ornithine carbamoyltransferase	Yes
2.1.3.2	Aspartate transcarbamoylase	No
3.5.2.3	Dihydroorotase	No
1.3.3.1	Dihydroorotate oxidase (pyrD)	Yes
2.4.2.10	Orotate phosphoribosyltransferase	Yes
2.7.6.1	phosphoribosyl pyrophosphate	Yes
4.1.1.23	Orotidine 5'-phosphate decarboxylase (pyrF)	Yes
2.7.4.14	Cytidylate kinase	Yes
2.7.4.-	UMP kinase (pyrH)	Yes
2.7.4.6	Nucleoside diphosphate kinase	Yes
6.3.4.2	CTP synthase (pyrG)	No
1.17.4.2	ribonucleoside-triphosphate reductase (nrdD)	Yes
3.5.4.13	dCTP deaminase	Yes
2.4.2.9	Uracil phosphoribosyltransferase (upp)	Yes
3.5.4.1	Cytosine deaminase	Yes
2.4.2.3	Uridine phosphorylase (UDRPase) (udp)	Yes
2.7.1.48	Uridine/cytidine kinase (udk)	Yes
3.5.4.5	Cytidine deaminase (cytidine aminohydrolase) (cdd)	Yes
2.7.4.6	Nucleoside diphosphate kinase (ndk)	Yes
2.1.1.45	Thymidylate kinase (dTMP kinase) (tmk)	Yes
2.1.1.45	Thymidylate synthase (thyA)	Yes
3.6.1.23	Deoxyuridine 5'-triphosphate nucleotidohydrolase (dut)	Yes
1.17.4.1	Ribonucleoside diphosphate reductase, (alpha chain) (nrdA)	Yes
1.17.4.1	Ribonucleoside diphosphate reductase, (beta chain) (nrdB)	Yes
3.5.4.13	(deoxycytidine triphosphate deaminase) (dcd)	Yes
nupC	nucleoside permease except guanosine	Yes
uraA	Uracil transport, ABC transporter	Yes
2.7.2.2	Carbamate kinase	Yes
1.17.4.2	ribonucleoside-triphosphate reductase (nrdD)	Yes

Table 3.17. Enzymes involved in pyrimidine metabolism.

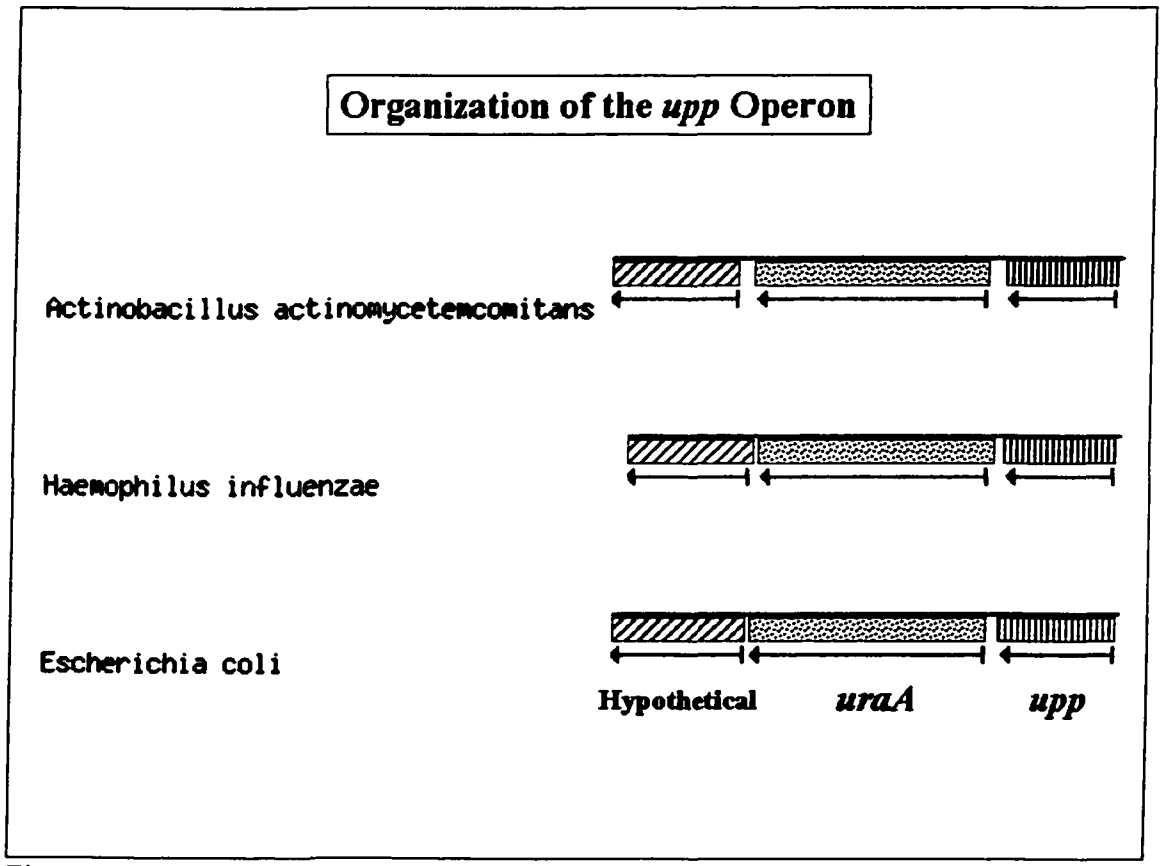


Figure 3.29. The conservation of the *upp* operon across different species.

3.3.4.2 Purine nucleotides

A. actinomycetemcomitans does not possess the genes that codes for the enzymes needed for the *de novo* synthesis of purines (Figure 3.30, Table 3.18). Therefore, as is required for pyrimidine metabolism, *A. actinomycetemcomitans* must import the required purine nucleotides through a transport system (Figure 3.31).

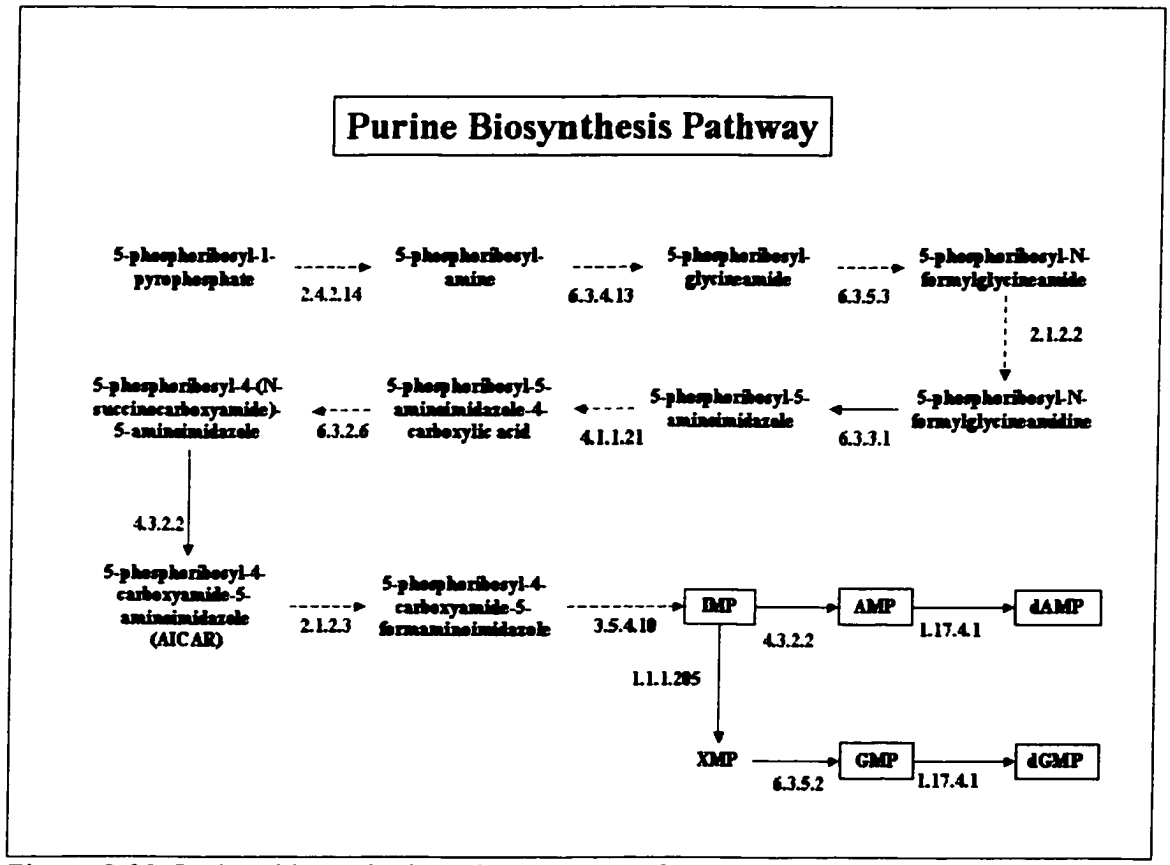


Figure 3.30. Purines biosynthetic pathway in *E. coli*. Dotted arrows are enzymes that were not found in *A. actinomycetemcomitans* and solid arrows represent enzymes whose genes are present.

EC#	Description	Genes in *A. actinomycetemcomitans*
2.4.2.14	Amidophosphoribosyltransferase (purF)	No
6.3.4.13	Phosphoribosylamine--glycine ligase (GARS) (purD)	No
6.3.5.3	Phosphoribosylformylglycinamidine synthase (purL)	No
2.1.2.2	Phosphoribosylglycinamide formyltransferase (purN)	No
6.3.3.1	Phosphoribosylformylglycinamidine cyclo-ligase (purM)	No
4.1.1.21	Phosphoribosylaminoimidazole carboxylase (purE)	No
6.3.2.6	Phosphoribosylaminoimidazole-succinocarboxamide synthase (purC)	No
4.3.2.2	Adenylosuccinate lyase (adenylosuccinase) (purB)	Yes
2.1.2.3	Phosphoribosylaminoimidazolecarboxamide formyltransferase	No
3.5.4.10	IMP cyclohydrolase (inosinicase) (IMP synthetase) (purH)	No
1.17.4.1	Ribonucleoside diphosphate reductase, (alpha chain) (nrdA)	No
1.17.4.1	Ribonucleoside diphosphate reductase, (beta chain) (nrdB)	No
1.1.1.205	Inosine-5'-monophosphate dehydrogenase (guaB)	No
6.3.5.2	GMP synthase (glutamine-hydrolyzing) (guaA)	No
2.4.2.7	Adenine phosphoribosyltransferase (apt)	Yes
6.3.4.4	Adenylosuccinate synthetase (IMP--aspartate ligase) (purA)	Yes
2.4.2.1	Purine nucleoside phosphorylase (inosine phosphorylase) (deoD)	Yes
1.1.1.205	Inosine-5'-monophosphate dehydrogenase (IMP dehydrogenase) (IMPDH) (IMPD) (guaB)	Yes
6.3.5.2	GMP synthase (glutamine-hydrolyzing) (glutamine amidotransferase)(guaA)	Yes
2.4.2.8	Hypoxanthine phosphoribosyltransferase (HPRT)(hpt)	Yes

Table 3.18. Enzymes involved in purine metabolism.

As discussed above, the NupC tranport system and nucleoside permease orthologue likely provide *A.actinomycetemcomitans* with the nucleosides needed for the salvage pathway synthesis of purines.

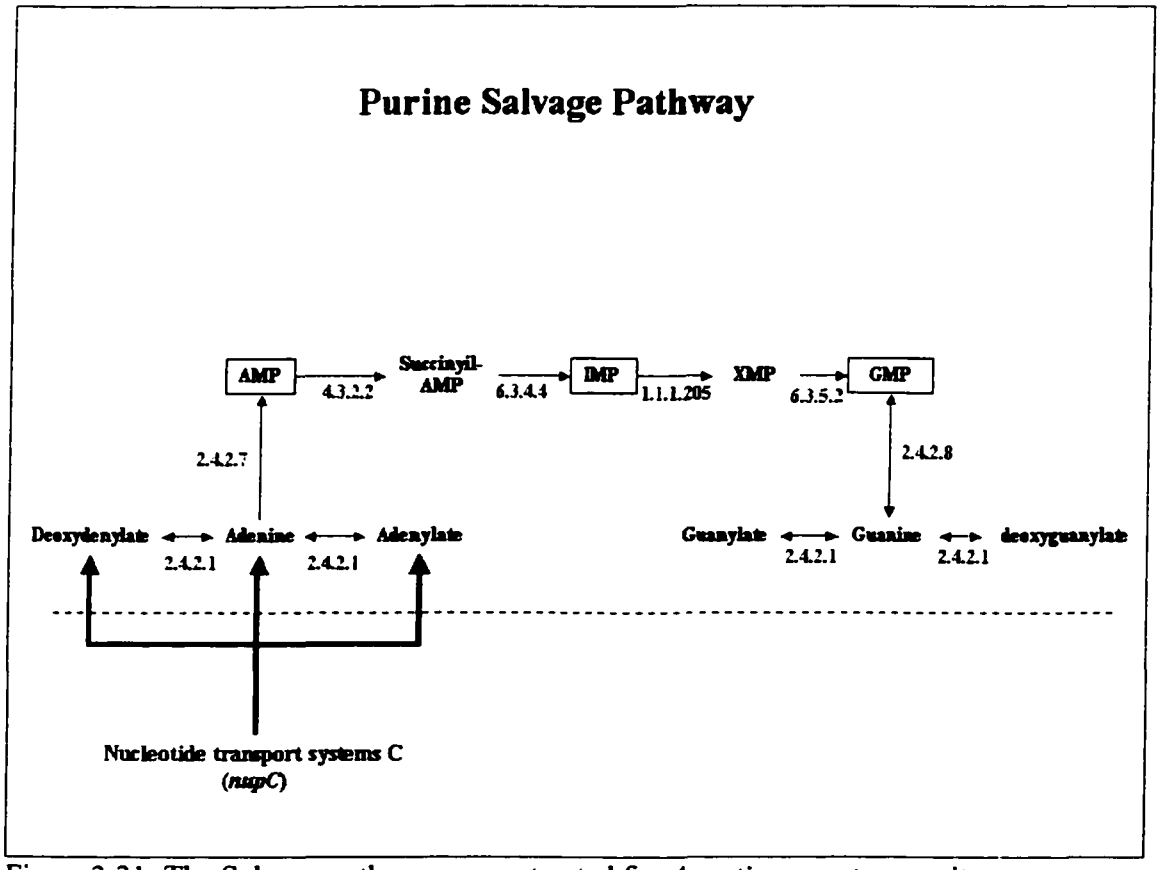


Figure 3.31. The Salvage pathway reconstructed for *A. actinomycetemcomitans*.

3.3.5 Biosynthesis of Cofactors and Vitamins

Vitamins and cofactors play a vital role in most cellular activities. Although *A. actinomycetemcomitans* is an opportunistic pathogen, analysis of its genome reveals that *A. actinomycetemcomitans* is capable of synthesizing riboflavin, folate, molybdenum cofactor, and biotin, while other cofactors such as pantothenate and thiamine likely are imported from the environment.

3.3.5.1 Riboflavin

Riboflavin, an essential cofactor in bacteria, participates in both bioenergetics and biosynthetic reactions. The enzymes of the riboflavin pathway therefore often are considered as useful antibacterial targets (**Bacher et al., 2001**). As shown below in figure 3.32 and table 3.19, all of the genes that encode the enzymes required for the synthesis of riboflavin are present in the *A. actinomycetemcomitans* genome.

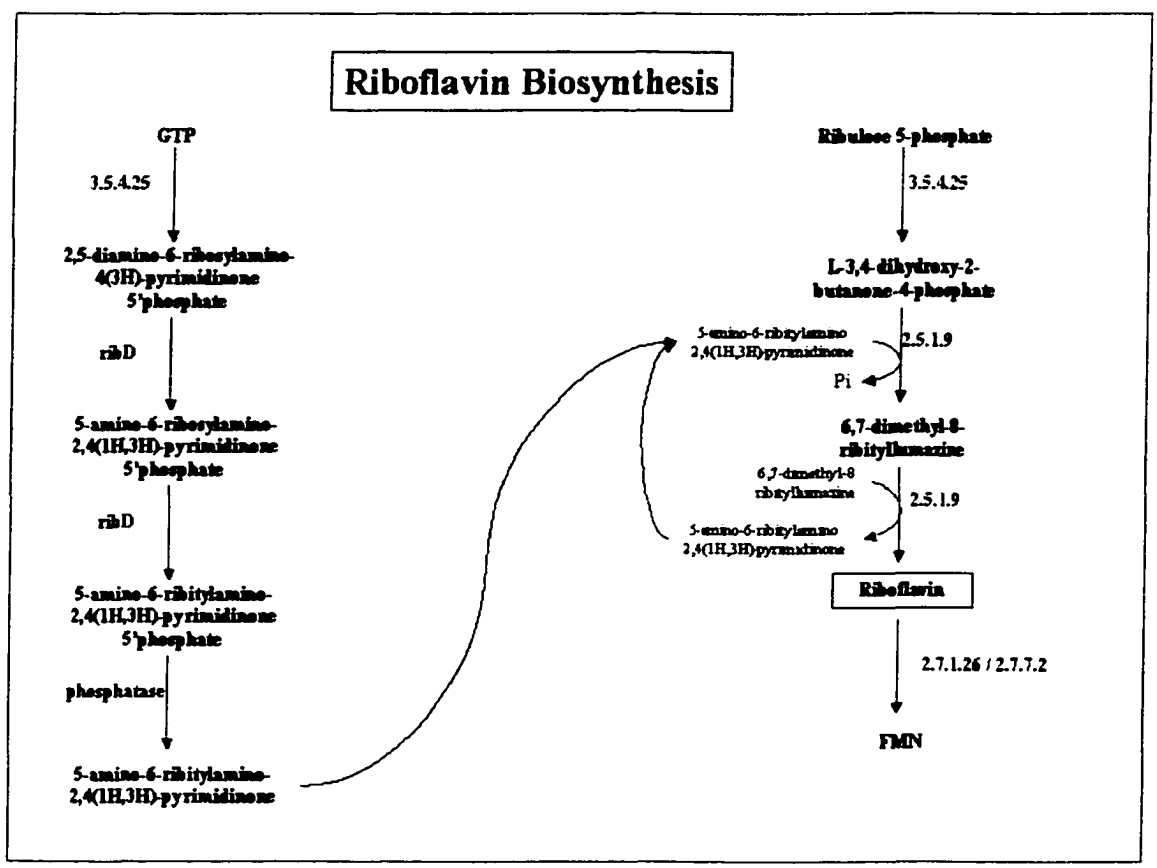


Figure 3.32. Riboflavin Biosynthesis.

EC#	Description	Genes in *A. actinomycetemcomitans*
3.5.4.25	GTP cyclohydrolase II (ribA)	Yes
3.5.4.-	Riboflavin-specific deaminase; pyrimidine deaminase (ribD)	Yes
3.5.4.25	3,4-dihydroxy-2-butanone 4-phosphate synthase (ribB)	Yes
2.5.1.9	Riboflavin synthase alpha chain (ribC)	Yes
2.5.1.9	Riboflavin synthase beta chain (ribE)	Yes
2.7.1.26 / 2.7.7.2	Riboflavin kinase / FMN adenylyltransferase	Yes

Table 3.19. Enzymes involved in the riboflavin synthesis pathway.

3.3.5.2 Folate Biosynthesis

Folic acid derivatives play an important role in a variety of cellular functions that include one-carbon transfer and examination of the *A. actinomycetemcomitans* genomic sequence reveals that the genes for all the folate biosynthetic pathway enzymes are present (Figure 3.33 and table 3.20) and that the genomic organization of the genes is similar to that of *H. influenzae* **(Sutton et al., 1995 and KEGG)**. Tetrahydrofolate serves as a one-carbon recipient during the synthsis of glycine from serine while methylation of dUMP to dTMP needs 5,10-methylenetetrahydrofolate, and 10-formyltetrahydrofolate is required for formylation of fMET-tRNA-fmet as well as being a one-carbon donor for purine biosynthesis. Folic acid derivatives also play other roles in cellular processes, such as pyrimidine dimer repair by DNA photolyase **(Green et al., 1996)**

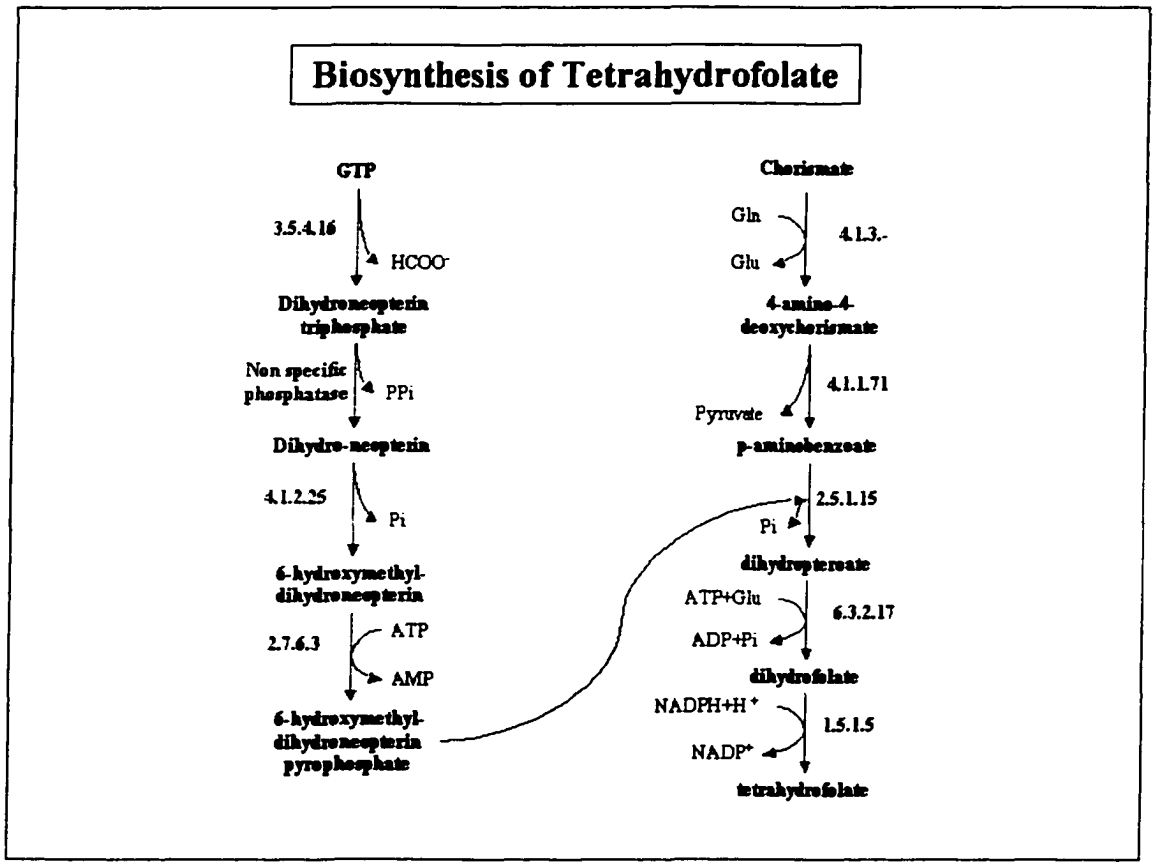


Figure 3.33. Biosynthesis of tetrahydrofolate.

EC#	Description	Genes in *A. actinomycetemcomitans*
3.5.4.16	GTP cyclohydrolase I(folE)	Yes
4.1.2.25	Dihydroneopterin aldolase	Yes
2.7.6.3	2-Amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase (folK)	Yes
4.1.3.-	para-aminobenzoate synthase component I (pabB)	Yes
4.1.1.71	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase (menD)	Yes
2.5.1.15	Dihydropteroate synthase (folP)	Yes
6.3.2.17	dihydrofolate synthase (folC)	Yes
1.5.1.5	Methylenetetrahydrofolate dehydrogenase	Yes

Table 3.20. Enzymes involved in folate biosynthesis pathway.

Interestingly, the enzyme dihydrofolate reductase (1.5.1.5) has been used as a target for drug therapy. Specifically, the drug methoprim was developed to target this enzyme (**Green et al., 1996**). However, shortly after its introduction, two types of the

dihydrofolate reductase enzyme were discovered, both of which were encoded by and expressed from an R-plasmid. The type I enzyme was several thousand-fold less sensitive to the drug, while type II was not sensitive at all (**Green et al., 1996**).

3.3.5.3 Molybdenum Cofactor

The molybdopterin cofactor is a required component of every enzyme that contains molybdopterin (**Rajagopalan, 1996**). All of the genes needed to encode the enzymes involved in the biosynthesis of the molybdenum cofactor were found in the *A. actinomycetemcomitans* genome. Although it's instability has prevented its isolation in a pure form, studies on the decarboxamidomethyl derivative of molybdopterin have resulted in a proposed structure for this cofactor as illustrated in figure 3.34. Reactions catalyzed by molybdenum-containing enzymes are involved in transfer of an oxygen atom from or to water (**Kisker et al., 1997**). In *E. coli*, mutations in any or all of the genes involved in the biosynthesis of molybdopterin were shown not to be lethal because of the existence of anaerobic pathways with molybdopterin-independent enzymes (**Rajagopalan, 1996**).

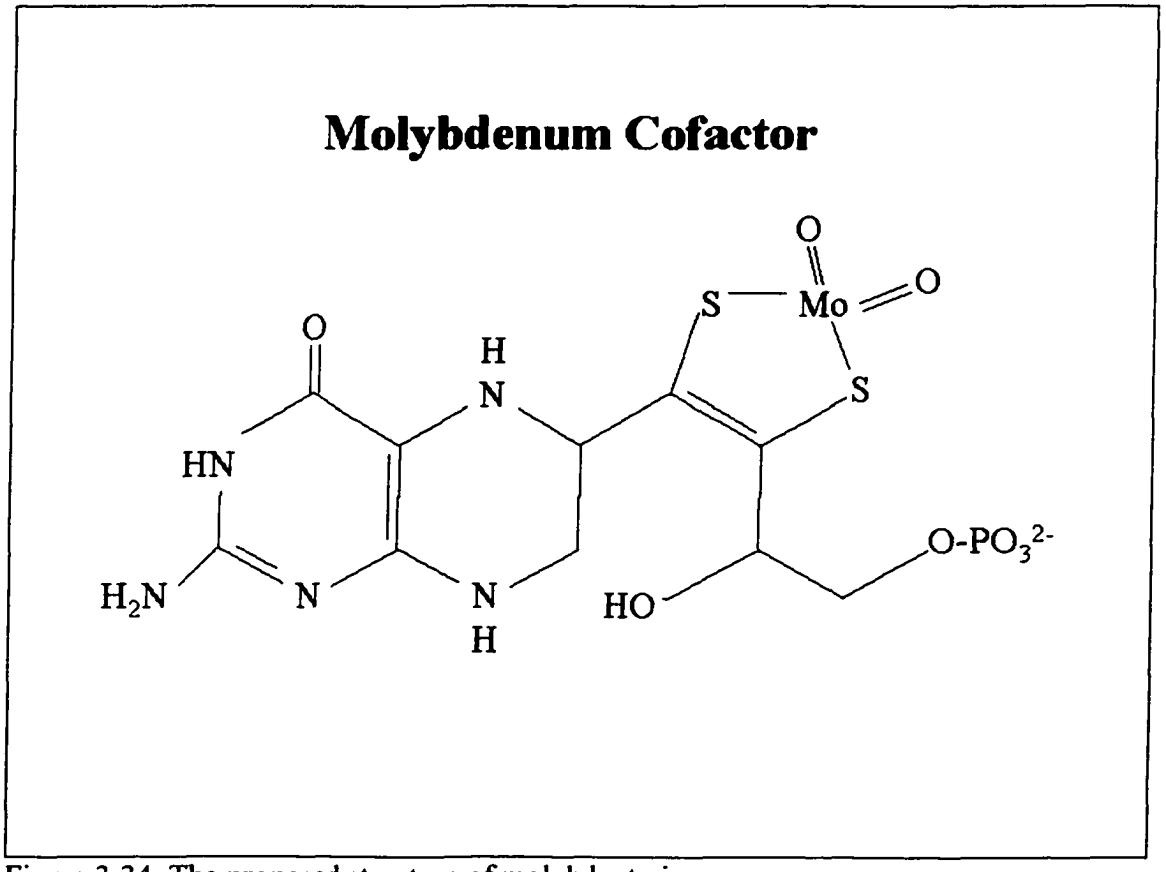


Figure 3.34. The proposed structure of molybdopterin.

3.3.5.4 Thiamine

Thiamine consists of two units, a pyrimidine unit (4-amino-5-hydroxymethyl-2-methylpyrimidine), and a thiazole unit (5(-2-hydroxymethyl)-4-methylthiazole). The thiamine biosynthetic pathway differs among bacteria in both to the synthesis of the individual units and the union of these two moieties to form the final thiamine product. Although thiamine derivatives are vital to electron transport and central metabolic pathways, not much is known about the *de novo* synthesis or salvage pathways (**Webb et al., 1998**). Recently, an operon of three thiamine ABC transporters designated *thiABC*,

was discovered by Webb and colleagues in *S. typhimurium* (**Webb et al., 1998**). Similarly, *Actinobacillus* also possess the genes for three thiamine transporters (Table 3.21), suggesting that *A. actinomycetemcomitans* is capable of acquiring thiamine and its phosphoester through transporters from the external environment. Interestingly, although the *H. influenzae* genome contains the genes for *de novo* thiamine biosynthesis (**Sutton et al., 1995 and KEGG**), these genes, except for the thiamine-phosphate kinase, are lacking in *A. actinomycetemcomitans*. Again, this might reflect the environmental growth differences between the two organisms. Figure 3.35 shows the reactions involved in thiamine metabolism with the *A. actinomycetemcomitans* and *H. influenzae* enzymes highlighted.

Gene name	Description	Genes in *A. actinomycetemcomitans*
tbpA	Thiamin ABC transporter, periplasmic-binding protein	Yes
ThiQ	Thiamin ABC transporter,ATP-binding protein	Yes
thiP	Thiamin ABC transporter, permease protein	Yes

Table 3.21. Enzymes involved in thiamine transport.

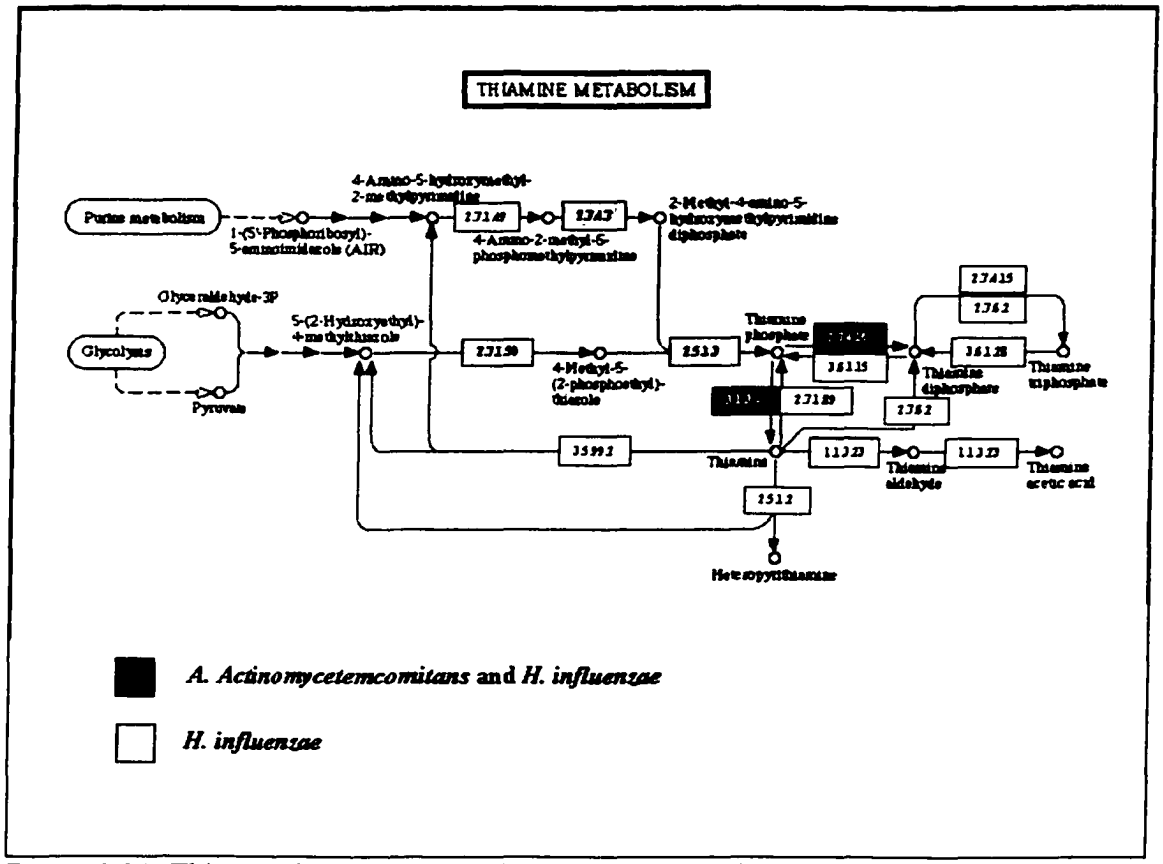


Figure 3.35. Thiamine biosynthesis pathway. The lighter boxes represent enzymes found in *H. influenzae*. The darker boxes represent enzymes found in both *H. influenzae* and *A. actinomycetemcomitans*.

3.3.5.5 Pantothenate and Coenzyme A

Pantothenate, a component of the vitamin B complex, serves as a precursor for coenzyme A and acyl carrier protein (**Jachowsky, 1996**). A thorough examination of the *A. actinomycetemcomitans* genomic sequence failed to reveal any of the genes that encode the enzymes involved in the biosynthetic pathway for pantothenate and CoA. This prompted a further search for an alternative means to produce these compounds and led to the identification of the presence of PanF-like gene. Pan F, which has been characterized in *E. coli*, is responsible for the sodium-dependent pantothenate uptake

from the environment (**Jackowski et al., 1990**). PanF is an integral membrane carrier with 13 predicted hydrophobic membrane spanning domains (**Jackowski et al., 1990**) as illustrated in figure 3.36 below, and therefore it is likely that this transmembrane protein is responsible for pantothenate uptake into *A. actinomycetemcomitans*.

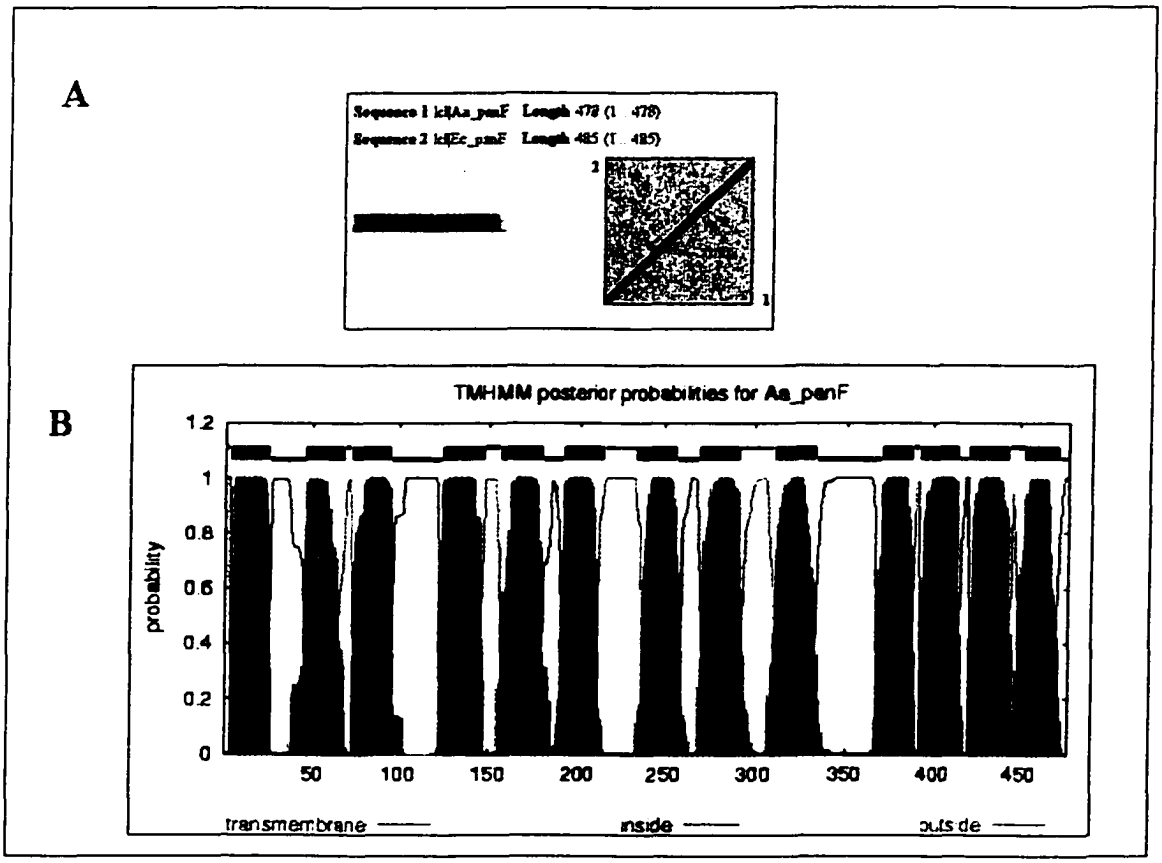


Figure 3.36. PanF protein. Panel A. alignment of PanF proteins from *A. actinomycetemcomitans* (Aa) and *E. coli* (Ec). B. The predicted transmembrane domains in *A. actinomycetemcomitans* ' PanF protein using the TMHMM algorithm.

Although the gene for ACP synthase was not detected in *A. actinomycetemcomitans* by a BlastP homology search, a COG analysis did reveal a putative ACP synthase. Based on this, a putative pathway for CoA and ACP synthesis

was reconstructed as illustrated in figure 3.37 and table 3.22. Therefore, it is very likely that coenzyme A and acyl carrier protein in *A. actinomycetemcomitans* are produced via a pantothenate.

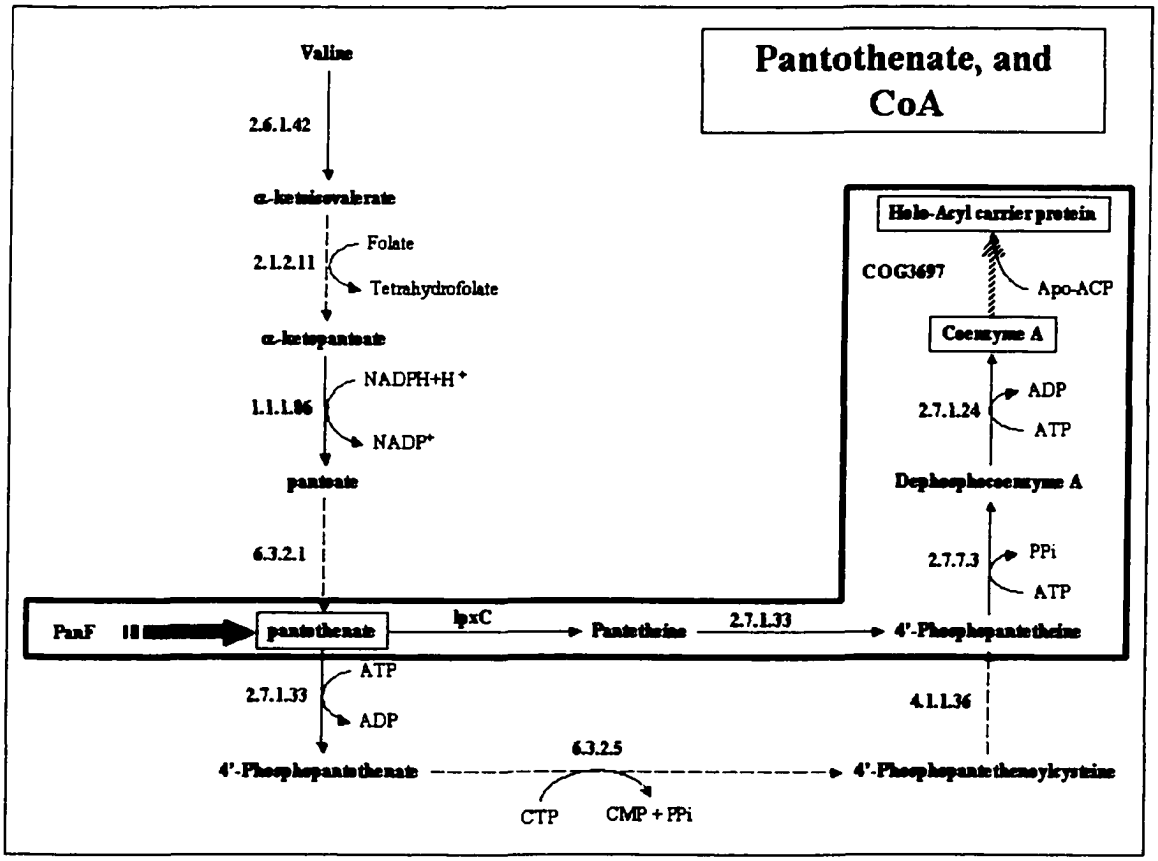


Figure 3.37. Biosynthsis of Pantothenate, CoA, and ACP in *A. actinomycetemcomitans*. The putative pathway for CoA, and ACP synthesis is indicated with a thick border. Dashed lines indicate enzymes missing from *A. actinomycetemcomitans* analysis. Solid lines indicate enzymes found from the sequence analysis. The putative ACP synthase is shown as COG3697.

EC#/Gene name	Description	Genes in *A. actinomycetemcomitans*
2.6.1.42	Branched-chain amino acid aminotransferase (ilvE)	Yes
2.1.2.11	3-Methyl-2-oxobutanoate hydroxymethyltransferase	No
1.1.1.86	ketol-acid reductoisomerase (ilvC)	Yes
6.3.2.1	Pantoate--beta-alanine ligase (panC)	No
2.7.1.33	Pantothenate kinase (panK)	Yes
6.3.2.5	Phosphopantothenate--cysteine ligase	No
4.1.1.36	Phosphopantothenoylcysteine decarboxylase	No
3.5.1.-	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase (lpxC)	Yes
2.7.1.33	Pantothenate kinase	Yes
2.7.7.3	Pantetheine-phosphate adenylyltransferase	Yes
2.7.1.24	Dephospho-CoA kinase	Yes
2.7.8.7	Holo-[acyl-carrier protein] synthase (holo-ACP synthase)(acpS, dpj)	No
PanF	Na/pantothenate symporter	Yes

Table 3.22. Enzymes envolved in pantothenate synthesis.

3.3.5.6 Biotin

Biotin is an essential cofactor in carboxyl transfer reactions (**DeMoll, 1996**). Although its biosynthetic pathway is not well characterized, its synthesis from pimeloyl-CoA has been reported and it appears to occur in several microorganisms (**DeMoll, 1996**). Analysis of the *A. actinomycetemcomitans* genomic sequence reveals the presence of the genes that encode the enzymes for this pathway (Figure 3.38 and table 3.23).

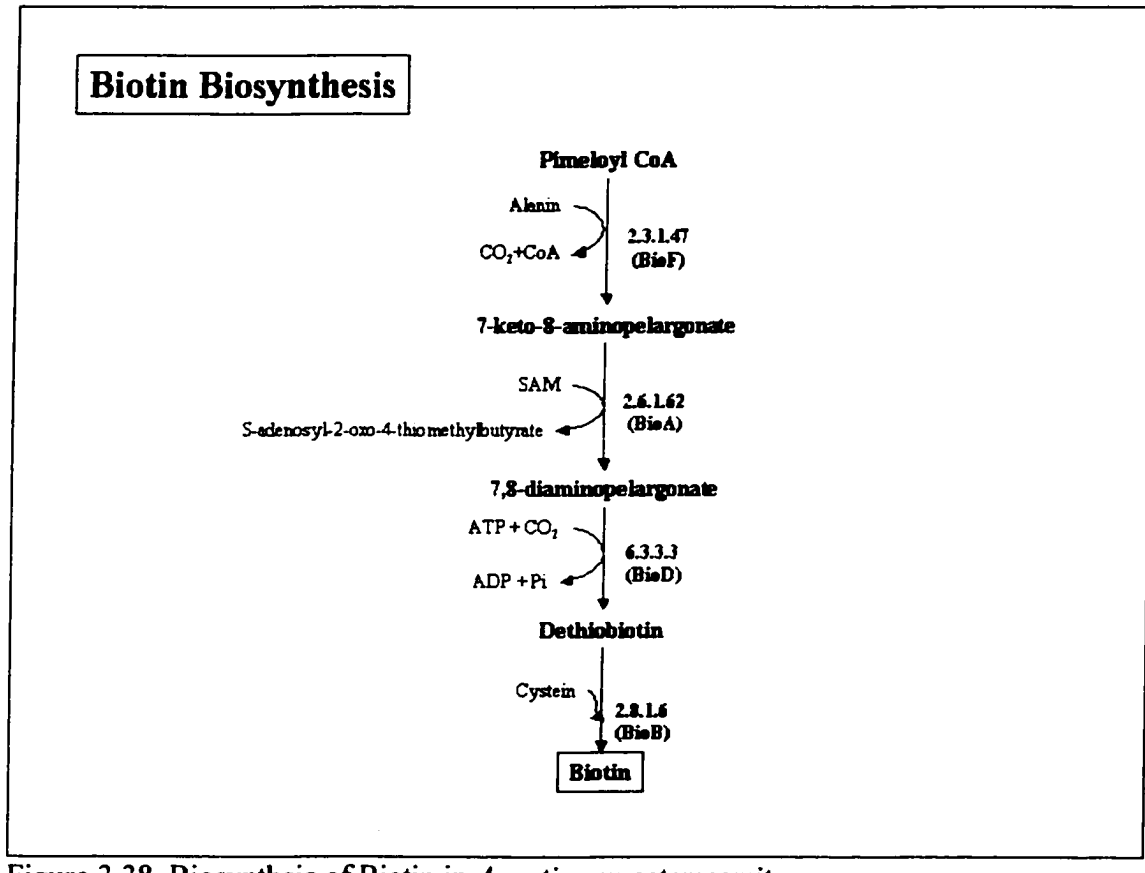


Figure 3.38. Biosynthsis of Biotin in *A. actinomycetemcomitans*.

EC#	Description	Genes in *A. actinomycetemcomitans*
2.3.1.47	8-amino-7-oxononanoate synthase (bioF)	Yes
2.6.1.62	Adenosylmethionine--8-amino-7-oxononanoate transaminase (bioA)	Yes
6.3.3.3	Dethiobiotin synthetase (bioD)	Yes
2.8.1.6	Biotin synthetase (bioB)	Yes

Table 3.23. Enzymes involved in biotin biosynthesis.

The genes for the enzymes involoved in biotin biosynthesis in *A. actinomycetemcomitans* are organized in an operon similar to that observed in *H. influenzae* but different from that in *E. coli*, where *bioA* is transcribed separately from *bioB, F, C*, and *D* **(DeMoll, 1996)**. In *H. influenzae* and *A. actinomycetemcomitans*, the *bioB* gene appears to be distinct from an operon where the *bioA, F, C*, and *D* genes are

colinear. Moreover, there is a hypothetical ORF between *bioF* and *bioC* that likely also codes for a protein that is involved in the biosynthetic pathway of biotin (Figure 3.39)

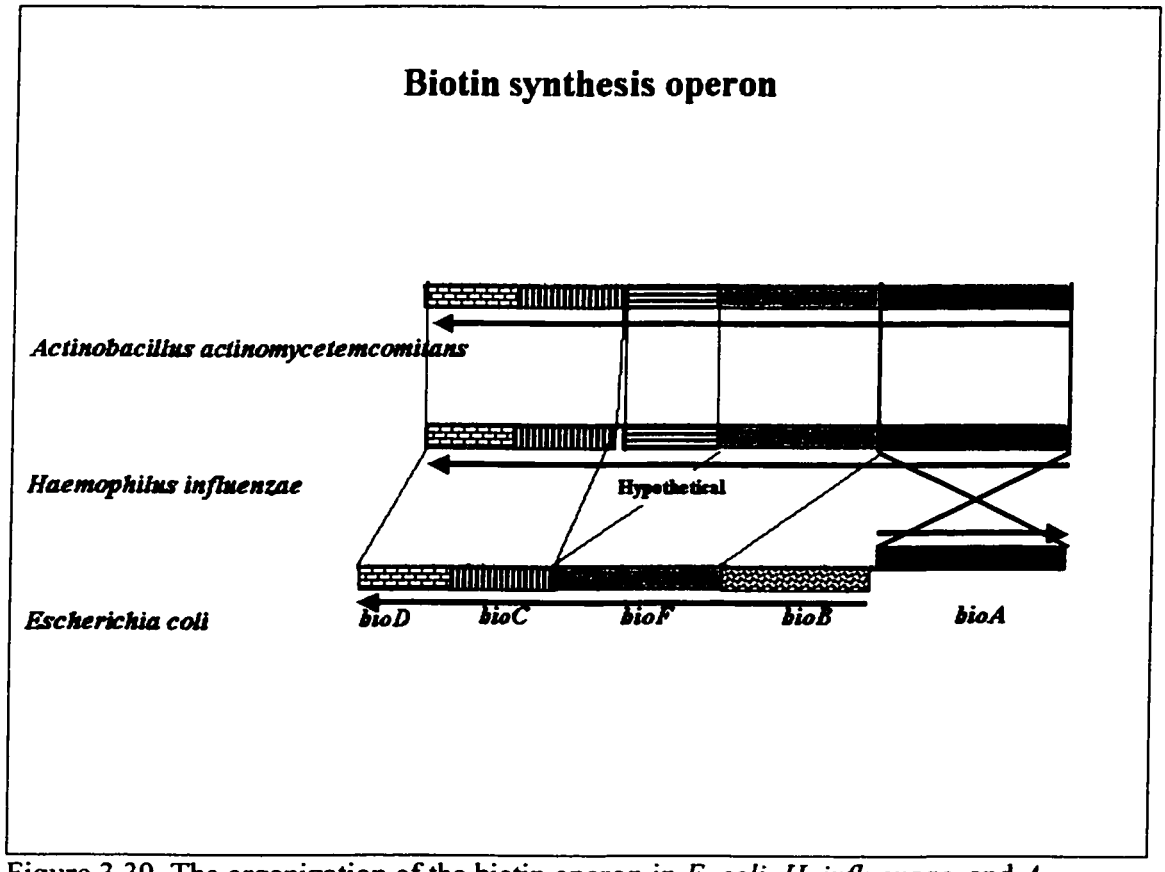


Figure 3.39. The organization of the biotin operon in *E. coli*, *H. influenzae*, and *A. actinomycetemcomitans*.

3.3.5.7 Nicotinamide

NAD (nicotinamide adenine dinucleotide) is a critical cofactor in redox reactions (**Martin et al., 2001**). In *E. coli*, and *S. typhimurium*, NAD is synthesized *de novo* from quinolinic acid via two possible pathways (**Foster et al., 1980**). Aerobic bacteria use tryptophane as a precursor for quinolate, whereas anerobic bacteria use aspartate with

fumarate or dihydroxyacetonephosphate as illustrated in figure 3.40 below (**Foster et al., 1980**).

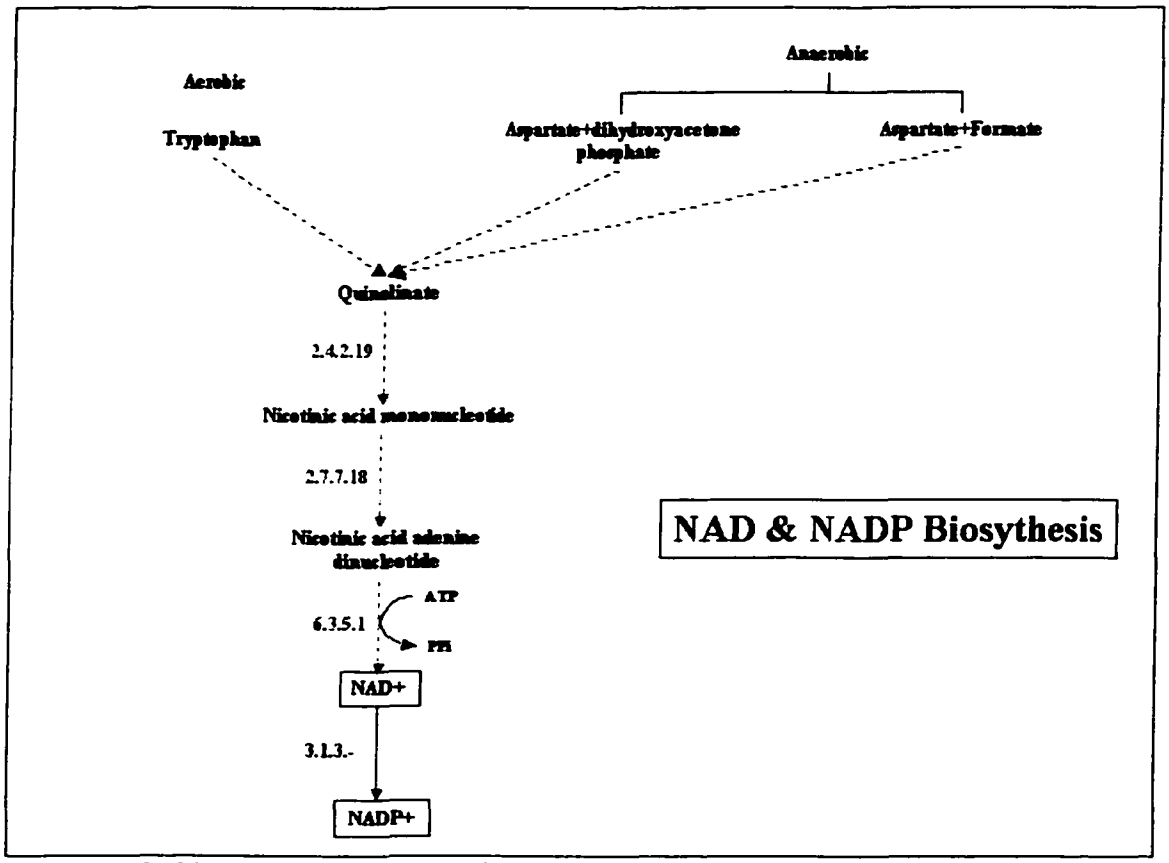


Figure 3.40. Biosynthesis pathway of NAD in bacteria except the family *Pasteurellaceae*.

Several members of the family *Pasteurellaceae* can not synthsize NAD *de novo* as they lack the enzymes that produce quinolinate (**Martin et al., 2001**). Thus, these organisms must acquire NAD from the environment either directly or as a limited number of NAD derivatives. The ability of a bacterial microorganism to uptake NAD has been used historically to distinguish between different genera of the family *Pasteurellaceae* (**Martin et al., 2001**) and results in dividing their species into two groups. One group,

called factor V-dependent, acquires NAD as a pyridine nucleotide precursor with an intact amidated carbonyl group and pyridine-ribose bond such as NAD, NMN (nicotinamide adenine mononucleotide) or NR (nicotinamide ribose) (**Martin et al., 2001**). Organisms, such as *H. influenzae*, belong to this group, since NAD must be present in the growth medium. The second group, is called factor V-independent, as it can acquire nictoinamide as a precursor for synthesis of NAD. This group includes *H. parainfluenzae* (**Cynamon et al., 1988**), *H. hemoglobinophilus* (**Kasarov et al., 1973**), *H. ducreyi* (**Martin et al., 2001**), *A. pleuropneumoniae* (**Martin et al., 2001**), and *A. actinomycetemcomitans* (**Martin et al., 2001**). The ability to utilize nicotinamide as a precursor to NAD requires nicotinamide phosphoribosyltransferase (2.4.2.12, NadV), an enzyme which only recently has been identified and its gene cloned in *H. ducreyi* (**Martin et al., 2001**).

Interestingly, nicotinamide phosphoribosyltransferase (NadV) occurs in eukaryotes as a preB-cell enhancing factor. In *A. actinomycetemcomitans* and *A. pleuropneumoniae*, this protein is chromosomally encoded while in *H. ducreyi* it is a plasmid encoded gene (**Martin et al., 2001**). Figure 3.41 illustrates the alignment of these two proteins from *A. actinomycetemcomitans* and *H. ducreyi.*

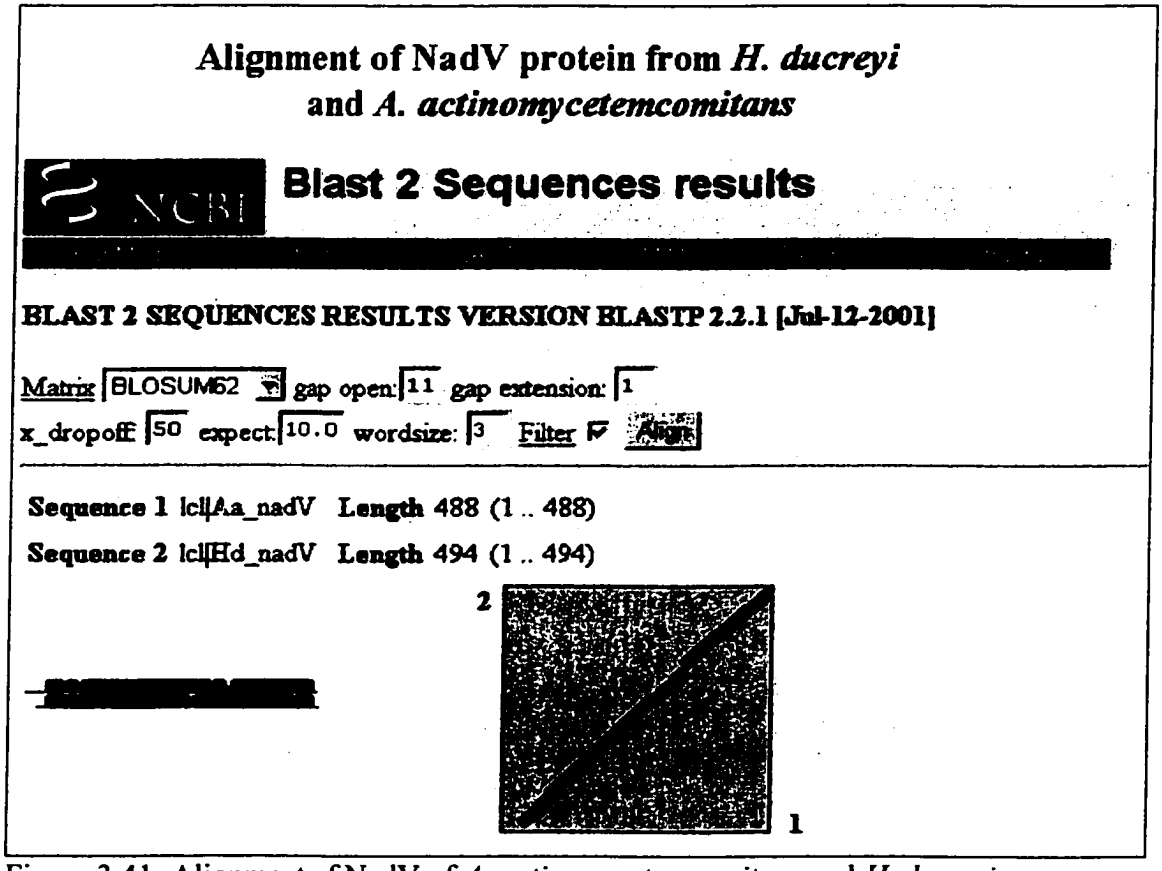


Figure 3.41. Alignment of NadV of *A. actinomycetemcomitans* and *H. ducreyi*.

Another protein involved in the synthesis of NAD, PnuC, is responsible for uptake of nicotinamide mononucleotide from the environment. The gene for PnuC was found through a COG search of the *A. actinomycetemcomitans* ORFs as it was not revealed by a BlastP search.

The gene for NadR encodes a protein that regulates the de novo synthesis of NAD and also regulates the NMN uptake protein PnuC (**Raffaelli et al., 1999**) which also was found encoded in the *A. actinomycetemcomitans* genome. A COG analysis of NadR in *A. actinomycetemcomitans* showed that the N-terminal of the protein is homologous to COG1056, a cluster representing the activity of nicotinamide mononucleotide

adenylyltransferase. The C-terminus, on the other hand, is homologous to COG3172, which represents the nicotinamide mononucleotide binding domain in different proteins (figure 3.42A). Additionally, a helix-turn-helix motif found in NadR suggesting a DNA binding domain associated with regulatory proteins (figure 3.42B), is missing from NadR of *H. influenzae*. Thus, the NadR homolog in *A. actinomycetemcomitans* is likely involved in both the uptake of NMN, the synthesis of NAD, and regulation of both. Figure 3.43 shows the putative pathway for NAD synthesis in *A. actinomycetemcomitans* and table 3.24 summarizes the proteins involved.

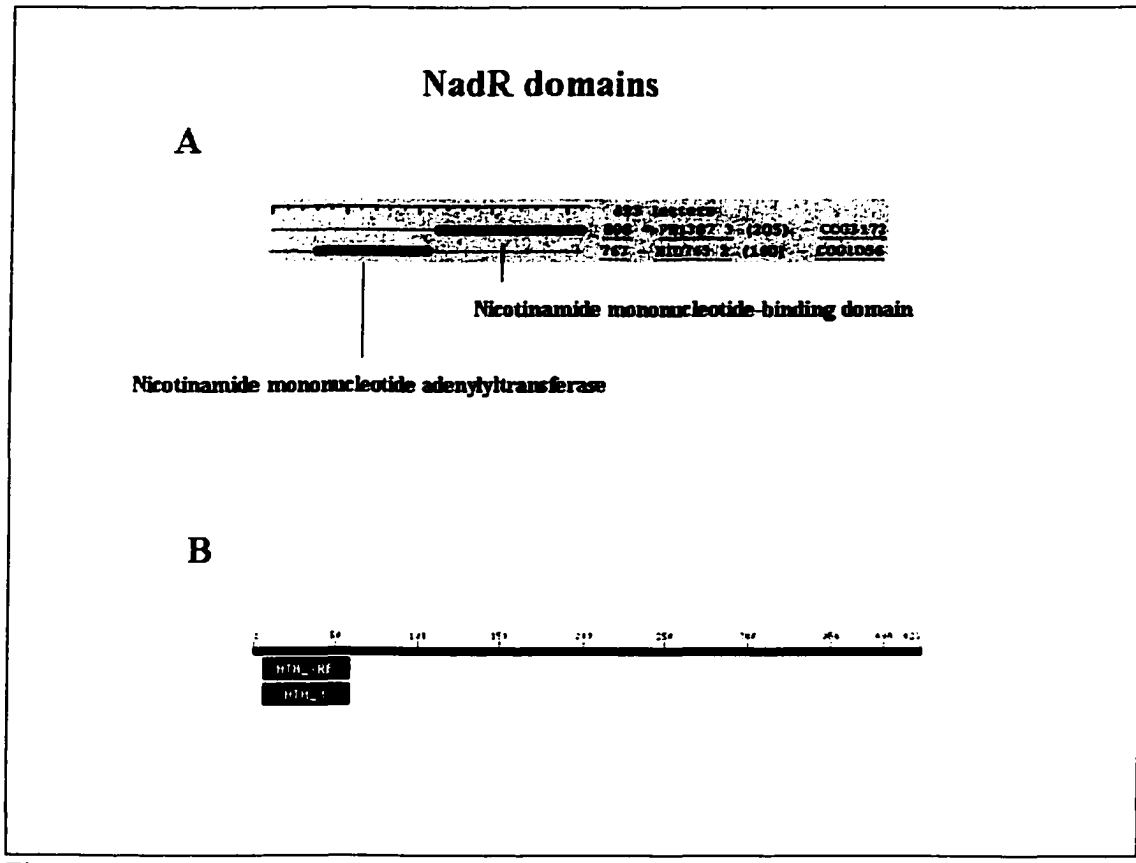


Figure 3.42. Schematic representation of the different domains in NadR protein. A. The results of COG analysis showing the two domains. B. Helix-turn-Helix motif for DNA binding.

EC#/Gene name	Description	Genes in *A. actinomycetemcomitans*
1.6.1.1	Pyridine nucleotide transhydrogenase	Yes
3.1.3.-	Alpha-ribazole-5'-phosphate phosphatase (cobC, phpB)	Yes
pnuC	Nicotinamide mononucleotide transporter (Coenzyme metabolism) -- COG3201.	Yes
2.4.2.12	Nicotinamide phosphoribosyltransferase	
NadR/ 2.7.7.1	Transcriptional regulator (nadI) / Nicotinamide-nucleotide adenylyltransferase – COG1056	Yes

Table 3.24. Proteins involved in NAD synthesis.

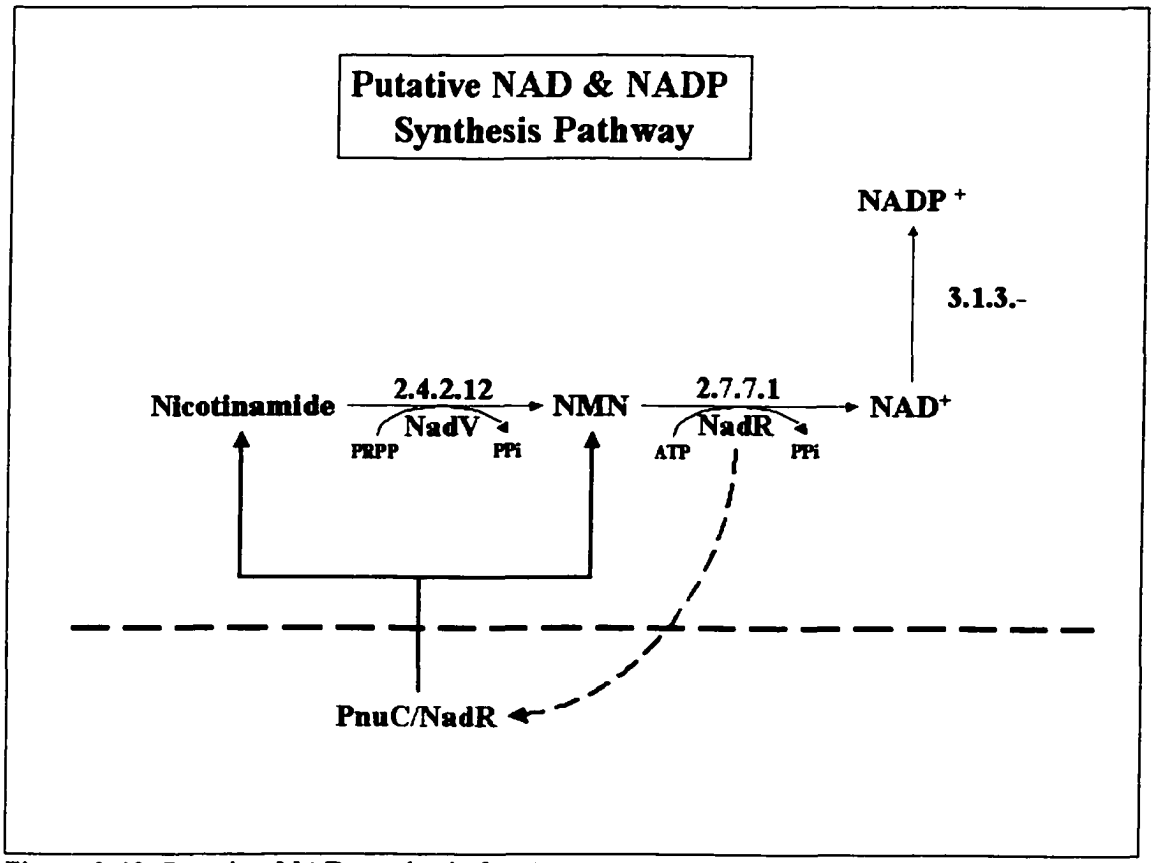


Figure 3.43. Putative NAD synthesis for *A. actinomycetemcomitans*.

3.3.6 Macromolecule Metabolism

Macromolecular metabolism broadly includes the metabolic processes involving DNA, RNA, and protein synthesis. Analysis of the *A. actinomycetemcomitans* genomic sequence suggests its replication, transcription, and translation machinary is similar to that in *E. coli* and most other bacteria. In addition, the *A. actinomycetemcomitans* genome analysis reveals the genes for multiple DNA repair pathways, similar to those observed in the KEGG database for *E. coli* and *H. influenzae* (**KEGG**).

3.3.6.1 DNA metabolism

Generally, replication in *A. actinomycetemcomitans* is similar to that of *E. coli* as suggested by the analysis of the *A. actinomycetemcomitans* genomic sequence. However, two proteins notably were absent in the *A. actinomycetemcomitans* genome. One of the missing genes encodes DnaC protein, a protein involved in the initiation complex in *E. coli* as part of the ABC complex which contains DnaA, and DnaB proteins (**White et al., 2000**). Interestingly, DnaC also is absent in *H. influenzae* suggesting that DnaC might not be essential for replication or that there is a functionally equivalent protein without homologue. In *E. coli*, this protein appears to be involved in forming a complex with DnaB and delivering it to the DNA-protein complex formed during the initiation of replication. Another gene that is absent is the gene for DNA polymerase II, the protein used in *E. coli* for DNA repair. This gene also is absent from several bacterial genomes, including *H. influenzae*, *Neisseria mengitidis*, *H. pylori*, and *Vibrio cholera*, indicating that repair can be accomplished using DNA polymerase I and III. More interestingly, the

epsilon subunit of DNA polymerase III, which is a 3'-5' exonuclease, also is absent from *A. actinomycetemcomitans*, *H. influenzae*, *N. meningitides*, *and H. pylori* suggesting that *A. actinomycetemcomitans* relys on the DNA polymerase I 3'-5' exonucleolytic activity.

Figure 3.44. summarizes DNA replication in *A. actinomycetemcomitans* and lists the proteins involved in the replication process. Appendix A contain a complete listings of all proteins involved in replication and other metabolic processes.

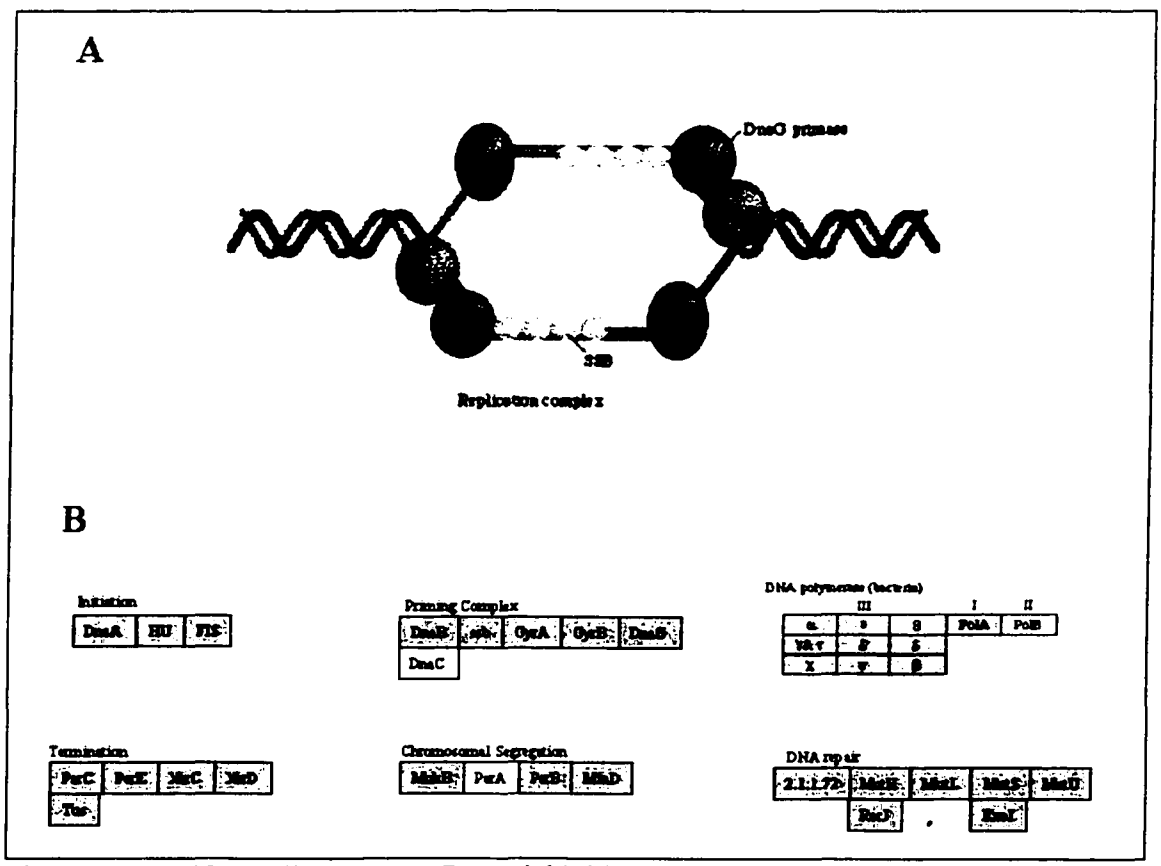


Figure 3.44. DNA replication. A. Bacterial initiation complex. B. *A. actinomycetemcomitans* genes involved in replication. The graphics were adapted from KEGG picture files.

3.3.6.1.1 Restriction and modification systems

The term restriction system was first coined to indicate the restriction of growth of bacteriophages in their respective bacterial host cells (**Redaschi et al., 1996**). These restriction systems are believed to have evolved to protect unicellular organisms from foreign invasion based on the principal of distinguishing self DNA from all others (**Redaschi et al., 1996**). There are two mechanisms by which bacteria can restrict foreign DNA (**Redaschi et al., 1996**). The first mechanism entails modifying its own DNA during and after replication. This is accomplished through the classical restriction-modification enzymes. Since foreign DNA is not modified, it can be identified and restricted. The second mechanisms involves restricting foreign DNA that has been modified differently than the host DNA. Here, host bacterial genomic DNA is protected because it is not modified (**Redaschi et al., 1996**).

The classical restriction-modification system consists of pairs of cognate enzymes that have the same DNA sequence specificity but different activity (**Redaschi et al., 1996**). One enzyme of the pair is a deoxyribonuclease, that cleaves the DNA at a specific, unmodified sequence. The other enzyme of the pair, methylates adenosyl or cytosyl in the same specific sequence recognized by the cognate endonuclease (**Redaschi et al., 1996**).

Restriction-modification systems are grouped into three classes. Type I modification and restriction enzymes are part of the same complex. They recognize asymmetric DNA sequences and they cut over 1000 bp from that sequence. In type II restriction-modification systems, the enzyme pair are separated. The sequence they recognize is symmetrical and so are the cleavage sites. The type III restriction-

modification system consists of one multifunctional enzyme with two separately encoded subunits.

The sequence of the *A. actinomycetemcomitans* genome reveals the presence of homologes to type I and type III restriction-modification enzymes but lacks genes for the type II system. Three homologs were found for the type I restriction-modification system, the hsdR (restriction activity), hsdM (modification activity), and hsdS (specificity). Another methylation system, the Dam system (DNA adenine methyltransferase) (**Marinus, 1996**), methylates adenosine to 6-methyladenosine in GATC sequences. It is apparent from the sequence of the *A. actinomycetemcomitans* genome that the Dam system is present and that it is similar to that of *E. coli* and *H. influenzae* (Figure 3.45). Table 3.25 below lists all modification-restriction systems found encoded in the *A. actinomycetemcomitans* genome.

EC#	Description	Genes in *A. actinomycetemcomitans*
2.1.1.72	DNA adenine methylase (deoxyadenosyl-methyltransferase)(dam)	Yes
3.1.21.3	Type I restriction enzyme EcoKI R protein (hsdR)	Yes
2.1.1.72	Type I restriction enzyme M protein (hsdM)	Yes
3.1.21.3	Type I restriction enzyme S protein (hsdS)	Yes
3.1.21.3	Type I restriction enzyme EcoR124II R protein	Yes
3.1.21.5	Type III restriction-modification enzyme homologus to EcoPI and LlaFI (deoxyribonuclease)	Yes
2.1.1.-	Modification methylase, type III R/M system.	Yes

Table 3.25. Restriction-Modification genes found in *A. actinomycetemcomitans*.

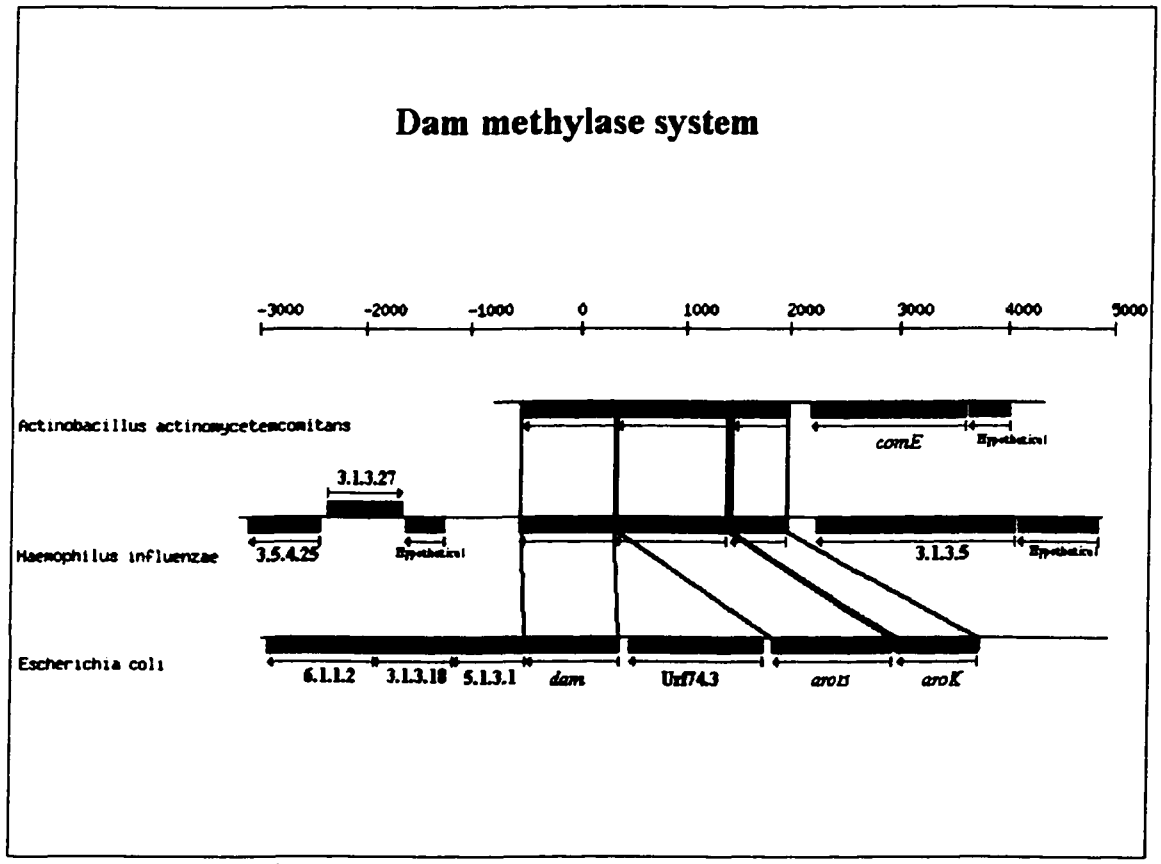


Figure 3.45. Comparison of the *dam* locus between *E. coli*, *H. influenzae*, and *A. actinomycetemcomitans*.

3.3.6.1.2 DNA Repair

Analysis of the sequence of the *A. actinomycetemcomitans* genome reveals that it encodes a full array of DNA repair proteins as listed in table 3.25. Based on this analysis, two types of DNA repair mechanisms were reconstructed for *A. actinomycetemcomitans*. The first type involves the reversal of DNA damage, by the Ada protein (2.1.1.63) (Figure 3.46A).

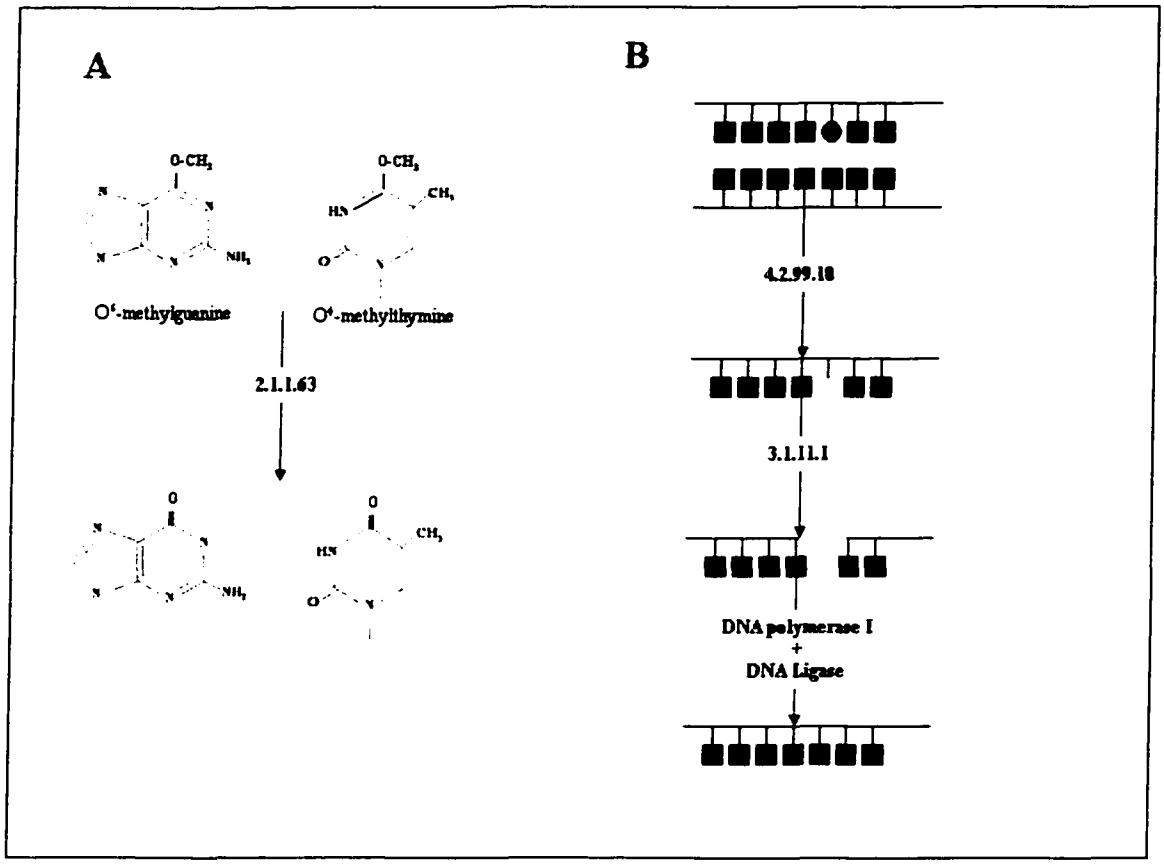


Figure 3.46. DNA repair. A. Damage reversal and B. base-excision repair pathways predicted in *A. actinomycetemcomitans.*

The second type of damage repair results from more complicated damage to the individual bases that can not be reversed by a single biochemical step. Here, there are several repair mechanisms which initially were discovered in *E. coli*. Analysis of the genomic DNA sequence reveales that the genes for this base excision repair pathway are encoded in the *A. actinomycetemcomitans* genome. Base-excision repair is represented in Figure 3.46B. Here, DNA glycosylase enzyme excises the damaged base creating an apurinic/apyrimidinic site in the DNA which then is followed by a single-stranded break through an endonuclease creating a 5'-deoxyribose terminal which then is removed by

DNA deoxyribophosphodiesterase. The resulting gap is closed by DNA polymerase and DNA ligase. Generally, DNA glycosylases recognize specific damaged bases, or specific types of damages in bases (**Friedberg et al., 1995**); however, some of the glycosylases have more relaxed damage recognition stringency.

A more elaborate DNA repair takes place when the damage encompasses several bases, such as the damage caused by UV irradiation, producing pyrimidine dimers, or damage by cross linking agents such as nitrogen mustard or psoralen. This damage can be repaired by a three-step nucleotide-excision repair mechanism which involves three enzymes that in *E. coli* have been identified by their UV-sensitive mutants (**Thomas et al., 1985**). Since the genes for all three enzymes, UvrA, UvrB, and UvrC were found in the *A. actinomycetemcomitans* genome, it is apparent that the capability for this repair mechanism exists in *A. actinomycetemcomitans*. The first step is the association of two molecules of UvrA and one molecule of UvrB to form a complex that binds to DNA and slides along it using the helicase activity (possibly provided by UvrB) searching for damaged sites. $UvrA_2B$ dissociates leaving UvrB at the damaged site. UvrC binds to UvrB and excises a nucleotide fragment extending ~7 nucleotides from the 5' direction and ~4 nucleotide from the 3' direction. The excised nucleotide is removed by UvrD helicase generating a gap which is filled and sealed with DNA polymerase and DNA ligase, respectively (Figure 3.47).

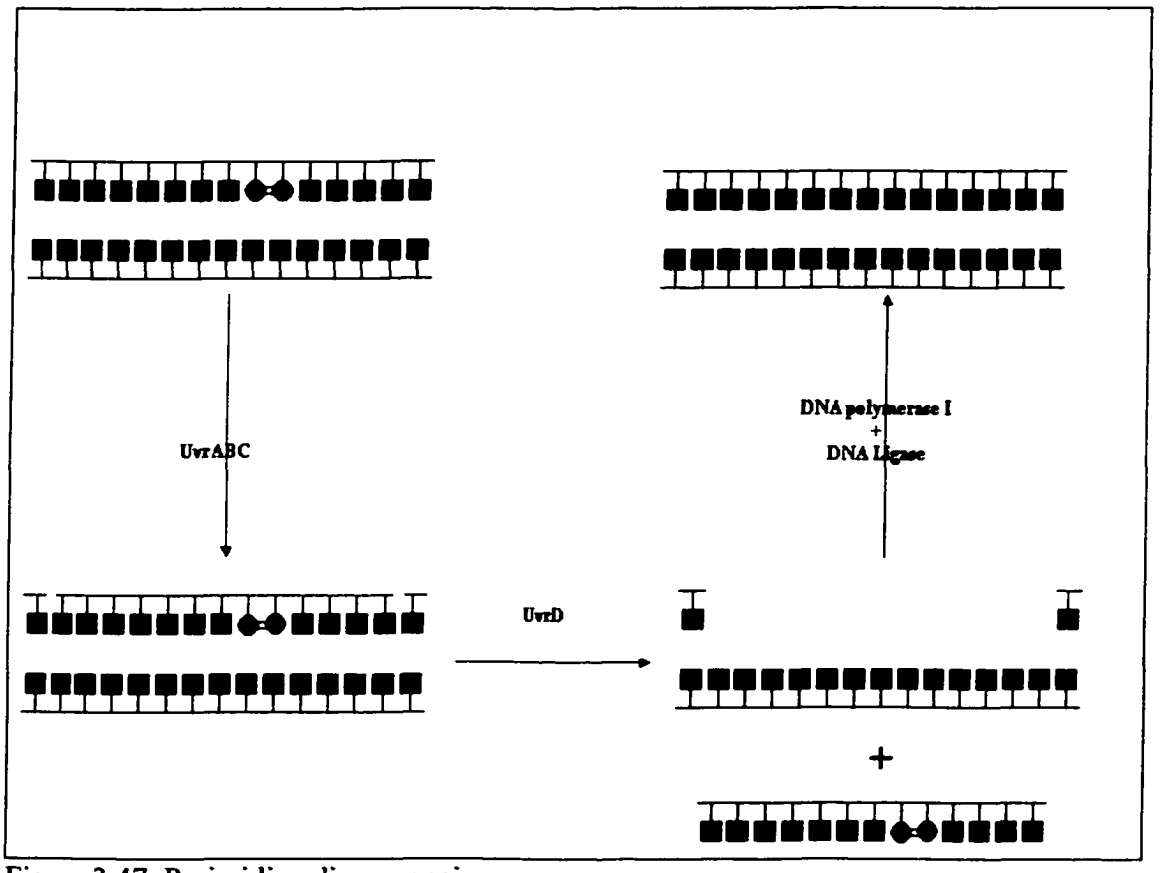


Figure 3.47. Pyrimidine dimer repair.

A second mechanism, the Mut system, for nucleotide-excision repair that also was reconstructed from the sequence data for *A. actinomycetemcomitans* is the long-patch methyl-directed mismatch repair which is associated with mismatched base pairing or damages that occur after replication but before the methylation of the newly synthesized strand (**Friedberg et al., 1995**). In *A. actinomycetemcomitans*, the two genes of the Mut system, *exoI* and *recJ*, both are encoded in the genome. This repair system relies on methylation of the original DNA strand by the restriction-modification system discussed earlier. Initially, the Mut system is responsible for the identification of the methylated strand and the site of damage and then exonuclease I or RecJ exonuclease create single

strand breaks on both sides of the damaged site. The incised oligonucleotide then is removed through UvrD, and DNA polymerase III fills the resulting gap generated (Figure 3.48)

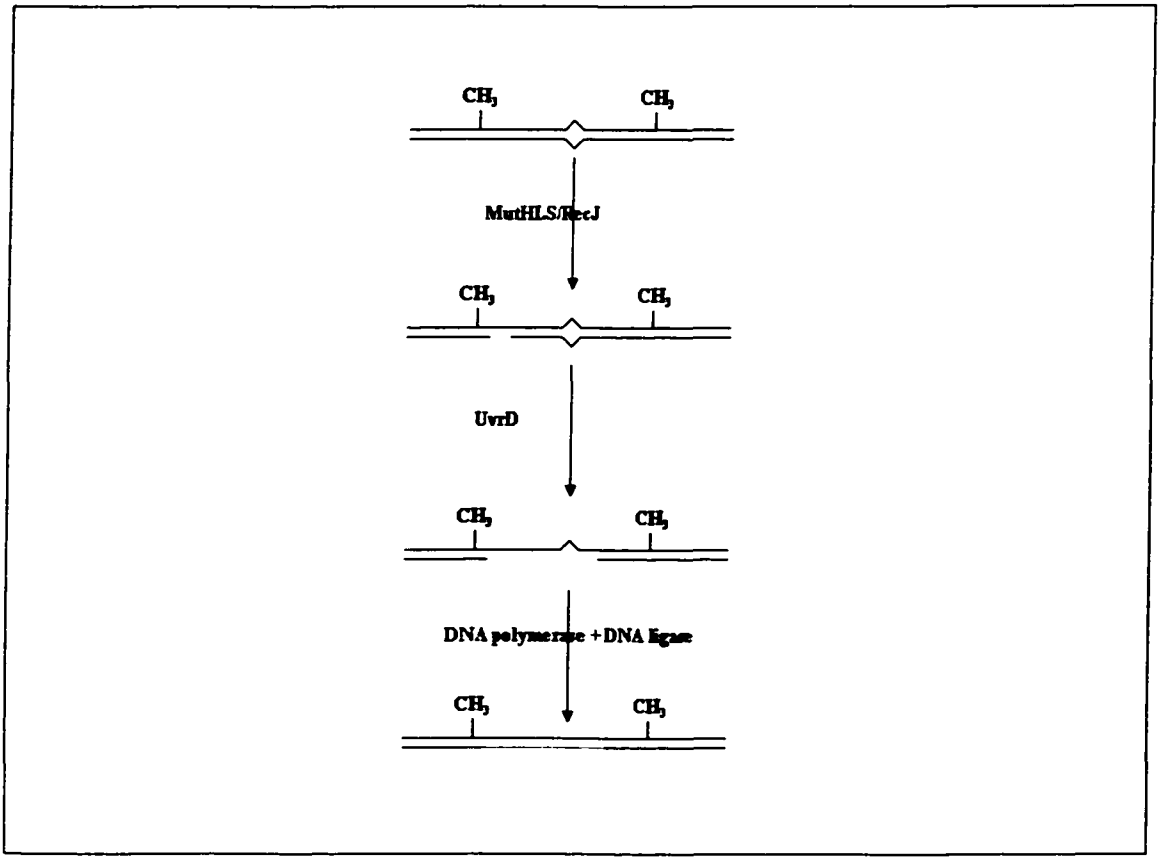


Figure 3.48. Long-patch methyl-directed mismatch repair.

Finally, *A. actinomycetemcomitans* appears to possess genes coding for the enzymes required for a third repair system that entails double-strand repair of DNA lesions that are caused by cross-linking agents. This type of repair, termed recombination repair, involves both Rec and Uvr proteins (**Friedberg et al., 1995**). As shown in figure 3.49, one of the damaged strands is cleaved and the 5'-3 exonucleotytic activity of DNA

polymerase I generates a gap that is used by RecA protein for homologous recombination. The homologous strand is extended by DNA polymerase I to fill the gap. Finally, an incision in the second strand generates another gap which is filled with DNA polymerase I.

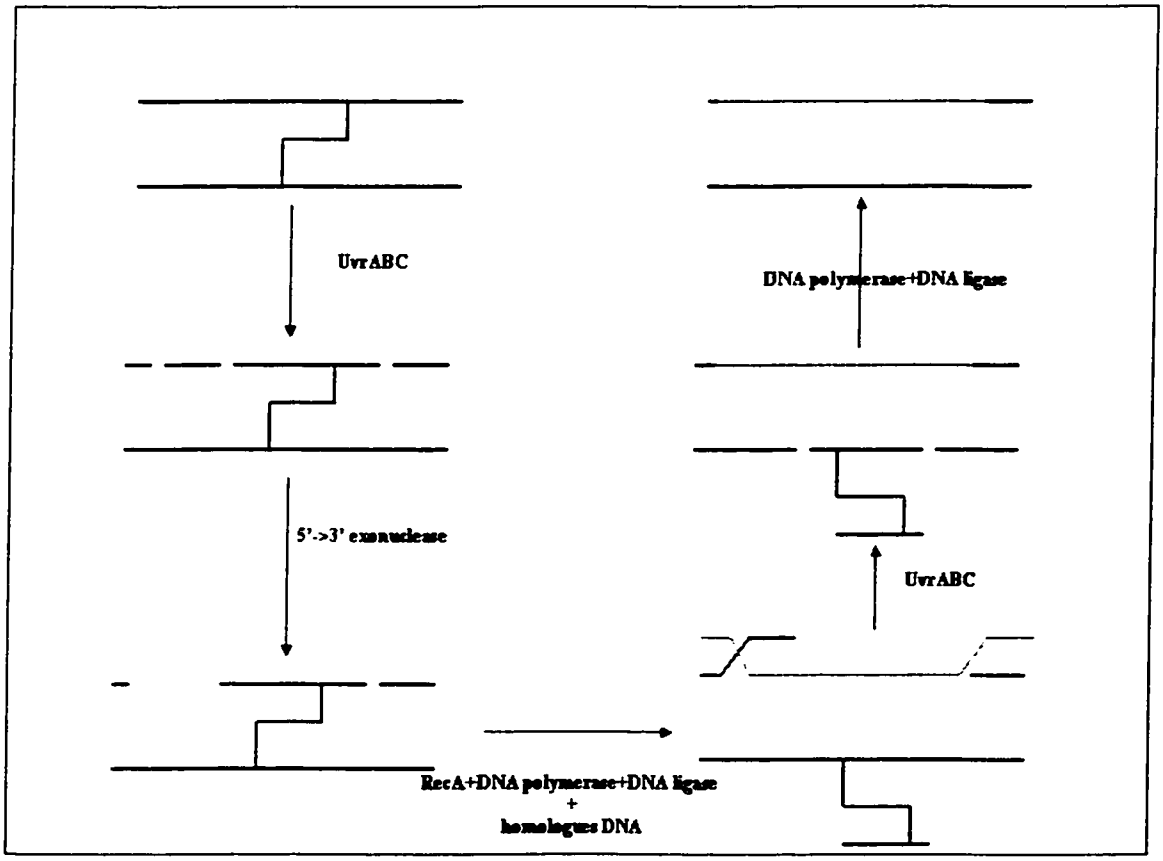


Figure 3.49. Long-patch methyl-directed mismatch repair.

EC#/gene name	Description	Genes in *A. actinomycetemcomitans*
2.1.1.63	6-O- methylguanine-DNA methyltransferase	Yes
4.2.99.18	Endonuclease III (DNA-(apurinic or apyrimidinic site) lyase); 8-oxoguanine DNA glycosylase (nth)	Yes
3.2.2.23	Formamidopyrimidine-DNA glycosylase (fpg)	Yes
3.2.2.-	uracil-DNA glycosylase (UDG) (ung)	Yes
3.1.11.1	Exodeoxyribonuclease I (exonuclease I) (DNA deoxyribophosphodiesterase) (DRPase) (sbcB)	Yes
uvrA	excinuclease ABC subunit A. Repair of UV damage to DNA (dinE)	Yes
uvrB	excinuclease ABC subunit B	Yes
uvrC	excinuclease ABC subunit C	Yes
uvrD	DNA helicase II	Yes
mutY	A/G-specific adenine glycosylase (mutY)	Yes
mutS	Methyl-directed mismatch repair	Yes
mutL	Methyl-directed mismatch repair	Yes
mutH	DNA mismatch repair protein	Yes
mutT	7,8-dihydro-8-oxoguanine-triphosphatase	Yes
recA	recombination protein, ATP-dependent coprotease	Yes
recX	Regulatory protein recX	Yes
recF	ssDNA and dsDNA binding protein	Yes
recG	DNA helicase, resolution of Holliday junctions	Yes
recN	Recombination and DNA repair	Yes
recO	Interaction with RecR and RecF	Yes
recJ	Single-stranded-DNA-specific exonuclease	Yes
recR	recombination protein recR	Yes

Table 3.26. Proteins involved in DNA repair.

3.3.6.2 RNA Metabolism

3.3.6.2.1 Transcription

Sequence analysis of the *A. actinomycetemcomitans* genome reveals that it contains a transcription machinery similar to that of *E. coli*. All of the genes for the proteins required for the transcription in *E. coli* are present in *A. actinomycetemcomitans*. In addition, there are two other sigma factors, the general sigma factor, sigma 70, which is needed for the transcription of most cellular genes, sigma 50, which is needed to

express the genes for nitrogen metabolism, and sigma 32, which is required for expressing heat-shock genes. Table 3.27 below lists transcription proteins encoded by the *A. actinomycetemcomitans* genome.

EC#/Gene name	Description	Genes in *A. actinomycetemcomitans*
rho	transcription termination factor	Yes
2.7.7.6	DNA-directed RNA polymerase alpha chain (rpoA)	Yes
2.7.7.6	DNA-directed RNA polymerase beta chain (rpoB)	Yes
2.7.7.6	DNA-directed RNA polymerase beta' chain (rpoC)	Yes
srmB	ATP-dependent RNA helicase (srmB)	Yes
sigG	RNA polymerase sigma factor	Yes
rpoD	RNA polymerase , sigma(70) factor; regulation of proteins induced at high temperatures	Yes
rpoE	RNA polymerase, sigma-E factor; heat shock and oxidative stress	Yes
hslU	Heat shock protein	Yes
rpoH	RNA polymerase, sigma(32) factor; regulation of proteins induced at high temperatures	Yes
rpoN	RNA polymerase, sigma(54 or 60) factor; nitrogen and fermentation regulation	Yes

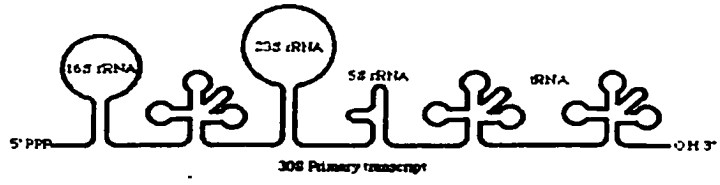
Table 3.27. Proteins involved in transcription encoded by the *A. actinomycetemcomitans* genome.

3.3.6.2.2 Processing of Ribosomal and Transfer RNAs

Generally, several ribosomal RNA operons are encoded in a bacterial genome. *E. coli* has seven independently transcribed ribosomal RNA operons that encode identical rRNAs but different tRNAs. Analysis of the *A. actinomycetemcomitans* genome reveals six independently transcribed rRNA operons encoding identical rRNAs but different tRNAs. As discussed earlier, due to their extreme similarities, assembly of rRNA operons is difficult at this stage. Figure 3.50 represents an attempt to decipher the rRNA operons that were assembled from shotgun sequence data. Initial PCR-based attempts to resolve the six rRNA operons in *A. actinomycetemcomitans* were unsuccessful.

Additional studies, such as long range PCR, will be needed to clarify the sequences of these six regions.

The genes encoding the two enzymes needed for processing the rRNA operon primary transcript, the ribonuclease III which cleaves 3' to the tRNA, and the ribonuclease P which cleaves 5' of the tRNA genes (Figure 3.51) also were present in the *A. actinomycetemcomitans* genome.



16S	tRNA	23S-5S	tRNA
	Ile Ala		
	Ile Ala		
	Glu Gly		
			Ser Asp Trp

Figure 3.50. Ribosomal RNA operons in *A. actinomycetemcomitans* genome. At least three different species were deduced from the sequence data, but additional long range PCR and DNA sequencing is needed to resolve these six regions.

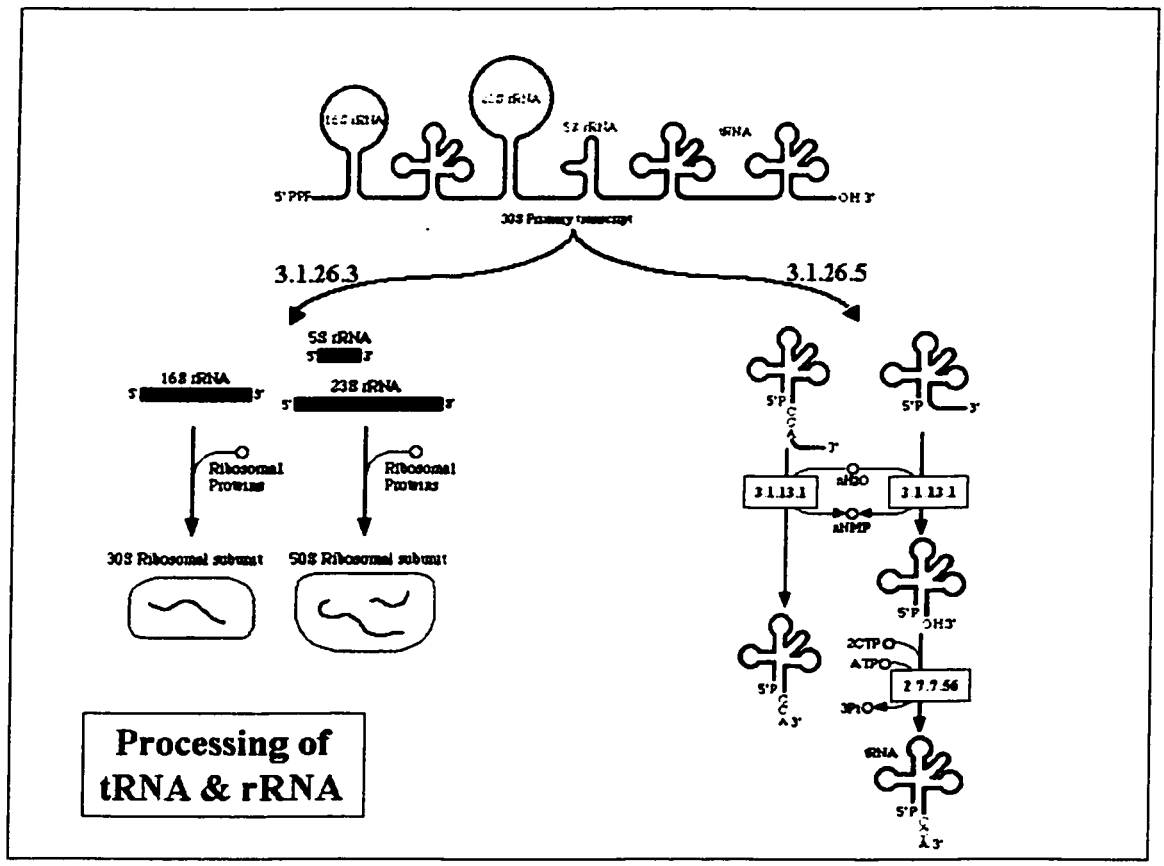


Figure 3.51. Schematic illustration of the processing steps of rRNA and tRNA. (adapted from KEGG).

As in *E. coli*, the ribosomes are composed of three rRNAs, 16S, 23S, and 5S, and more than 50 ribosomal proteins (**Keener et al., 1996**). Ribosomal proteins S1 to S21 makeup the 30S ribosomal subunit while ribosomal proteins L1 to L36 are components of the 50S ribosomal subunit. As shown in figure 3.51, the ribosomal operons contain three rRNA genes in the same order from the 5' end, 16S, 23S, and 5S. The region between the 16S and 23S rRNA genes and that 3' to the 5S rRNA contains tRNA gene(s) between the rRNA operons.

3.3.6.2.3 Post transcriptional modification of ribosomal and transfer RNA precursors to mature RNAs

The post transcriptional modification of RNA in bacteria requires a large number of enzymes (**Bjork et al., 1996 Bjork**). As in *E. coli*, and shown in *A. actinomycetemcomitans* in table 3.28, there are several rRNA modification enzymes , some of which modify functional sites in the rRNA. Very little is known about many of the enzymes involved in this process, however, in *E. coli*, it is known that the modification of 16S rRNA takes place before the assembly of the ribosomes. On the other hand, the modification of 23S rRNA is a late event that takes place after the assembly of the ribosomes. One gene that has been identified by mutagenesis studies in *E. coli* as an rRNA metylase is KsgA (S-Adenosylmethionine-6-N', N'-adenosyl (rRNA) dimethyltransferase) (**Van Buul et al., 1985**), and the gene for KsgA also is present in *A. actinomycetemcomitans*.

EC#/Gene name	Description	Genes in *A. actinomycetemcomitans*
2.1.1.31	tRNA (guanine-n1)-methyltransferase (trmD)	Yes
2.7.7.25	tRNA nucleotidyltransferase (cca)	Yes
4.2.1.70	Pseudouridylate synthase I	Yes
5.4.99.12	tRNA-pseudouridine synthase I (truA)	Yes
5.4.99.-	tRNA-pseudouridine synthase (truB)	Yes
2.5.1.8	tRNA delta(2)-isopentenylpyrophosphate transferase (miaA)	Yes
muvA	Uridine thiolation factor A activity	No
muvC	4-thiouridine modification of tRNA; near UV sensitivity and resistance	No
pheM	Phenylalanine tRNA synthetase leader peptide	No
5.-.-.-	S-adenosylmethionine:tRNA ribosyltransferase-isomerase (queA)	Yes
2.4.2.29	tRNA-guanine transglycosylase (tgt)	Yes
2.1.1.35	tRNA (uracil-5-)-methyltransferase (trmA)	Yes
trmB	tRNA (guanine-7-)-methyltransferase (tRNA(M-5-u54)-methyltransferase)	No
trmC	tRNA methytransferase; 5-methylaminoethyl-2-thiouridine biosynthesis	No
trmE	tRNA methytransferase; 5-methylaminoethyl-2-thiouridine biosynthesis (trmE)	Yes
trmF	tRNA methytransferase; 5-methylaminoethyl-2-thiouridine biosynthesis	Yes
ycfB	tRNA(5-methylaminomethyl-2-thiouridylate)-methyltransferase	Yes
ksgA	S-Adenosylmethionine-6-N', N'-adenosyl (rRNA) dimethyltransferase	Yes
ygcA	RNA methyltransferase	Yes

Table 3.28. RNA modification enzymes.

For tRNA modification, as in *E. coli*, *A. actinomycetemcomitans* encodes at least 12 tRNA modification enzymes (Table 3.28). Analysis of the *A. actinomycetemcomitans* genomic sequence also reveals that many of these tRNA modification proteins are encoded in operons with a gene order similar to that observed in *E. coli*, and *H. influenzae*. Figure 3.52 shows some of the studied operons in *E. coli* and the extent of similarities with *A. actinomycetemcomitans* and *H. influenzae*.

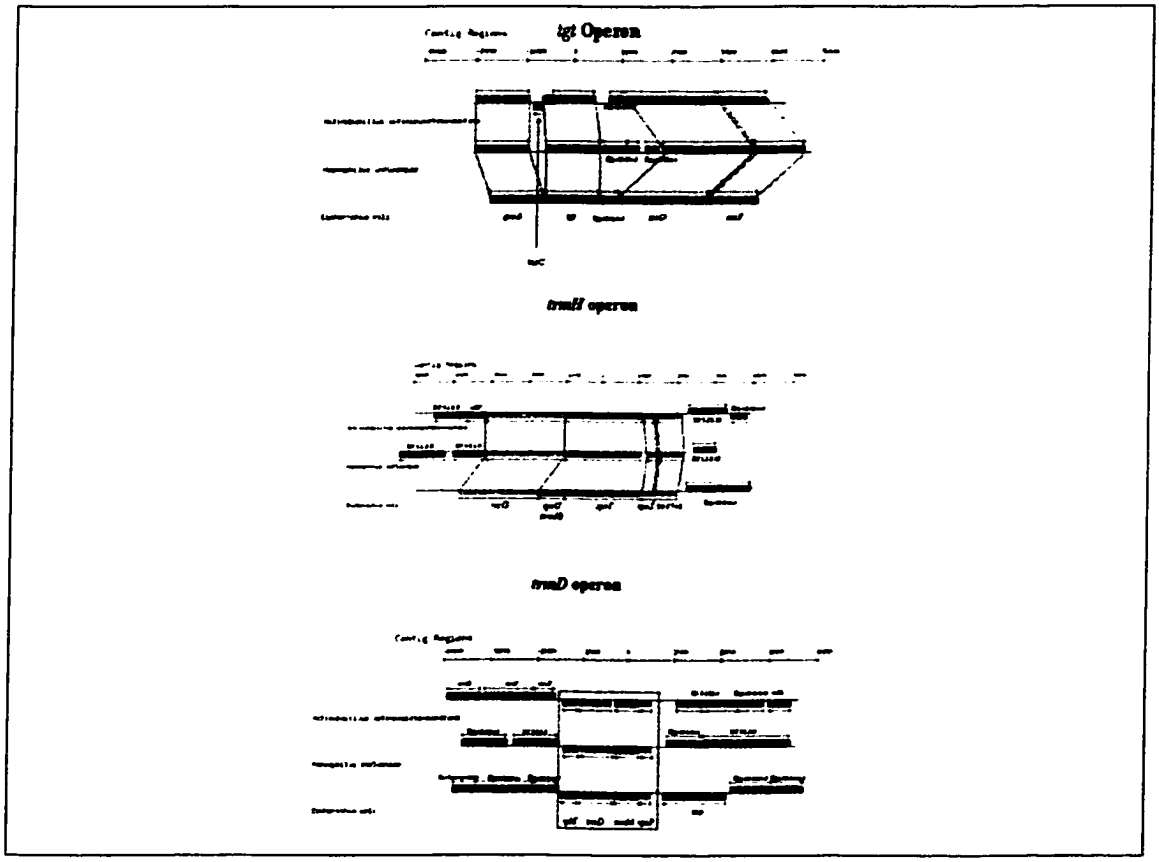


Figure 3.52. Comparison of the organization of RNA modification operons in *A. actinomycetemcomitans*, *H. influenzae*, and *E. coli*.

3.3.6.3 Aminoacyl-tRNA synthesis

Aminoacyl-tRNA synthetases (AARS) catalyze the esterification of an amino acid to its cognate tRNA (**Ibba et al., 1997**) (Figure 3.53). The genes for all twenty enzymes were found in *A. actinomycetemcomitans*. There historically have been two classes of aminoacyl-tRNA synthetases, class I and class II. Class I enzymes catalyze the transacylation via nucleophilic attack of the 2'-OH group of the terminal adenylate of the tRNA on the carbonyl of the amino acid which necessitates a transesterification to the 3'-OH of the terminal adenylate. Class II enzymes do not require the transesterification step

since the amino acid acylation initially occurs on the 3'OH group of the terminal adenylate residue directly (**Arnez et al., 1997**). The genes for both classes were observed in *A. actinomycetemcomitans* genome.

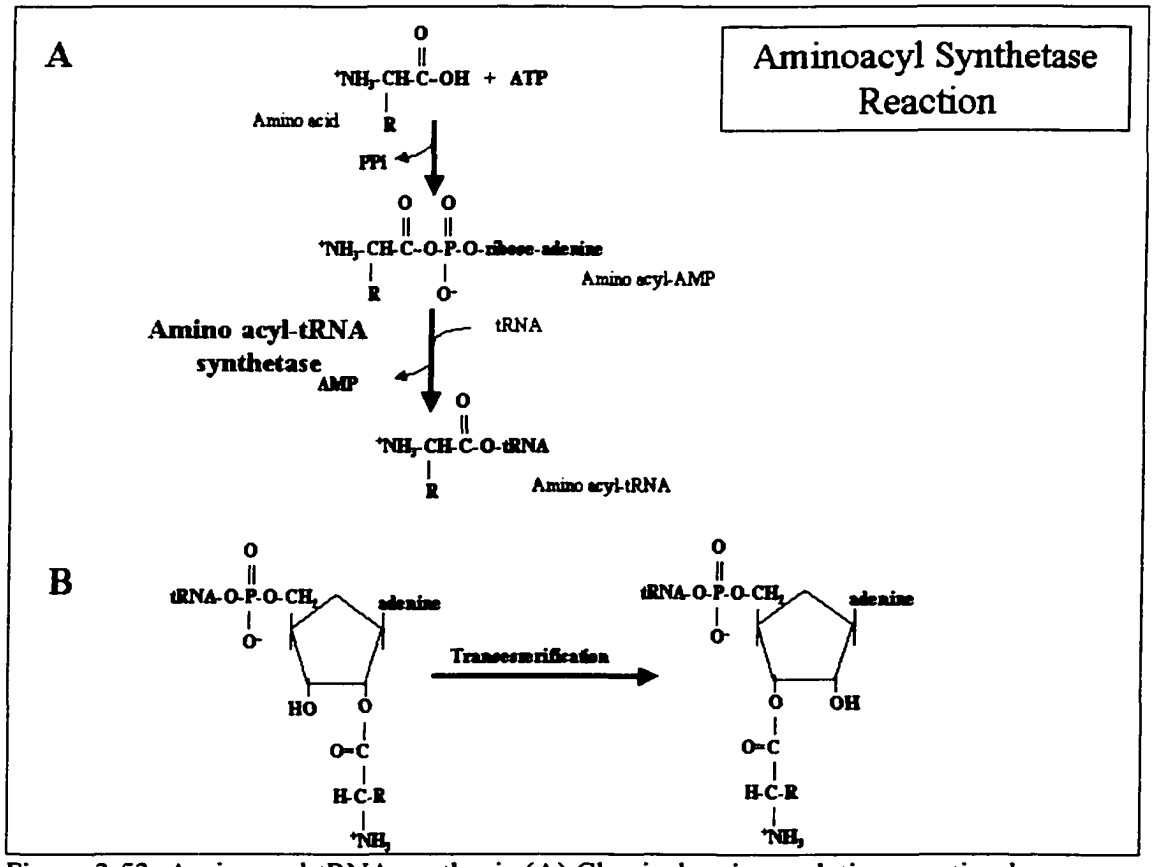


Figure 3.53. Aminoacyl-tRNA synthesis (A) Classical aminoacylation reaction by aminoacyl-tRNA synthetases. (B) The transesterification reaction required by Class I synthetases.

Initiation of protein synthesis in prokaryotes employs fmet- $tRNA^{fmet}$, where the esterified methionine is formylated. Initially, the methionine is esterified to $tRNA^{fmet}$ by the methionyl-tRNA synthetase and then the esterified methionine is formylated via the

enzyme methionyl-tRNA formyltransferase. The genes for both of these enzymes is encoded in the *A. actinomycetemcomitans* genome.

Selenocysteine is a metaloamino acid present in bacterial proteins. Interestingly, selenocysteine is produced by conversion of a seryl-tRNA^{Sec} to selenocysteinyl-tRNA^{Sec} as shown in figure 3.54 (**Bock et al., 1996**). All enzymes required for this pathway are present in *A. actinomycetemcomitans*.

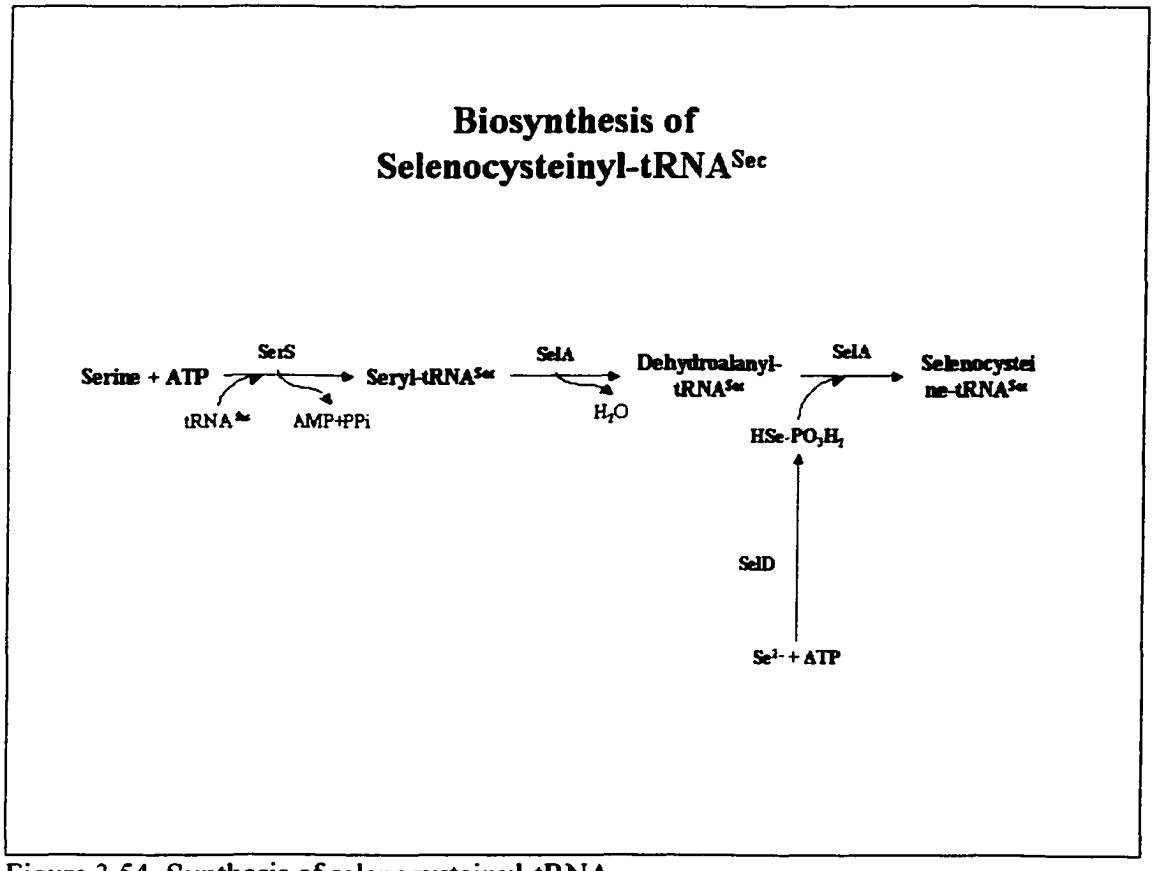


Figure 3.54. Synthesis of selenocysteinyl-tRNA.

3.3.6.4 Protein Synthesis (Translation)

The translation process in bacteria occurs in three stages, initiation, elongation, and termination.

3.3.6.4.1 Initiation.

The initiation complex is formed in the initiation phase, Here, the 30S ribosomal subunit is attached to mRNA and initiation factors 1, and 3 (IF1, and IF3). The binding of the 30S ribosomal subunit to the mRNA is facilitated by the presence of a short sequence upstream the start codon called the Shine-Delgarno sequence that is complimentary to a sequence near the 3' end of the 16S rRNA in the 30S ribosomal subunit (**Ringquist et al., 1992**). Then, formylmethionine-$tRNA^{fmet}$, in conjunction with an initiation factor 2 –GTP complex, is attached to what will become the ribosomal "P" site. Finally, the 50S ribosomal subunit binds to the complex displacing the initiation factors (Figure 3.55 and table 3.29). All genes coding for the enzymes required for initiation have been observed in *A. actinomycetemcomitans*.

3.3.6.4.2 Elongation

The elongation step starts when an aminoacyl-tRNA binds to the "A" site of the ribosome in the initiation complex. This requires elongation factor Tu-GTP to guide the aminoacyl-tRNA into the "A" site. Then, elongation factor Tu then is released with the hydrolysis of GTP to GDP and elongation factor Ts regenerates Tu-GDP to Tu-GTP. The peptide bond is formed when an electrophilic attack occurs through the free α-amino

group of the aminoacyl-tRNA in the "A" site by peptidyl transferase, which resides in the 50S ribosomal subunit. Finally, the ribosome translocates along the mRNA to the next codon accompanied by the hydrolysis of the GDP associated with elongation factor EF-G. This frees the "A" site for the next aminoacyl-tRNA to bind while the growing peptide chain is residing on the "P" site. The elongation cycle continues until the ribosome reaches a termination codon (**Laalami et al., 1996**) as shown in table 3.29, the genes for EF-Tu, and EF-Ts (*tufA*, and *tufB*) were observed in the *A. actinomycetemcomitans* genome.

3.3.6.4.3 Termination

When the ribosome reaches a stop codon (UAA, UAG, or UGA), release factors bind to the stop codon causing the dissociation of the ribosome and the release of the completed peptide from the tRNA. RF-1 binds to UAA and UAG, RF-2 binds to UAA, and UGA, RF-3 facilitates the activity of RF-1 and RF-2, and prevents the ribosomal subunits from re-associating before the next initiation stage (**Laalami et al., 1996**). All three release factors likely are present in *A. actinomycetemcomitans* as the genes encoding them were observed in the *A. actinomycetemcomitans* genomic sequence, as listed in table 3.29.

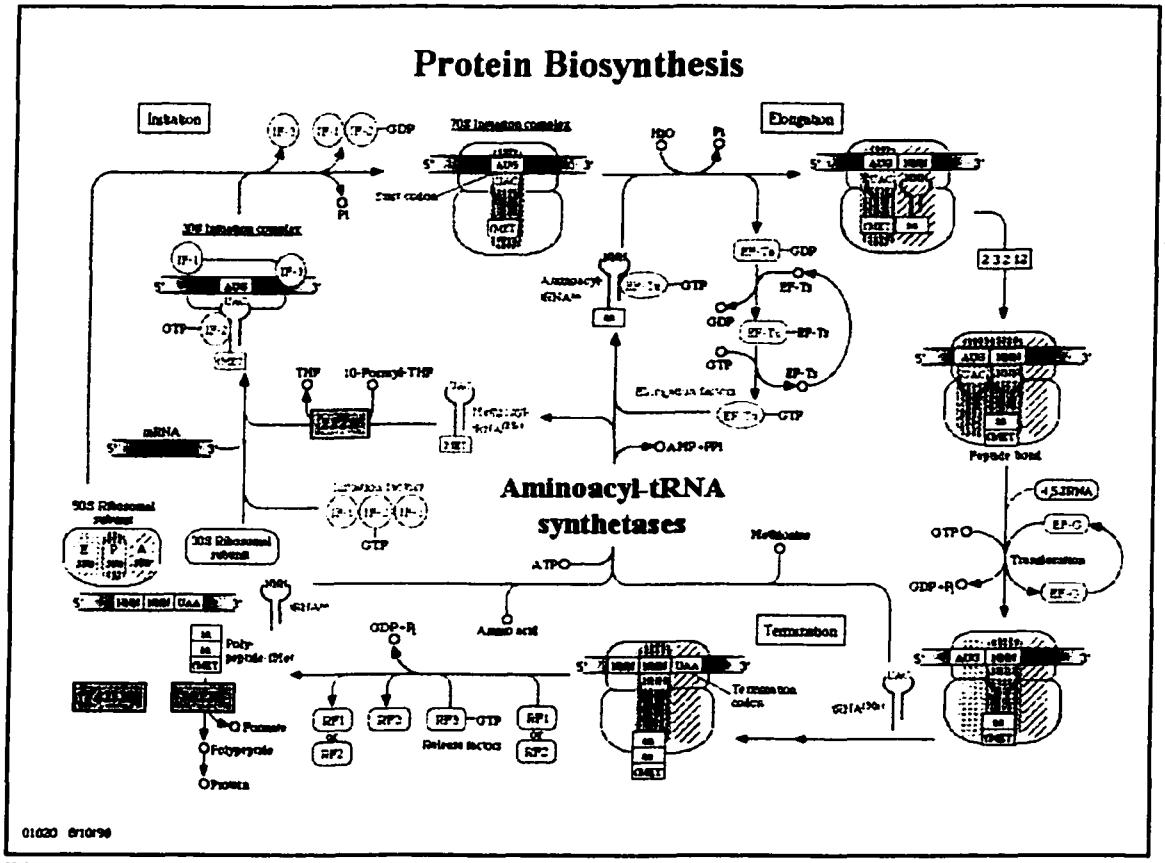


Figure 3.55. Schematic representation of the translation process as found in *A. actinomycetemcomitans*. (adapted from KEGG).

EC#/Gene name	Description	Genes in *A. actinomycetemcomitans*
3.5.1.31	Polypeptide deformylase (def)	Yes
2.1.2.9	Methionyl-tRNA formyltransferase (fmt)	Yes
2.1.2.9	Methionyl-tRNA formyltransferase (fmt)	Yes
infA	translation initiation factor IF-1	Yes
infB	translation initiation factor IF-2	Yes
infC	translation initiation factor IF-3	Yes
3.4.11.18	methionine aminopeptidase (MAP) (peptidase M) (map)	Yes
prfA	peptide chain release factor RF-1	Yes
prfB	peptide chain release factor RF-2	Yes
prfC	peptide chain release factor RF-3	Yes
tsf	elongation factor EF-Ts	Yes
tufA	elongation factor EF-Tu	Yes
tufB	elongation factor EF-Tu	Yes

Table 3.29. Proteins involves in protein synthesis in *A. actinomycetemcomitans*.

3.3.6.5 Protein Degradation

Hydrolysis of peptides to free amino acids is one of the central activities within a cell (**Miller, 1996**). The degradation of proteins is an important mechanism for regulating many pathways (**Miller, 1996**), for degrading miss-folded proteins, and for degrading proteins during starvation to provide amino acids for energy (**Miller, 1996**). Analysis of *A. actinomycetemcomitans* genome reveals the genes for several proteases which are listed in table 3.30.

EC#/Gene name	Description	Genes in *A. actinomycetemcomitans*
hycI	Hydrogenase3maturationprotease	Yes
3.4.21.53	ATP-dependentproteaseLa(lon,capR,deg,muc,lopA)	Yes
recA	Recombinationprotein,ATP-dependentcoprotease	Yes
3.4.21.92	ATPasesubunitofATP-dependentprotease(clpB)	Yes
3.4.21.-	ATP-dependentClpprotease,ATP-bindingsubunit(clpX)	Yes
3.4.99.-	Zincprotease(pqqL);(yddC)	Yes
3.4.21.-	Periplasmicserineprotease;heatshockprotein(htrA)	Yes
3.4.21.-	carboxyl-terminalproteaseforpenicillinbindingprotein(prc)	Yes
3.4.21.-	Putative protease (sohB)	Yes
3.4.24.13	ImmunoglobulinA1protease;IgA-	Yes
3.4.11.4	Tripeptideaminopeptidase (PeptidaseB)	Yes
2.3.2.2	gamma-glutamyltranspeptidaseprecursor(ggt)	Yes
3.4.11.-	Aminotripeptidase;peptidaseT(pepT)	Yes
3.4.11.18	methionineaminopeptidase(MAP)(peptidaseM)(map)	Yes
3.4.24.15	OligopeptidaseA;thimetoligopeptidase(prlC)	Yes
3.4.24.57	O-sialoglycoproteinendopeptidase(gcp)	Yes
3.4.11.1	AminopeptidaseA/I(pepA)(leucylaminopeptidase)(lap)	Yes
3.4.13.3	Aminoacyl-histidinedipeptidase (peptidaseD)	Yes
3.4.-.-	PeptidaseE (alpha-aspartyldipeptidase)	Yes
3.4.11.2	AminopeptidaseN (pepN)	Yes
3.4.11.9	aminopeptidasePII(pepP)	Yes
3.4.-.-	ProteaseIV,asignalpeptidepeptidase (sppA)	Yes
3.4.16.4	D-alanyl-D-alaninecarboxypeptidasefractionA	Yes
3.4.99.-	D-alanyl-D-alanine-endopeptidase)	Yes
3.4.23.36	Lipoproteinsignalpeptidase	Yes

Table 3.30. Proteases revealed from the *A. acinomystemcomitans* sequence data.

3.3.7 Cell Wall

3.3.7.1 Peptidoglycan

Peptidoglycan is a polymer chains containing alternating N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) moieties (**Van Heijenoort, 1996**) linked through their NAM residues via L-alanine – D-glutamate – L-amino acid – D-alanine tetra peptide units. The synthesis of peptidoglycan occurs in different stages. The first stage involves synthesis of the amino sugars NAM and NAG (**Van Heijenoort, 1996**). Then, D-alanine is synthesized via alanine racemase (5.1.1.1) and D-glutamate is synthsized by glutamate racemase (5.1.1.3) (Table 3.31).

EC#/Gene name	Description	Genes in *A. actinomycetemcomitans*
2.6.1.16	L-Glutamine:D-fructose-6-phosphate aminotransferase	Yes
2.3.1.4	Glucosamine-phosphate N-acetyltransferase	No
5.4.2.2	Phosphoglucomutase (Glucose phosphomutase)	Yes
msbB	Lipid A biosynthesis (kdo)2-(lauroyl)-lipid IVa acyltransferase	Yes
5.4.2.3	Phosphoacetylglucosamine mutase	Yes
2.7.7.23	N-acetylglucoseamine-1-phosphate uridyltransferase	Yes
2.5.1.7	UDP-N-acetylglucosamine enolpyruvyl transferase	Yes
1.1.1.158	UDP-N-acetylenolpyruvoylglucosamine reductase	Yes
murC	UDP-N-acetylmuramate--alanine ligase	Yes
murD	UDP-N-acetylmuramoylalanine--D-glutamate ligase	Yes
murF	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl- D-alanyl ligase	Yes
murE	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	Yes
5.1.1.1	Alanine racemase	Yes
5.1.1.3	glutamate racemase	Yes
6.3.2.4	D-alanine--D-alanine ligase	Yes

Table 3.31. Enzymes involved in the first step of peptidoglycan synthesis.

The second stage involves the transfer of the amino sugars through the inner membrane to the periplasm (Table 3.32). This is accomplished through the lipid carrier, undecaprenyl-phosphate, which is a C_{50} isoprenoid phosphate (**Rogers et al., 1980**). This

lipid can bind and transport different cell wall intermediates such as LPS, and teichoic acid to the site of peptidoglycan synthesis (**Rogers et al., 1980**). The genes for all of the enzymes involved in this process are present in *A. actinomycetemcomitans* genome except undecaprenyl-diphosphatase. Interestingly, this gene, which encodes a protein responsible for recycling undecaprenyl-pyrophosphate to undecaprenyl-phosphate, also is not present in many other Gram-negative bacteria including *H. influenzae*, *E. coli*, and *H. pylori* as described in KEGG..

EC#	Description	Genes in *A. actinomycetemcomitans*
2.7.8.13	phospho-N-acetylmuramoyl-pentapeptide-transferase	Yes
2.4.1.-	UDP-N-acetylglucosamine--N-acetylmuramyl (murG)	Yes
3.6.1.27	Undecaprenyl-diphosphatase	No
tgtase	Monofunctional biosyntheticpeptidoglycan transglycosylase (monofunctional tgase); gene name from *B. subtilis*	Yes
3.4.16.4	Penicillin binding proteins	Yes

Table 3.32. Enzymes involved in stages 2 and 3 of peptidoglycan synthesis.

The third stage of peptidoglycan synthesis is the polymerization of peptidoglycan, a process which takes place on the periplasmic side of the inner membrane in Gram-negative organisms. All of the enzymes involved in this stage were found in *A. actinomycetemcomitans* including the transpeptidases required for cross-linking the peptidoglycans which also are termed penicillin-binding proteins. Figure 3.56 illustrates peptidoglycan synthesis as reconstructed in *A. actinomycetemcomitans*.

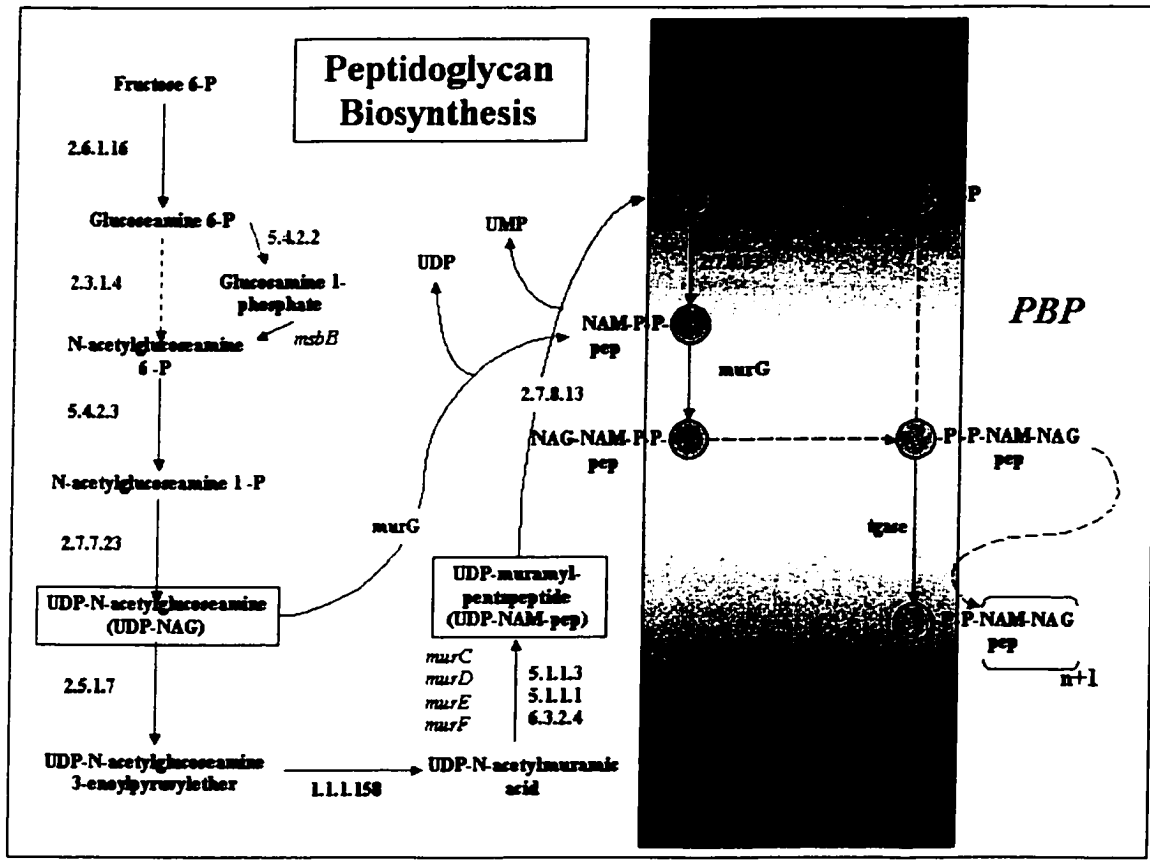


Figure 3.56. Peptidoglycan biosynthesis. The circle represent undecaprenyl moiety of undecaprenyl phosphate. Dotted arrows represent enzyme missing from *A. actinomycetemcomitans* genome.

3.3.7.2 Lipopolysaccharides

Gram-negative bacteria differ from Gram-positive bacteria by the presence of lipopolysaccharides (LPS), large amphophilic lipids that lie on the outer membrane of the bacteria (**Heinrichs et al. 1998**). LPS functions as a barrier against heavy metals and lytic enzymes produced by the host or competitive organism in the same environment. An LPS contain three components. The first is lipid A, a hydrophobic lipid which anchors LPS in the outer membrane. Lipid A typically consists of β-hydroxymyristic acid (C_{14}) linked to a backbone of two glucoseamine residues via an ester bond through

the hydroxyl groups and amide bond to the nitrogen of the glucoseamine to two of the β-hydroxymyristic acid, lauric acid (C_{12}), and myristic acid (C_{14}).

The second component in LPS is the core which is connected to lipid A through 3-hydroxy-D-manno-octulosonate (KDO). This part is further divided into two regions, the inner region which consists of KDO, heptose and phosphate, and an outer region which consists of hexoses. The hexoses are added through specified glycosyl transferase enzymes for each hexose from UDP derivatives. These enzymes are membrane-bound. The inner region biosynthesis is not yet understood.

The third component, O-antigen, consists of repeating units of four to six sugars that can be repeated up to 30 times depending on the bacterial species.

Synthesis of lipid A starts from UDP-NAG described above in figure 3.56. The remainder of the synthesis process occurs as illustrated in figure 3.57 and listed in table 3.33 (**Raetz et al., 1987**). Here, the core and lipid A are translocated to the periplasmic surface of the inner membrane. O-antigen is synthesized in a similar manner to peptidoglycan. The sugar residues are synthesized as units of sugars that are assembled on an undecaprenoyl carrier. It has been suggested that the final assembly of the completed LPS takes place in the periplasm where the entire LPS is translocated to the outer membrane (**Heinrichs et al., 1998**). The translocation is postulated to take place through undecaprenol-phosphate that serves as a transporter of the LPS to the cell membrane. There is evidence that LPS actually is translocated to the cell membrane via an ABC transporter (**Raetz, 1996**). Although the genes for these specific ABC transporters have not yet been found in *A. actinomycetemcomitans genome* sequence, as

in most bacteria in the *Pasteurellaceae* family, the *A. actinomycetemcomitans* genome does contain the genes for the proteins of the entire LPS synthesis pathway.

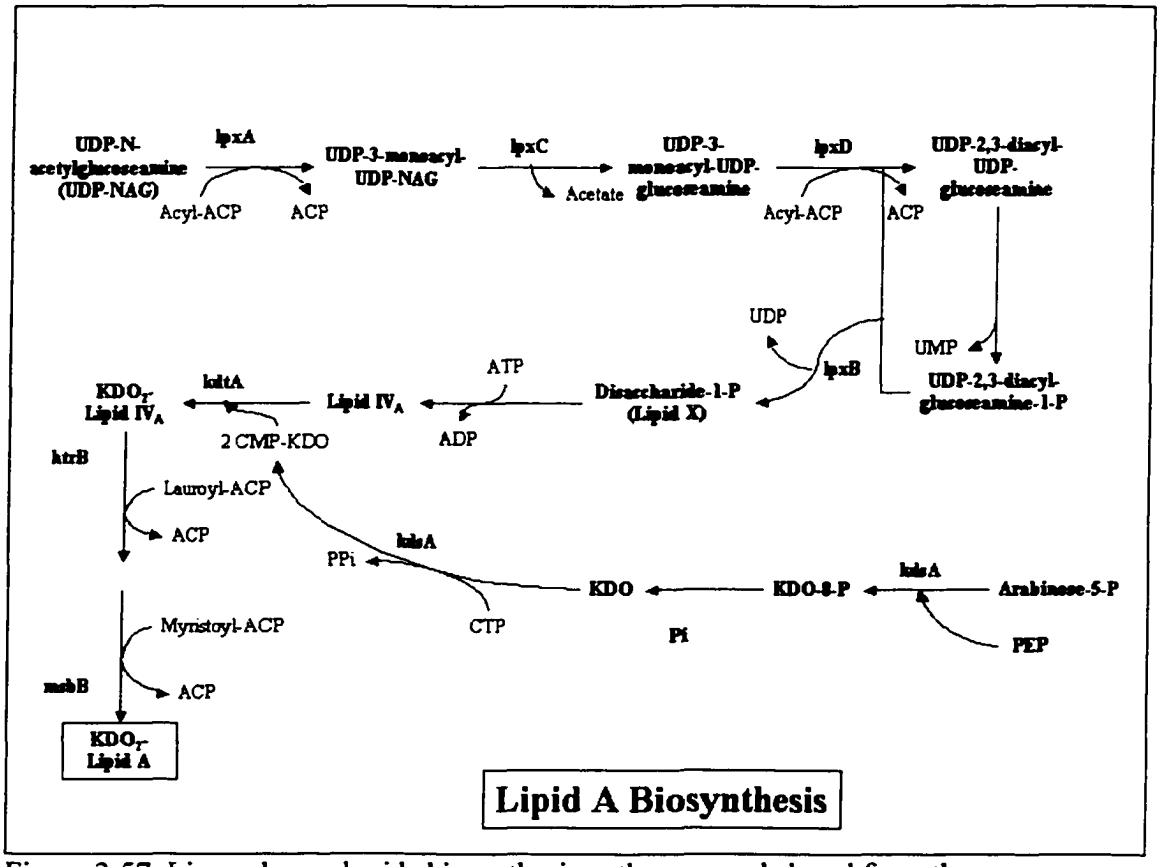


Figure 3.57. Lipopolysaccharide biosynthesis pathways as deduced from the sequence analysis of *A. actinomycetemcomitans*.

EC#	Description	Genes in *A. actinomycetemcomitans*
2.3.1.129	UDP-N-acetylglucosamine acyltransferase (lpxA)	Yes
3.5.1.-	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase (lpxC)	Yes
2.3.1.-	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase (lpxD)	Yes
2.4.1.182	Lipid-A-disaccharide synthase (lpxB)	Yes
2.4.99.-	3-deoxy-D-manno-octulosonic-acid transferase (KDO transferase) (kdtA)	
2.3.1.-	Lipid A biosynthesis lauroyl acyltransferase;lauroylacyltransferase (htrB)	Yes
2.3.1.-	Lipid A biosynthesis (kdo)2-(lauroyl)-lipid IVa acyltransferase (msbB)	Yes
2.7.7.38	3-deoxy-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase) (kdsB)	Yes
4.1.2.16	3-deoxy-D-manno-octulosonic acid synthetase (KDO-8-PHOSPHATE SYNTHETASE) (kdsA)	Yes

Table 3.33. Enzymes involved in Lipid A and KDO biosynthesis.

3.3.8 Transport proteins

Transport proteins are responsible for the passage of substances, from small atoms to large molecules, accross the cell membrane. Therefore, transport proteins can compensate for the incomplete metabolic pathways by providing the cell with needed precursors for these reactions. The comprehensive list of all annotated *A. actinomycetemcomitans* genes given in appendex A contains a listing of the putative transporter genes which can be classified into the functional groups discussed below, some of which are listed, in table 3.34.

Gene name	Description	Genes in *A. actinomycetemcomitans*
artI	arginine transport system periplasmic binding protein	Yes
artM	arginine transport system permease protein	Yes
artP	Arginine transport system	Yes
artQ	Arginine ABC transport system	Yes
brnQ	branched chain transport system	Yes
gltS	Sodium--glutamate symport carrier protein	Yes
potA	Spermidine/putrrscine transport protein	Yes
potB	Spermidine/putrrscine transport protein	Yes
potC	Spermidine/putrrscine transport protein	Yes
potD	Spermidine/putrrscine transport protein	Yes
sdaC	Probable ser transport	Yes
tnaB	Low-affinity trp permease	Yes
tyrP	Tyr-specific transport system	Yes
PfeA	Ferric enterobactin receptor	Yes
corA	Mg++ transport system	Yes
exbB	uptake of enterochelin	Yes
exbD	uptake of enterochelin	Yes
fecB	Citrate-dependent Fe transport, periplasmic	Yes
fecC	Citrate-dependent Fe (III) transport, cytosolic	Yes
fecD	Citrate-dependent Fe transport, membrane-bound protein	Yes
fecE	Citrate-dependent Fe (III) transport, membrane bound	Yes
nhaB	Na+/H antiporter, pH independent	Yes
nhaC	Na+/H antiporter, pH independent	Yes
nikA	periplasmic binding protein for Ni	Yes
panF	Na/pantothenate symporter	Yes
putP	Major Na/pro symporter	Yes
rsgA	Ferritin-like protein	Yes
trkA	Transport of K	Yes
trkH	K uptake	Yes
yfeA	Mn++ transport protein; iron (chelated) ABC transporter	Yes
3.6.1.-	K+/Cu++ transporting ATPase (copA)	Yes
snf	Sodium-dependent transporter	Yes
znuA	zinc ABC transporter	Yes
znuB	zinc ABC transporter	Yes
Na_Sulfate	putative Na_Sulfate (or dicarboxilate) transporter	Yes
cysZ	Cysteine synthetas; required for sulfate transport	Yes
modA	Molybdate uptake (ABC)	Yes
modE	Molybdenum transport protein	Yes
modB	Molybdate uptake	Yes
modC	Molybdate uptake (ABC) (chlD)	Yes
molB	Molybdenum transport protein	Yes
pitA	low-affinity inorganic phosphate transporter 1 (pit)	Yes
eriC	putative chloride-channel protein	Yes

Table 3.34. Transporters found in *A. actinomycetemcomitans* genome.

3.3.8.1 Cation and Anion transporters

Analysis of the *A. actinomycetemcomitans* genomic sequence reveals the presence of the genes for several ion transporters including those for potassium, sodium, magnesium, molybdate and zinc as listed in table 3.35 below.

EC#/Gene name	Description	Genes in *A. actinomycetemcomitans*
corA	Mg++ transport system	Yes
nhaB	Na+/H antiporter, pH independent	Yes
nhaC	Na+/H antiporter, pH independent	Yes
nikA	periplasmic binding protein for Ni	Yes
panF	Na/pantothenate symporter	Yes
putP	Major Na/pro symporter	Yes
trkA	Transport of K	Yes
trkH	K uptake	Yes
yfeA	Mn++ transport protein; iron (chelated)	Yes
3.6.1.-	K+/Cu++ transporting ATPase (copA)	Yes
snf	Sodium-dependent transporter	Yes
znuA	zinc ABC transporter	Yes
znuB	zinc ABC transporter	Yes
Na_Sulfate	putative Na_Sulfate (or dicarboxilate) transporter	Yes
modA	Molybdate uptake (ABC)	Yes
modE	Molybdenum transport protein	Yes
modB	Molybdate uptake	Yes
modC	Molybdate uptake (ABC) (chlD)	Yes
molB	Molybdenum transport protein	Yes
pitA	low-affinity inorganic phosphate transporter	Yes
eriC	putative chloride-channel protein	Yes

Table 3.35. Cation and anion transporters found in *A. actinomycetemcomitans* genome.

3.3.8.2 Phosphotransferase system (PTS)

The PTS system usually is dedicated to carbohydrate transport and is commonly found in facultative and obligate anaerobic bacteria (**Postma et al., 1996**). This system transports carbohydrates into the cell as phosphorylated derivatives rather than free sugars. The phosphate donor is phosphoenol pyruvate (PEP) (**Postma et al., 1996**). There are different configurations of the PTS system but they are all quite similar no

matter which carbohydrate is being translocated (**Postma et al., 1996**). The PTS is composed of three different enzymes. Enzyme I (EI), catalyzes the phosphate transfer from PEP to the enzyme to generate phosphorylated enzyme I (EI-P). Then EI-P transfers the phosphate to a protein called HPr (Histidine protein). Finally, HPr-P phosphorylates enzyme II (EII) which in turn phosphorylates the carbohydrate as it is transported into the cell (**Postma et al., 1996**). EII has three subunits, A, B, and C. HPr phosphorylates subunite A which in turn, phosphorylates B. Subunit B, then phosphorylates the carbohydrate as subunit C translocates it into the cell. The difference between the systems lies mostly in the third enzyme of the system (EII) (**Postma et al., 1996**) as it can have different subunit arrangements. For example, although it can be composed of three domains that are membrane bound, it also may be composed of membrane-bound components B and C, and a soluble A component (**White, 2000**). Additionally, A and B can be soluble while C forms a membrane-bound channel with yet another protein, designated D. Analysis of the *A. actinomycetemcomitans* genome reveals the genes required to code for the proteins of the PTS transport systems listed in table 3.36 and illustrated in figure 3.58.

Transporter	Description	Specificity
PtsG	pts system, glucose-specific IIBC component (EIIBC-Glc) (ptsG)	Glucose, Sorbose, glucose amine
crr	PTS system, Glucose phosphotransferase enzyme III glucose-specific IIA component (EIIA-Glc) (crr, gsr, iex, tgs)	Can be the "A" component for many PTS proteins in the Glucose family
fruB	pts system, fructose-specific iia/fpr component (EIIA-fru)	Fructose, Xylitol, Sorbose, Glucose, Mannose
fruA	pts system, fructose-specific IIBC component (EIIBC-fru)	
manX	pts system, mannose-specific IIab component (EIIBC -man) (mannose-permease iiab component)	Mannose, Trehalose, Glucose, Glucoseamine, N-acetylglucose amine
manY	pts system, mannose-specific IIC component (EIIC-Man) (mannose- permease IIC component) (phosphotransferase enzyme II, C component) (ptsP, pel, manY)	
manZ	pts system, mannose-specific IID component (EIID-Man) (mannose- permease IID component)(ptsM, gptB)	

Table 3.36. The PTS system in *A. actinomycetemcomitans*.

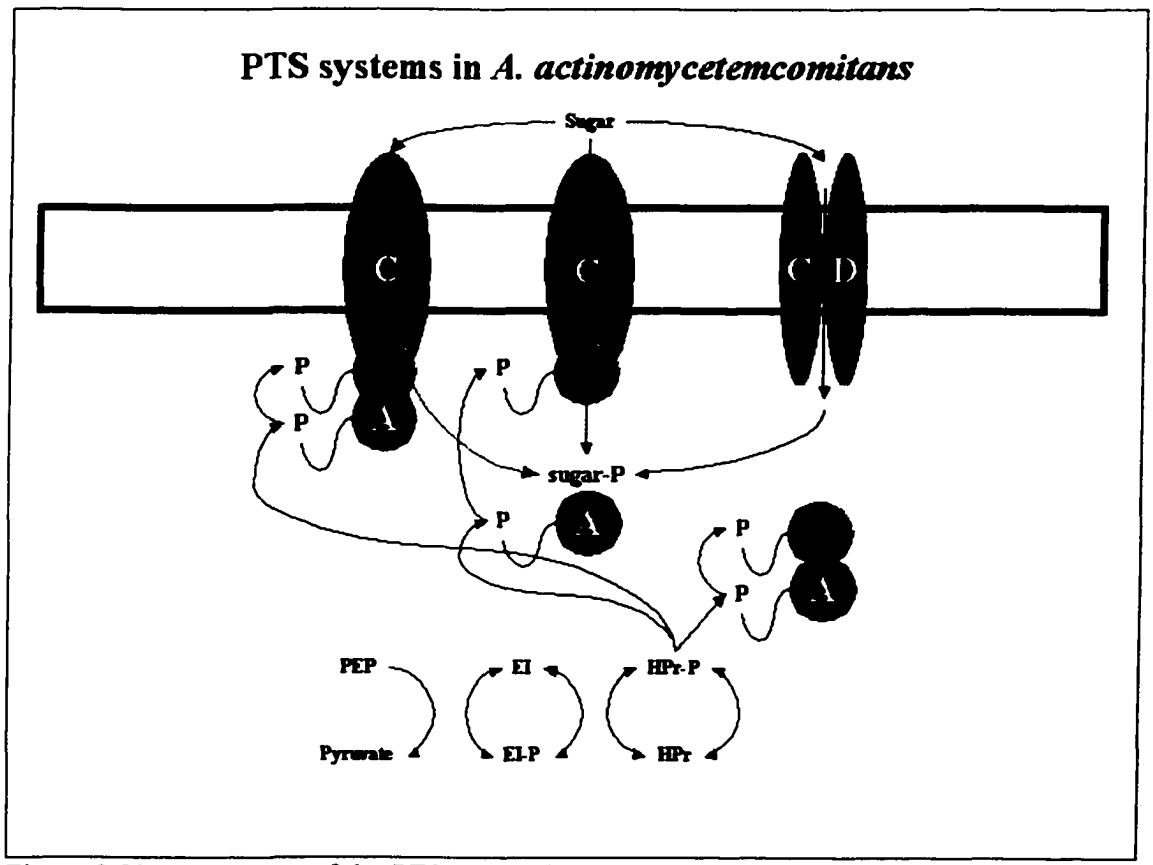


Figure 3.58. Summaery of the PTS system in *A. actinomycetemcomitans*. Note the different arrangements of EIII.

3.3.9 Protein Export and Secretion

One of the common features in bacteria is the export of proteins from "their site of synthesis to other destinations inside or outside the cell" (**Tjalsma et al. 2000**).

As with *E. coli* and other Gram-negative bacteria, the *A. actinomycetemcomitans* genome contains the genes for the general secretion system, the Sec pathway, which is shown in figure 3.59 and reviewd by White and Murphy et al. (**White et al., 2000 and Murphy et al., 1996**), where the ffh protein might play a role in recognizing the signal sequence. SecB then binds to the internal domain of the preprotein and transfers it to SecA which, in turn, delivers the preprotein to the SecY/G/E membrane translocase complex. SecA protein is an ATPase that utilizes ATP hydrolysis to translocate the preprotein through the membrane. It is hypothesized that the signal sequence loops inside the membrane through its h-region while the n-region remains bound to the phodpholipids in the inner leaflet of the inner membrane. SecA then spools the protein through the channel formed by SecY/E. ATP hydrolysis is required for the initiation of the translocation process. The remainder of the protein then translocates across the membrane via proton motif force. SecY appears to be an integral membrane protein with 10 transmembrane spanning domains. It is believed that SecY forms a large pore through the membrane to accommodate the translocating preprotein. SecG and SecE both are part of the translocase machinery and they appear to be membrane proteins as well. Mutations in SecE severely affects the translocation, while SecG mutations decreases the efficiency of export. SecD and SecF function is not clear but they both appear to have soluble domains in the periplasm. It is speculated that these two proteins may play a role in the maintaining the electrochemical gradient required for the translocation process.

This pathway appears to be complete in *A. actinomycetemcomitans* as the genes encoding all the proteins involved were found in the genome. The recent observation that the insertion of secreted proteins into the membrane required YidC proein, as mutations in this protein decreases the efficieny of transporting protein (**Samuelson et al., 2000**), prompted the search for its gene, and the discovery of a YidC gene homolog in the *A. actinomycetemcomitans* genome.

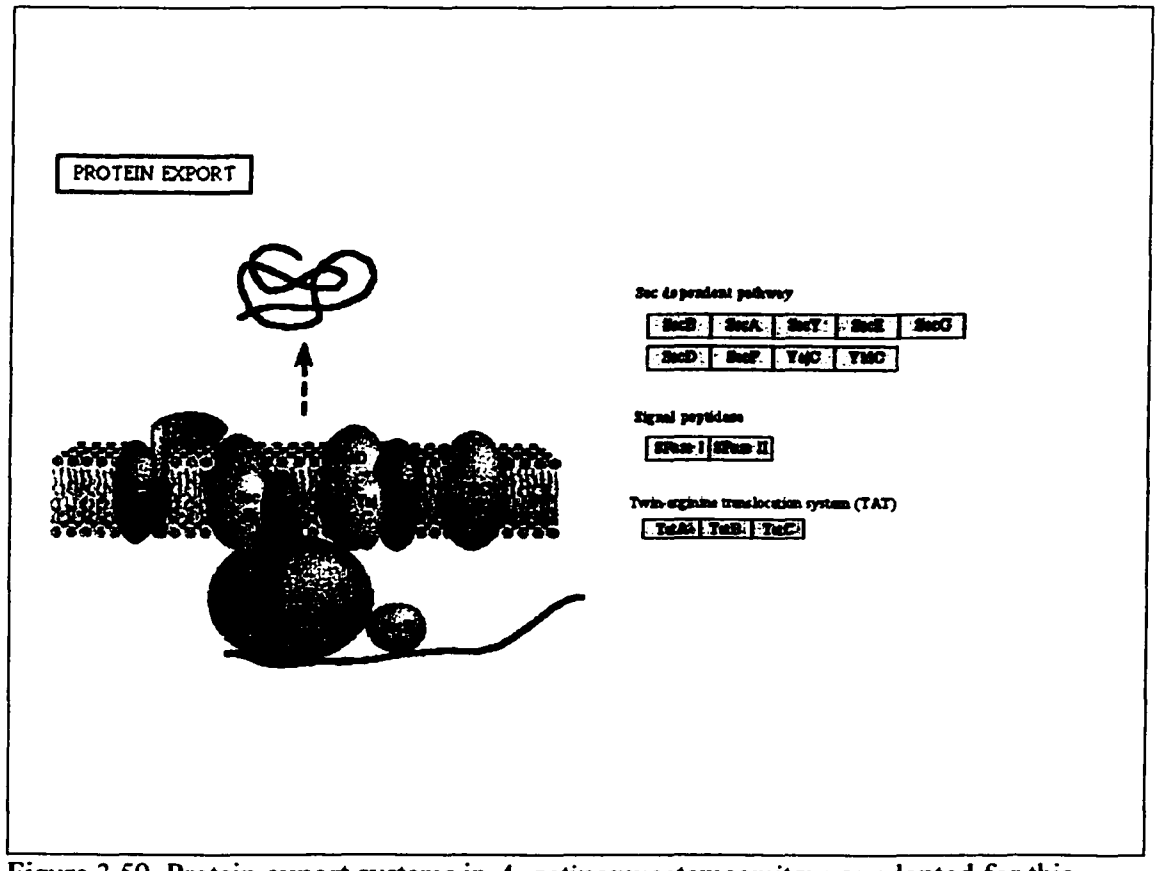


Figure 3.59. Protein export systems in *A. actinomycetemcomitans* as adapted for this organism using KEGG's reconstruction option as discussed in Materials and Methods. Later the diagrams were refined manually to signify the systems as present in *A. actinomycetemcomitans*.

Analysis of the *A. actinomycetemcomitans* genomic sequence also reveals the presence of the genes encoding the twin arginine translocation (TAT) system that was predicted by Berks in 1996 (**Berks, 1996**) and discovered by Bogsch and colleagues in 1998 (**Bogsch et al., 1998**). Interestingly, the TAT system is homologous to the ΔpH protein transport system in chloroplasts (**Berks et al., 2000**), and it is believed that this transport system is responsible for whole protein transport across the membrane (**Berks et al., 2000**). Unlike the Sec transport system, where proteins are prevented from folding until this tranlsocation takes place, this system allows the cofactor to be inserted into the protein and the final 3-D confirmation to be established prior to the translocation. TAT-exported proteins contain a signal sequence with a structure similar to those of Sec-exported proteins. The n-region of the signal peptide in TAT-exported proteins is much longer than that of Sec. The h-region is shorter in Tat, and the c-region contains charged amino acids as opposed to polar amino acids in the system. The most important characteristic of the TAT-dependent signal peptide is the presence of two consecutive arginine residues in the junction between the n- and the h- regions (**Berks et al., 2000**). In *E. coli*, the consensus sequence of signal is (S/T)-R-R-x-F-L-K where S/T is either serine or threonine, and x is any polar amino acid (**Voordouw, 2000**). The signal peptide in TAT-exported proteins is 26-60 amino acids long as opposed to the Sec-exported proteins signal sequence which is 26 amino acids. Not much is known about the mechanism or the organization of the proteins involved. However, it is believed that the energy source for this system is the electrochemical gradient across the membrane (**Berks et al., 2000**). The sequence data suggests that all components of the TAT system exist in *A. actinomycetemcomitans* and figure 3.60 illustrates the TAT system components.

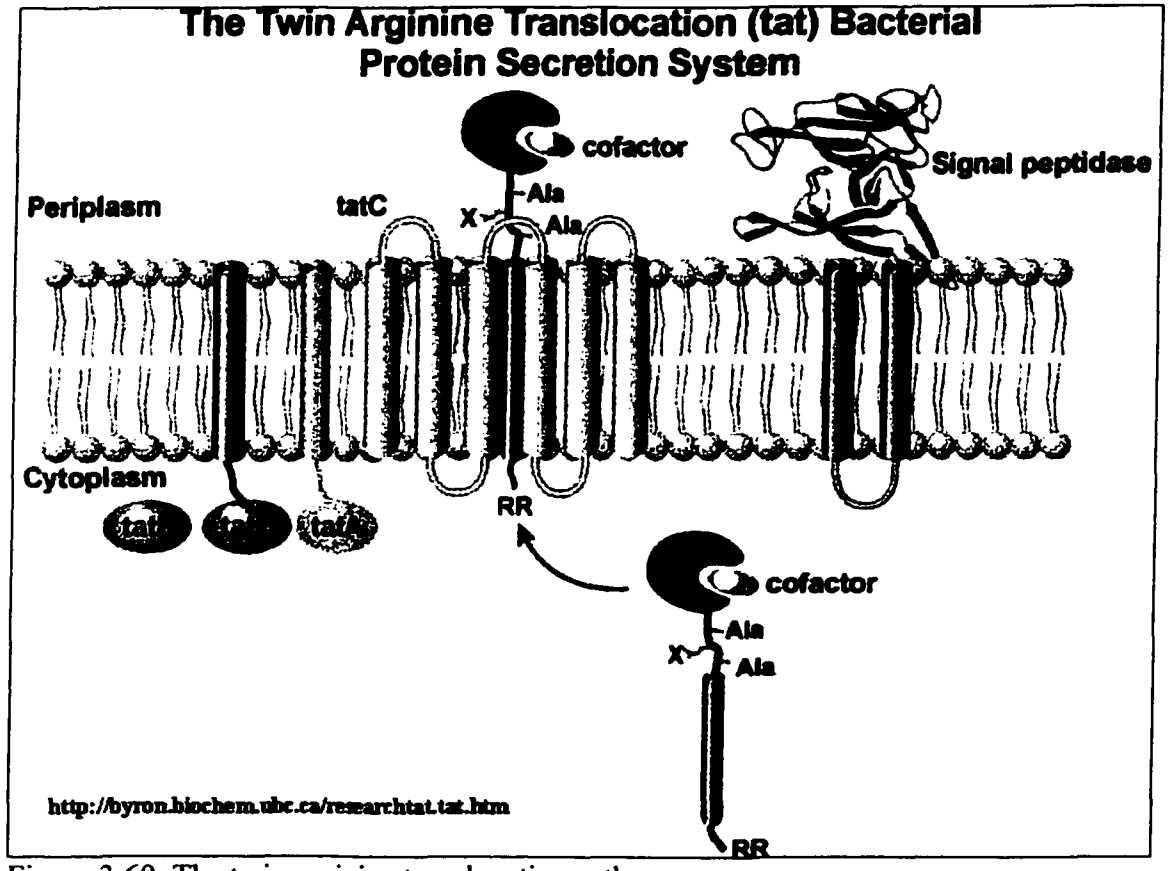


Figure 3.60. The twin arginine translocation pathway.

3.4 Virulence Factors

Analysis of the sequence of *A. actinomycetemcoitans* genome revealed the presence of numerous putative virulence factors, many of which have been observed in other pathogens (Table 1.2). These predicted virulence factors included proteins involved in iron acquisition, adhesion, invasion, and mechanisms to evade the host immune response. Virulence factors typically are involved in enabling and maintaining the infection process of a pathogenic organism interacting with its host

3.4.1 Iron acquisition and utilization

A. actinomycetemcomitans contains the genes for the hemoglobin and haptoglobin-hemoglobin binding proteins, and the ferrated transferrin binding protein (**Schryvers et al., 1999**). Hemoglobin is rarely free in the body as haptoglobin rapidly binds to any free circulating hemoglobin (**Schryvers et al., 1999**). Any heme which is released from free hemoglobin is bound by either hemopexin or albumin, and cleared from the circulation. Although *A. actinomycetemcomitans* appears not to need heme (as opposed to *H. influenzae*), since its genomic sequence indicates that it possesses the genes for all the enzymes necessary for synthesizing heme as shown in figure 3.61 and listed in table 3.37, *A. actinomycetemcomitans* still requires an iron source. Hemoglobin is one of the sources available to *A. actinomycetemcomitans*, and analysis of the genomic sequence reveals three hemoglobin-haptoglobin binding proteins homologous to the *hgp*ABC proteins in *H. influenzae* (**Morton et al., 1999**). The proteins encoded by these genes have been shown to bind the hemoglobin-haptoglobin complex in *H. influenzae* and appear to be regulated by a slip-strand synthesis mechanism as observed and suggested from the CCAA tetranucleotide repeat found after the signal peptide sequence of each of the three open reading frames (**Jinet et al., 1996**). It is suggested that the CCAA repeat is responsible for phase variation in *H. influenzae* through a slip-strand synthesis mechanism (**Ren et al., 1999**). Interestingly, the CCAA repeat could not be detected in the *hgpABC* homologues in *A. actinomycetemcomitans* sugesting that this operon likely is not regulated in a similar manner.

Gram-negative bacteria also acquire iron from the host through transferrin binding protein, a TfpA homologue. In *A. actinomycetemcomitans*, as in *Neisseria gonorrheae*,

TfpA and TfpB proteins which facilitate transferrin binding to TfpA, are encoded in a single operon, which also is present in *A. actinomycetemcomitans*. The *A. actinomycetemcomitans* genome also encodes the *afuABC* iron uptake system, which is homologous to the *H. influenzae hit*ABC and *Neisseria gonorrhoeae fbp*ABC systems (**Willemsen et al., 1997**). Chin and colleagues (**Chin et al. 1996**) first discovered the *afuABC* locus in *A. pleuropneumoniae* and later, Graber and colleagues (**Graber et al. 1998**) detected this activity in *A. actinomycetemcomitans*. The *afuABC* system is encoded by three genes assembled into an operon and expressed as one transcriptional unit (**Willemsen et al., 1997**). In *N. gonorrhoeae*, this system has the ability to transport iron to the cytoplasm through the *fbp*A gene product, a periplasmic protein with a permease motif, the *fbp*B gene product, a hydrophobic protein which is located in the cytoplasmic membrane, and the *fbp*C gene product, a hydrophilic protein with a nucleotide binding motif, that is peripheral to the cytoplasm (**Adhikari et al., 1996**).

In addition to the above, *A. actinomycetemcomitans* also encodes a homolog to the *yfe*ACD system of *Yersinia pestis* (**Bearden et al., 1998**). In *Y. pestis*, the *yfe*ACD system is encoded by two operons, *yfe*A-D, and *yfe*E. The *yfe*A-D operon encodes YfeA, a periplasmic protein, YfeB, an ABC transporter, and both YfeC and YfeD which are integral membrane proteins. The *yfe*E operon, which is transcribed in opposite orientation than the *yfe*A-D operon, encodes for a cell envelope protein (**Bearden et al., 1998**). Analysis of the *A. actinomycetemcomitans* genomic sequence reveals the presence of a *yfe* iron uptake system similar in organization to that observed in *Y. pestis*.

Additionally, an enterobactin binding protein, a BfeA homologue also observed in the *A. actinomycetemcomitans* genome. This gene product has been studied by Dean and colleagues in *P. aerogenosa* (**Dean et al., 1993**) and from *B. pertussis* by Beall and colleagues (**Beall et al., 1995**). Since *A. actinomycetemcomitans* also encodes a BfeA homolog, it too can scavenge iron from other organisms which synthesize siderophores and can co-infect the oral cavity. For example, *P. aerogenosa* is an oral pathogen, that synthesizes two iron-chelating agents, pyoverdin and pyochelin, which might be scavenged by *A. actinomycetemcomitans*.

Finally, a Fec ferric citrate transport system homologue also was detected in the sequence of *A. actinomycetemcomitans*. This system was discovered in *E. coli* (**Wagegg et al, 1981, and Enz et al, 2000**). It transports ferric citrate by binding it from the environment using FecA protein that then releases it to FecB protein in the periplasm. Subsequently, ferric citrate is transported through the cytoplasmic membrane through FecC, D, and E, all of which are present in *A. actinomycetemcomitans* genomic sequence.

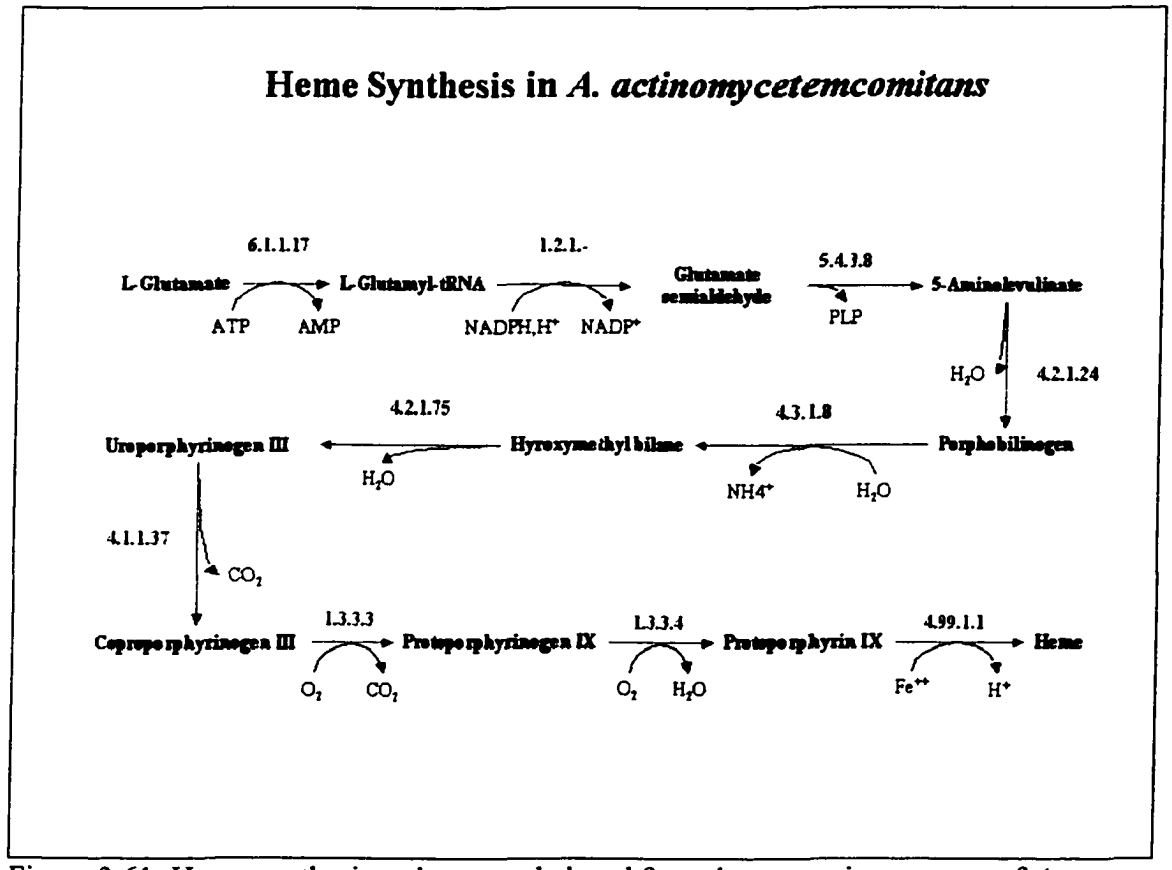


Figure 3.61. Heme synthesis pathway as deduced from the genomic sequence of *A. actinomycetemcomitans*.

EC#	Description	Genes in *A. actinomycetemcomitans*
6.1.1.17	Glutamyl-tRNA synthetase (gltX)	Yes
1.2.1.-	glutamyl-tRNA reductase (hemA)	Yes
5.4.3.8	glutamate-1-semialdehyde 2,1-aminomutase (hemL)	Yes
4.2.1.24	porphobilinogen synthase (emB)	Yes
4.3.1.8	porphobilinogen deaminase (hemC)	Yes
4.2.1.75	uroporphyrinogen-III synthase (hemD)	Yes
4.1.1.37	uroporphyrinogen decarboxylase (UPD) (hemE)	Yes
1.3.3.3	coproporphyrinogen III oxidase (hemF)	Yes
1.3.3.4	protoporphyrinogen oxidase (hemG)	Yes
4.99.1.1	ferrochelatase (hemE synthetase) (visA)	Yes

Table 3.37. Enzymes involved in the heme synthesis pathway in *A. actinomycetemcomitans*.

3.4.2 Lipopolysacharides and Phosphorylcholine

As mentioned above, Gram-negative bacteria contain LPSs which function as a barrier against heavy metals and the lytic enzymes produced by the host or a competitive organism in the same environment. As such, LPSs are considered to be components of a primitive bacterial defense system. They likely evolved long before mammalian hosts evolved any defense systems (**Rietschel et al. 1994**). Since *A. actinomycetemcomitans* encodes the complete pathway for LPS biosynthesis (Figure 3.57, and table 3.33), it is likely that these lipids contain a highly variable carbohydrate moiety (O-antigen) that is an antigenic determinant presented to the host during infection (**Moxon et al.. 1998**). The high variability of the O-antigen is one of the key factors that results in a successful evasion of the host immune system of the host by the pathogen (**Moxon et al. 1998**). LPS therefore was considered as one of the first suspected virulence factors in Gram-negative bacteria (**Moxon et al.. 1998**). Since LPS an integral bacterial component, it is considered an endotoxin as opposed to an exotoxin that is secreted (**Rietschel et al. 1994**). It is well documented that injecting animals with LPS extracts from pathogenic bacteria can induce the symptoms of the disease fulfilling the fifth postulate of Koch (**Rietschel et al. 1994**). Therefore, it is likely that the LPS in *A. actinomycetemcomitans* play roles in its ability to evade the host's immune system, its toxicity, and its iron acquisition ability by a mechanism similar to that in other organisms (**Turner et al. 1998**).

Phosphoryl choline is another relatively recently discovered virulence factor which decorates the LPS in *Streptococcus pneumoniae* and other bacteria (**Gillespie et**

al. 1993). Since *A. actinomycetemcomitans* can import choline (**Schenkein et al. 2000**), this precursor for phosphorylcholine likely is present in this bacterium. Although the exact function of phosphorylcholine-decorated LPS is unknown, it is thought to be involved in the host invasion process in most species (**Weiser et al. 1998**) and has been implicated in maintaining a persistent infection of *H. influenzae* in the respiratory tract (**Weiser et al. 1998**). A similar observation was noted earlier for *A. actinomycetemcomitans* (**Zhang et al. 1999**), and phosphorylcholine (PC) also was found to enhance its invasiveness. It recently has been reported that *A. actinomycetemcomitas* might utilize platelet-activating factor receptor in a manner similar to *S. pneumoniae* (**Schenkein et al. 2000**), prompting these authors to hypothesize that *A. actinomycetemcomitas* gains entry to the circulatory system through the invasion of endothelial cells and transmigration.

3.4.3 Adhesion factors

For a pathogen to establish disease symptoms on a site, it has to adhere and colonize that site. Although primary colonization factors such as LPS play a role in initial attachment to the site, eventually, the pathogen must establish a tight adhesion to the host cell to exert its influence on the site of infection. In the case of *A. actinomycetemcomitans*, LPS is one of the initial factors involved in adhesion.

Recently, an operon that contains seven genes (*tadA-G*) has been characterized and found to be essential for tight nonspecific adhesion (**Kachlany et al. 2000**). These genes were discovered from transposon mutants that were smooth in appearance and that

could not aggregate in liquid culture. Under the microscope, these mutants did not display the fibril bundles seen on the wild type cells (**Kachlany et al. 2000**). Earlier, two genes linked to the rough colony morphology, *rcpA* and *rcpB* were discovered in *A. actinomycetemcomitans* and subsequently cloned and sequenced (**Haase et al. 1999**). In addition, the Flp protein, which was discovered earlier in *A. actinomycetemcomitans* (**Inoue et al. 1998**), was suggested to be the major fimbriae protein. All components of this locus are required to complement spontaneous smooth strains of *A. actinomycetemcomitans* (**Kachlany et al. 2000**). Analysis of the *A. actinomycetemcomitas* genomic sequence reveals that all of the above adhesion factor encoding genes are located in the same region of the genome as illustrated in figure 3.62 below. Since the GC% of the entire area is lower than the average GC% of the genome (37% versus 42.5%), it may be that these genes were introduced into the genome by horizontal transfer. This hypothesis is supported by the observation that pathogens encoding this operon (*tadA-G*) have a similar differences in the GC% compared to the GC% of the resident genomes (**Kachlany et al. 2000**).

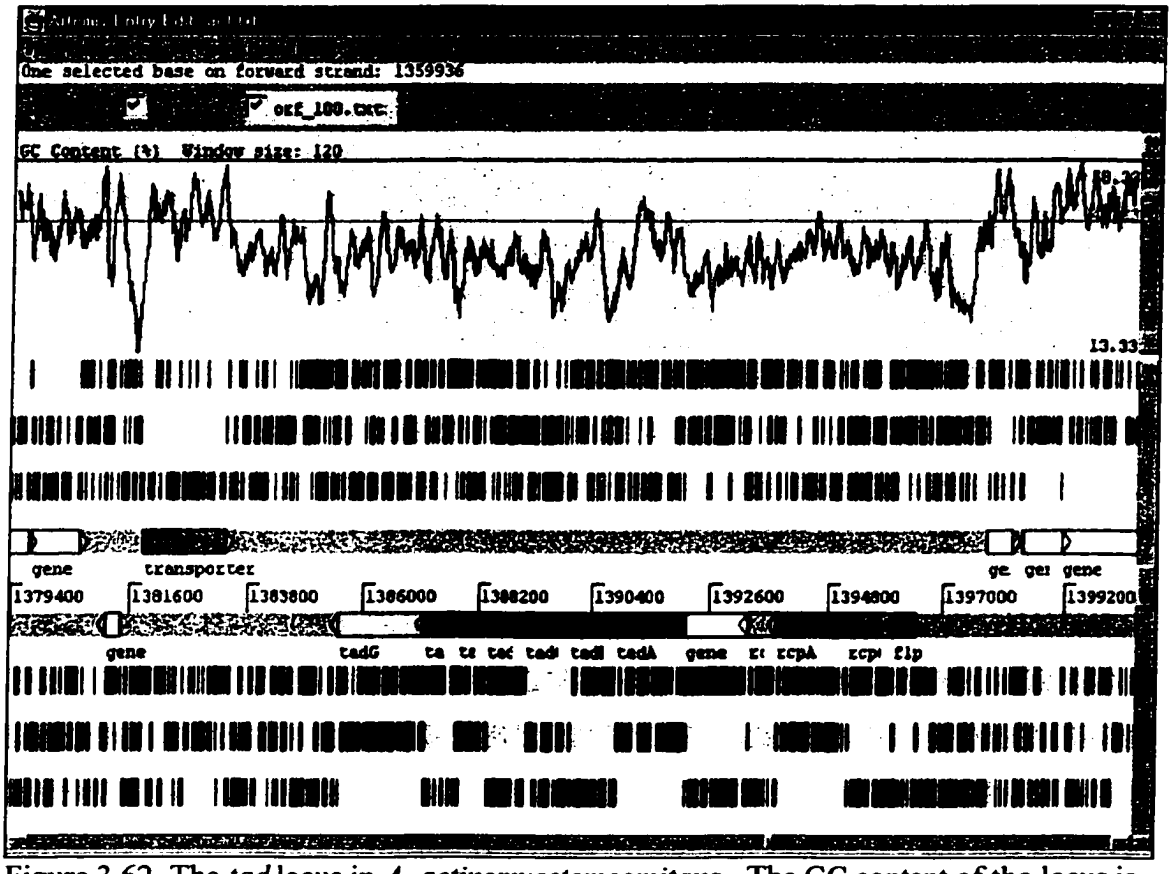


Figure 3.62. The *tad* locus in *A. actinomycetemcomitans*. The GC content of the locus is below the average which suggests a horizontal transfer.

Another known adhesion specific gene product, Hsf, also is encoded in the *A. actinomycetemcomitans* genome. The *H. influenzae* Hsf protein seems to be responsible for specific adhesion to epithelial cells (**St.Geme III et al. 1996**) and hence, Hsf likely has a similar function in *A. actinomycetemcomitans*, as well as in *E. coli*, and *S. typhimurium* where it also has been found.

Another possible adhesion factor is the YadA homolog, which has similarities to UspA1 (Ubiquitous Surface Protein A1) from *Moraxella sp.*, and a putative adhesin factor from *N. meningitidis*. In *Yersinia pestis*, YadA protein, a plasmid-encoded protein,

is optimally expressed at 37^0C and polymerizes to cover the outer membrane with fibrillar matrix (**Kapperud, et al. 1987**). This increases the negative surface charge presumably facilitates binding to extracellular matrix such as collagen and fibronectin (**Lachica et al. 1984, Tamm et al. 1993, Tertti et al. 1992, and Schulze et al. 1992**). YadA also acts in association with Yop proteins in protecting the phagocytosed *Yersinia sp.* from antimicrobial agents and lysozymes produced by granulocytes (**Visser et al. 1996**). In addition, YadA in combination with the Yop proteins induces apoptosis (**Monack et al. 1997**). Figure 3.63 below shows a Blast2 pairwise alignment (**Thompson et al. 1994**) of YadA from *Y. entercolitica* and its homolog from *A. actinomycetemcomitans*. This alignment suggests that the N-terminal of *A. actinomycetemcomitans* YadA homolog is truncated while the C-terminal is conserved. Since residues 29-81 in YadA from *Yersinia* are responsible for surface polymerization and binding to extracellular matrix proteins while the C-terminal is responsible for serum resistance and agglutination (**Roggenkamp et al. 1996**), it is likely that the YadA homolog in *A. actinomycetemcomitans* is not multifunctional and hence might not be involved in adhesion, at least in the same fashion, as YadA.

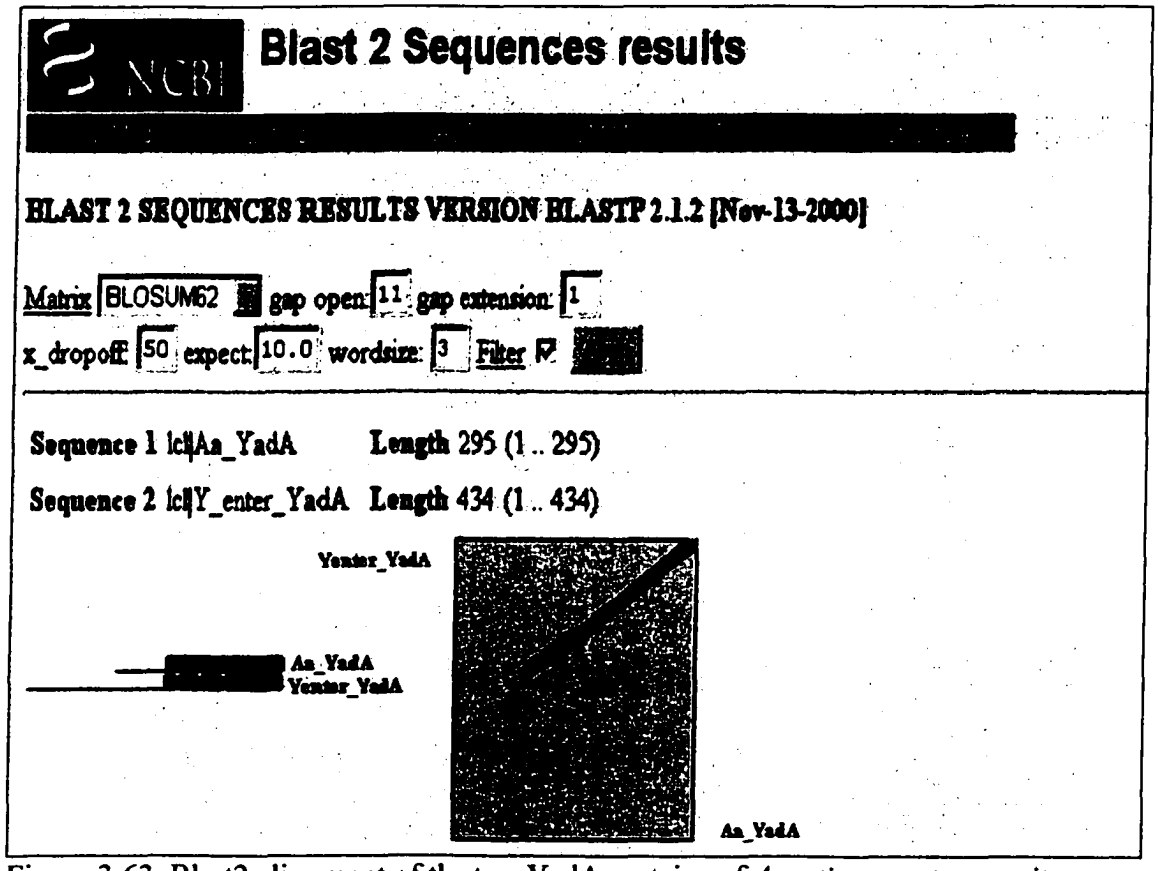


Figure 3.63. Blast2 alignment of the two YadA proteins of *A. actinomycetemcomitans* and *Y. entercolitica*. The C-terminal of both proteins is conserved while the N-terminal is truncated.

3.4.4 Invasion

Although *A. actinomycetemcomitans* is an invasive pathogen to date, no factors directly related to its invasion have been identified (**Christersson et al. 1987**). It has been observed that *A. actinomycetemcomitans* invades epithelial cells in a manner similar to *Salmonella typhimurium* by creating a lip-rimmed aperture through which it is believed to gain entry (**Meyer et al., 1996, and Ginocchio et al., 1992**). The entry of *S. typhimurium* into epithelial cells is linked to four genes, *invA, B, C, invD, and invE* (**Galan et al. 1989 and Ginocchio et al. 1992**) and movement into the host cell by *A.*

actinomycetemcomitans occurs via a receptor-mediated endocytosis which involves actin polymerization (**Meyer et al. 1991**). During analysis of the *A. actinomycetemcomitans* genomic sequence, a homolog of the invasion-associated gene *ialA* (*invA*) was discovered. The IalA protein, which also observed in *E. coli, Rickettsia prowazekii, H. influenzae, H. pylori*, and *Bortonella bacilliformis*, can increase *E. coli* invasiveness dramatically (**Mitchelle et al. 1995**). The IalA protein is characterized by its MutT domain that is used to catalyze the hydrolysis of nucleoside diphosphates (**Conyers et al., 1999**) and has been implicated as a signaling molecule in a wide variety of physiological responses including heat-shock (**Conyers et al., 1999**). Therefore, it is possible that the InvA homolog in *A. actinomycetemcomitans* also likely is involved manipulating signal transduction in the host cell to facilitate its invasion ans is the first invasion-related factor to be found in *A. actinomycetemcomitans*. Figure 5.64 illustrates the alignment of invA homolog from *A. actinomycetemcomitans* with those of *B. clarridgeiae*, and *B. bacilliformis*, in support of the above discovery.

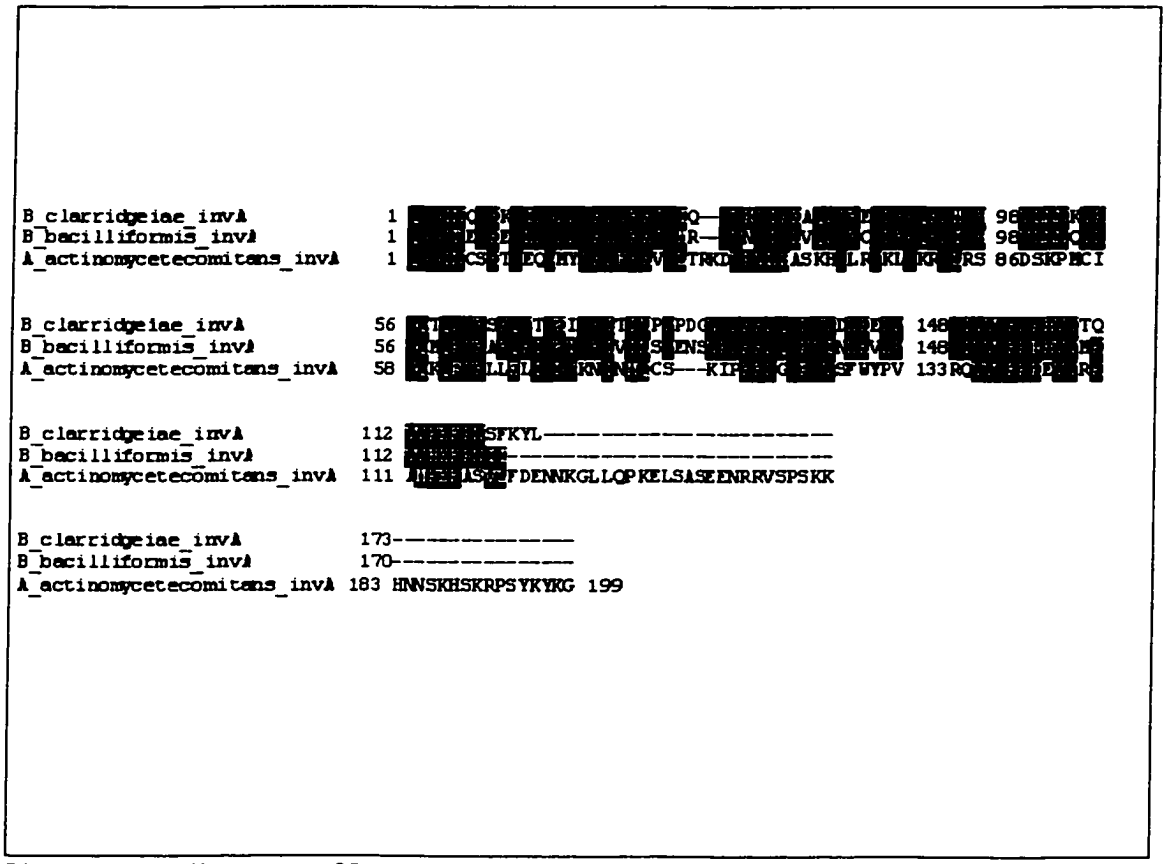


Figure 3.64. Alignment of InvA protein from *B. clarridgeiae, B. bacilliformis* and *A. actinomycetemcomitans.*

A second possible protein involved in colonization of the cavity by *A. actinomycetemcomitans* is the DksA homolog (Figure 3.65). DksA initially was discovered as a suppressor protein for mutations in heat-shock proteins such as DnaK, DnaJ and GrpE in *E. coli* (**Kang et al. 1990**) and is responsible for regulation of rRNA expression in stationary phase. DksA mutants require glutamate and glutamine to survive, suggesting a possible role in the stringent response (**Turner et al. 1998**). More recently, DksA protein has been shown to be involved in suppressing mutations in the origin of replication in plasmid pSC101 suggesting an additional role in replication

(**Ohkubo et al. 1997**) and *dksA* mutants in *Salmonella typhimurium* also reduce its colonization efficiency (**Turner et al. 1998**). Therefore, it is very likely that the DksA homolog also behave similarly, and therfore is a second putative invasiov-related protein discovered as a result of these genomic sequenceing studies.

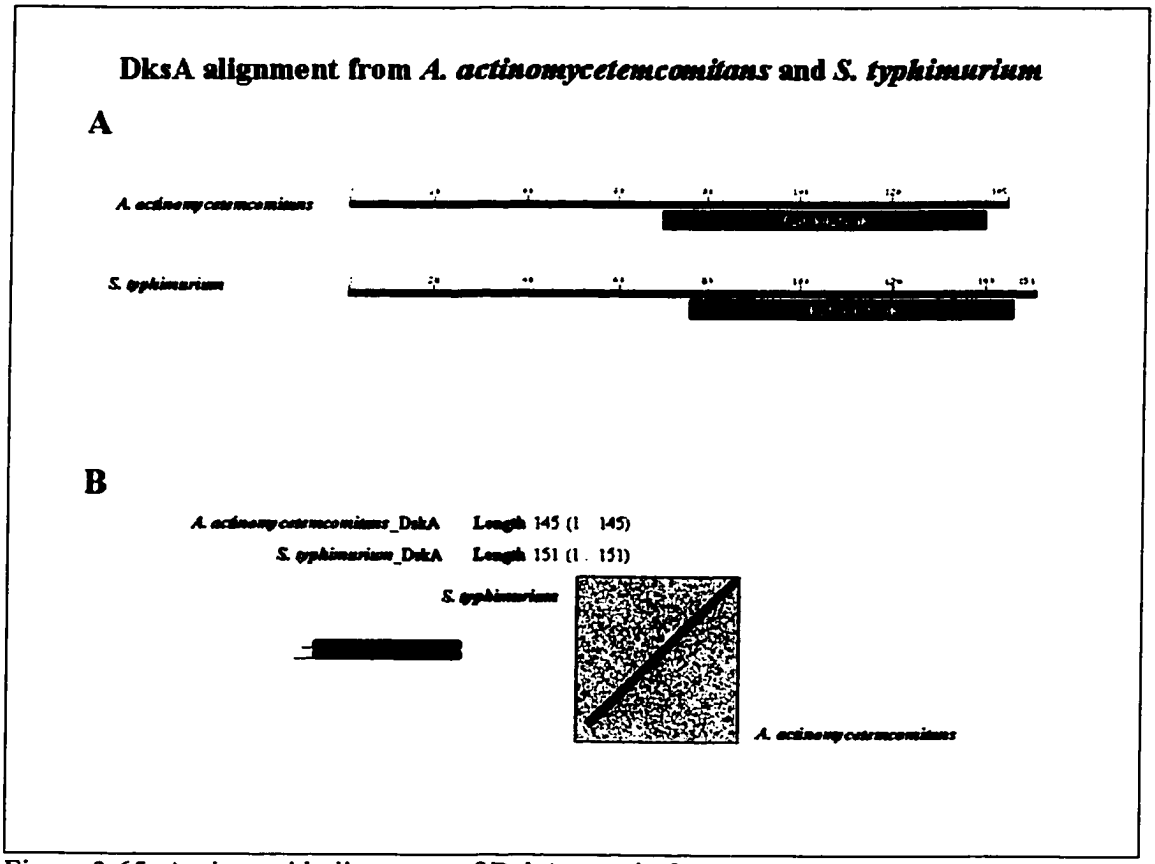


Figure 3.65. Amino acid alignment of DskA protein from *A. actinomycetemcomitans* and *S. typhimurium*. Panel A shows the conserved Zinc finger domain both homologues demonstrate. Panel B shows blastP alignment of both homologues.

3.4.4.1 DTDP-L-rhamnose pathway

Hexose dTDP-L-rhamnose is a 6-deoxyhexose glycoconjugate that is found as a cell wall component of most pathogenic bacteria (**Allard et al. 2001**) . This sugar,

synthesized through a four enzyme pathway as illustrated in figure 3.66 and listed in table 3.38 (**Tsukioka et al. 1997**), plays an important role in the invasiveness of different pathogens (**Li et al. 2000**). And analysis of the *A. actinomycetemcomitans* sequence reveals that the genes for these four enzymes are encoded in the genome. Interestingly, this pathway and its four enzymes, which do not exist in humans, and may represent them as potential target for drug developments (**Ma et al. 2001**)

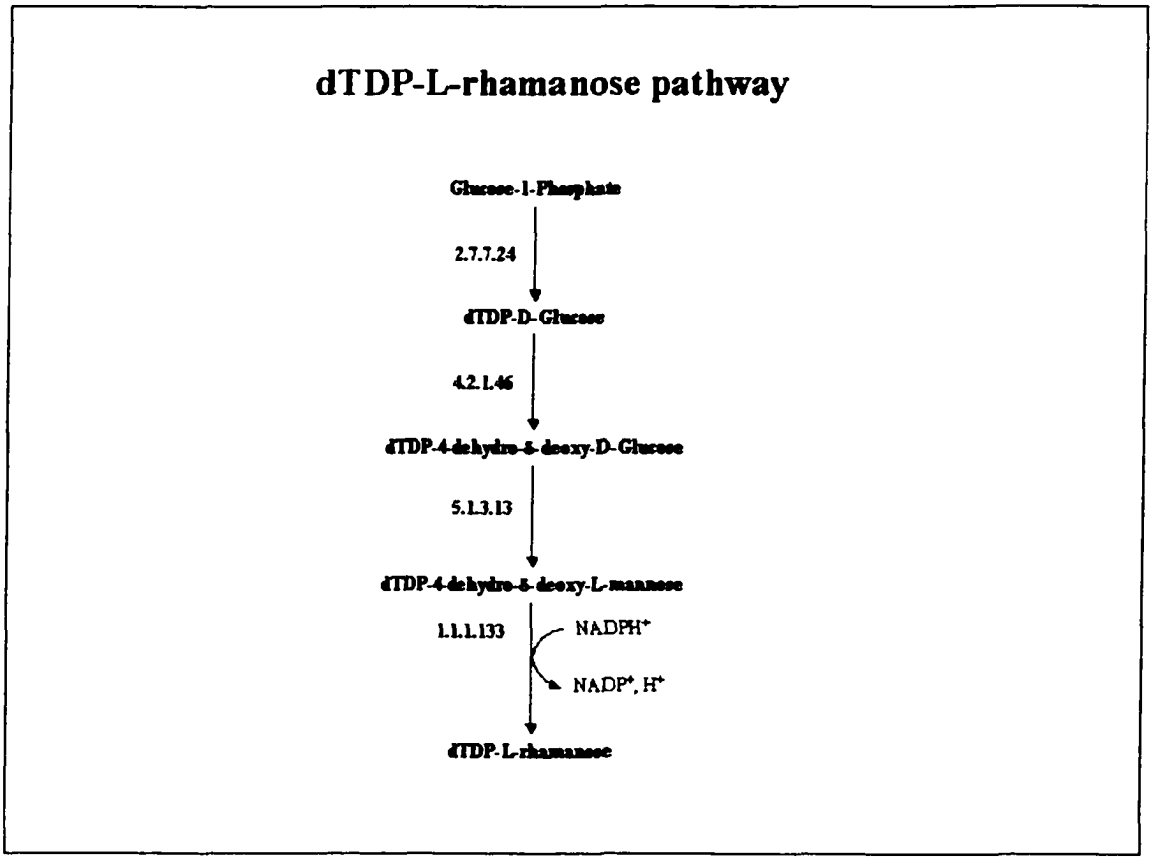


Figure 3.66. The dTDP-L-Rhamanose biosynthesis pathway as reconstructed from *A. actinomycetemcomitans* sequence data.

EC#	Description	Genes in *A. actinomycetemcomitans*
2.7.7.24	glucose-1-phosphate thymidylyltransferase	Yes
4.2.1.46	dTDP-glucose-4,6 dehydratase	Yes
5.1.3.13	dTDP-4-dehydrorhamnose-3,5-epimerase	Yes
1.1.1.133	dTDP-4-dehydrorhamnose reductase	Yes

Table 3.38. Enzymes envolved in dTDP rhamanose pathways.

3.4.5 Toxins

3.4.5.1 RTX Leukotoxin

Leukotoxins, are members of RTX toxins (Repeats in ToXin) family, are characterized by the presence of multiple copies of a consensus sequence of GGXGXDX(L|I|V|W|Y|F)X that is involved in calcium ion binding (**Learet al. 1995**). In addition, they are the only known bacterial proteins that are modified by fatty acid acylation (**Stanley et. al. 1998**). This toxin is encoded by an operon that contains all of the proteins required for its modification. The four genes in this operon, which are remarkably conserved across Gram-negative bacteria, are the *lkxC* gene whose product is responsible for acylating LkxA, the active toxin, and the *lkxB* and the *lkxD* gene products which are responsible for localizing LkxA on the outside via a type I (TolC) secretion system (**Kachlany et al. 2000**). Analysis of the *A. actinomycetemcomitans* genomic sequence reveals the presence of this operon as shown in figure 3.67. However, since earlier studies indicated that these gene products differ from those of other leukotoxin producing bacteria, since the toxin remains associated with proteases in the outer membrane of the organism in the blebs, a membrane outpocket seen in many Gram-negative bacteria (**Ohtaet al. 1991**). Thus, it is likely that *A. actinomycetemcomitans* must be in a direct contact with the target cell to impose the killing influence of this toxin.

Two models have been proposed for the mechanism of action of leukotoxins. The first model suggests that the leukotoxin forms a pore on the host cell's membrane causing small molecules to flow freely into and out of the target cell triggering cell lysis (**Bhakdi,**

et al. 1986 and Lear et al. 1995). The second model proposes a slightly different mechanism where the leukotoxin is partially inserted into the membrane of the target cell generating a lateral pressure on the molecules in the outer leaflet of the membrane causing monolayer collapse (**Soloaga et al. 1999**). Analysis of the *A. actinomycetemcomitans* genomic sequence could not support one or the other model.

In *A. actinomycetemcomitans*, the GC% of the *lkt* operon is significantly lower than the average GC content of the genome (38% versus 44.2%), suggesting a recent horizontal transfer (Figure 3.67), and consistent with the conservation of this operon among different organisms. Interestingly, the GC% of this operon in *E. coli* is 46%, clearly much higher than that of *A. actinomycetemcomitans* suggesting a more distant evolutionary event in *E. coli* integration.

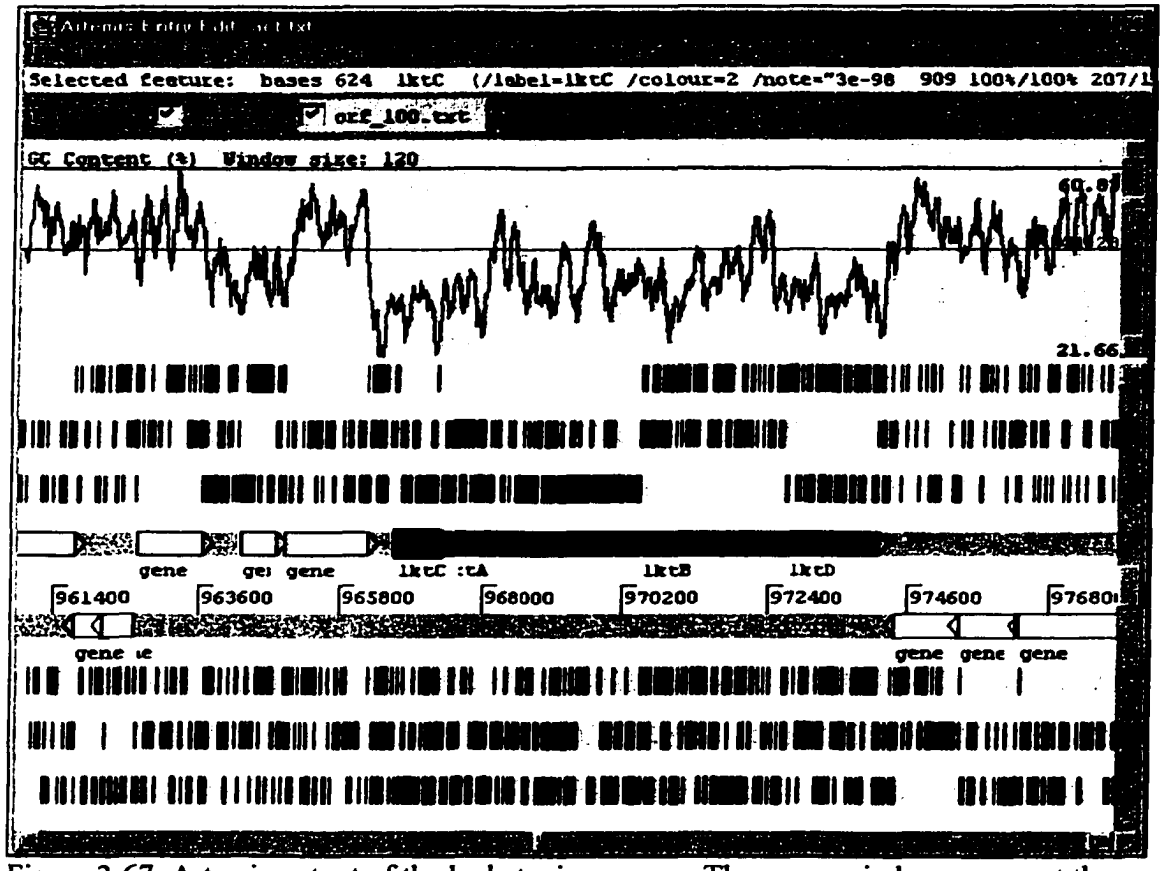


Figure 3.67. Artemis output of the leukotoxin operon. The upper window represent the GC content using a window size of 120 bp. As illustrated, the GC% drops significantly over the operon defining the start and the end of it.

3.4.5.2 Cytolethal distending toxin

Cytoleathal distending toxin (cdt) was first described when it was discovered by Johnson and colleagues in an *E. coli* strain isolated from children with diarrhea in 1987 (**Johnson et al. 1987**). Since then, this toxin has been found in several organisms, including *H. ducreyi* (**Cope et al. 1997**) and *A. actinomycetemcomitans* (**Sugai et al. 1998**). The *cdtABC* operon (Figure 3.68) that encodes the enzymes needed for the production of CDT is conserved among Gram-negative bacteria. Although *cdtC* gene

product is the active toxin, the function of the *cdtA* and *cdtB* gene products is yet to be determined. In *A. actinomycetemcomitans*, the cdt toxin has been shown to cause cell cycle arrest and distention in cell cultures, which leads to eventual cell death (**Sugai et al. 1998**), and therefore likely plays a role in the periodontal disease process.

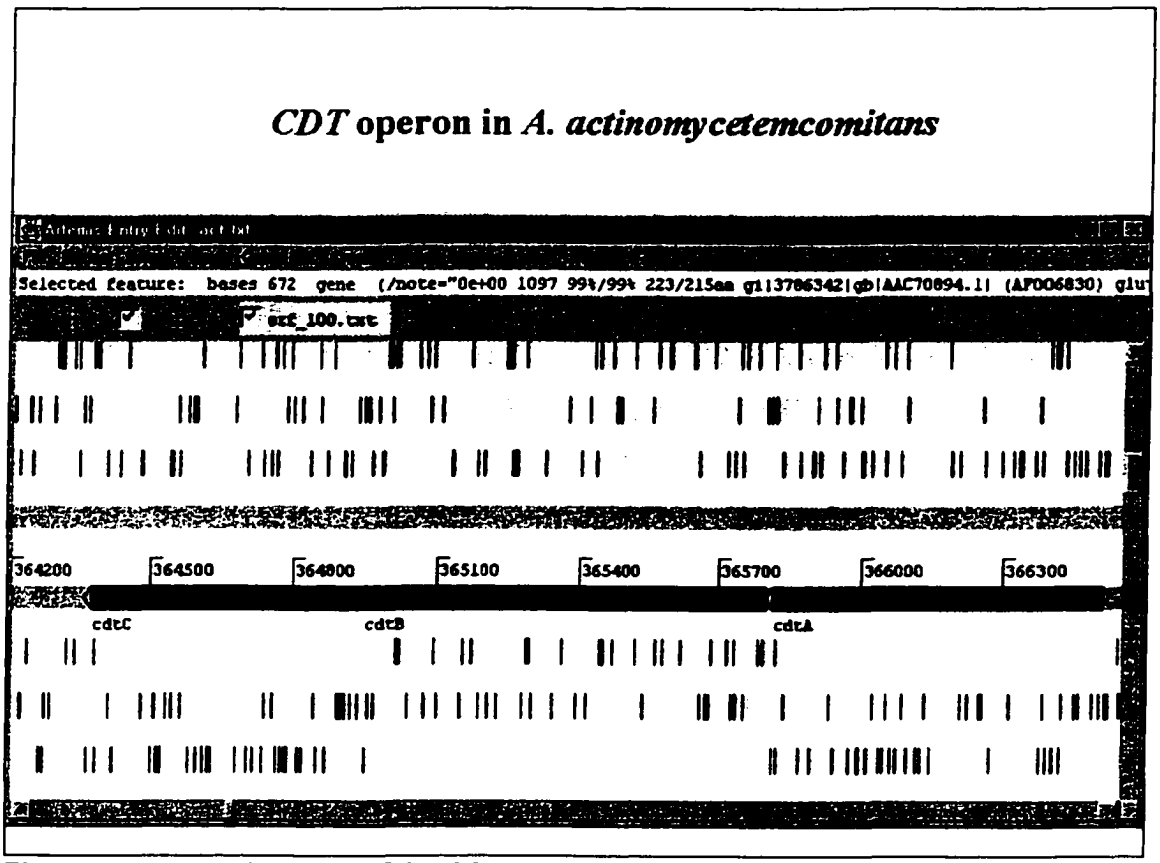


Figure 3.68. Artemis output of the CDT operon in *A. actinomycetemcomitans.*

3.4.6 Heat Shock Proteins

Heat-shock proteins such as members of the GroEL and DnaK families are cytosolic proteins mainly expressed when bacteria are exposed to environmental stress.

These proteins, also termed chaperons, bind to denatured proteins and assist them in achieving the proper functional conformation (**Bukau et al. 1998**). In addition to this functional role, heat shock proteins also assist the folding of newly translated proteins, guide these translocated proteins to their cellular functional sites, and aid in the degradation of proteins via proteolysis. In the last decade, heat shock proteins have been implicated in virulence as they are localized on the bacterial surface (**Paju et al. 2000**). Being associated with cell surface makes heat shock proteins good candidates for vaccine development. In addition, heat shock proteins have been used to diagnose disease in its early stages. Other evidence strongly suggests that heat shock proteins are contributing factors in the persistence of some diseases such as periodontitis. In *Haemophilus ducreyi* for example, mutations that affect the expression of GroEL-like proteins affect the bacteria's ability to adhere to epithelial surfaces (**Parsons et al. 1997**).

In *A. actinomycetemcomitans*, GroEL-like protein (Hsp70) is part of the cell surface associated material (**Hara et al. 2000**). This protein has demonstrated osteolysis (**Goulhen et al. 1998**) and epithelial cell proliferation (**Kirby et al. 1995**), both properties which are associated with the formation and maintenance of periodontal pockets.

Another gene revealed from the sequence of the *A. actinomycetemcomitans* genome shows strong homology to ClpB protein, a rather unique protein which belongs to a recently discovered family of heat-shock proteins that is widespread in all three kingdoms (**Squires et al. 1992**). This protein is part of the ClpB subfamily of Clp proteins that includes the ClpA and ClpC subfamilies, all of which initially were identified as heat-shock proteins (**Gottesmanet al.1990**). These proteins are

distinguished because they have two nucleotide-binding domains, a very rare occurrence in proteins (**Parsell et al. 1991**) (Figure 3.69). The only other example of proteins with two nucleotide-binding domains are the traffic ATPases that provide energy for permeation (**Ames et al. 1990**). ClpB protein seems to be associated with ClpP protease in *A. actinomycetemcomitans* so that denatured proteins can be degraded under heat shock conditions (**Squires et al. 1991**). Another function attributed to Clp family proteins is energy-dependent proteolysis, an extremely important process in regulating the availability of short-lived regulatory proteins and removing abnormal protein aggregates from the cell (**Gottesman et al.1990**). Mutations in the *clpB* gene cause accumulation of protein aggregates in yeast (**Parsell et al. 1991**) and increase the death rate in *E. coli* (**Squires et al. 1991**). The ClpP protein also has a nonspecific protease activity that is used in association with other members of the Clp family. This association gives ClpP the specificity needed to perform energy-dependent proteolysis (**Schweder et al. 1996**). ClpP associates with ClpX protein to regulate the availability of sigma factor 32 in *E. coli* during exponential phase (**Schweder et al. 1996**). Also, in *S. typhimurium*, clpB mutants lose their ability to attach and colonize 3-week-old chicks' alimentary tract (**Turner et al. 1998**).

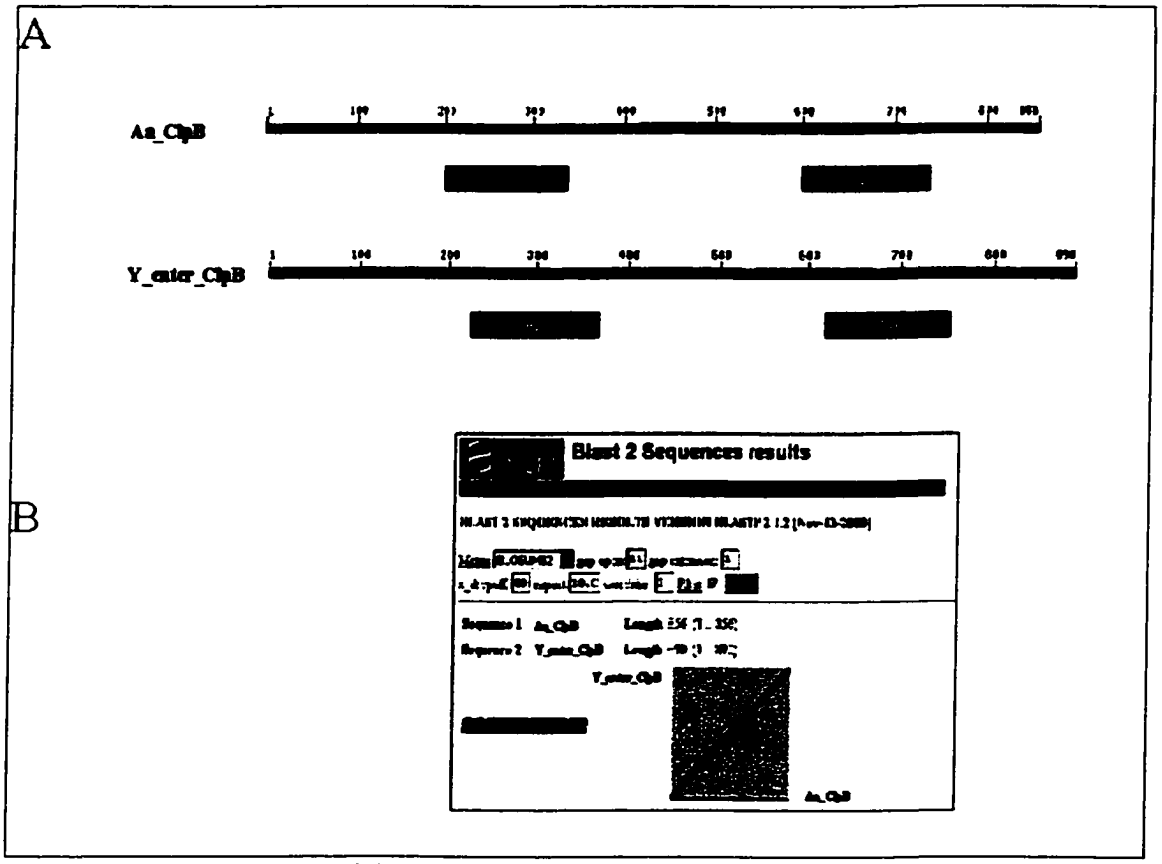


Figure 3.69. Comparison of the ClpB proteins from *A. actinomycetemcomitans* and *Y. enterocolitica*. A. The positions of the nucleotide-binding domains (NBD) on both homologs. B. the alignment of both proteins using blastP from NCBI.

In *Yersinia entercolitica*, *clpB* mutations affect both invasion and motility (**Badger et al. 2000**) as ClpB seems to regulate both invasin and flagellin expression at the transcription and translation levels (**Badger et al. 2000**). Also, it has been demonstrated that Clp protease can repress expression of Ail protein in *Y. entercolitica* (**Pederson et al. 1997**). In *Y. entercolitica*, mutational studies revealed that the attachment invasion locus protein Ail, is associated with evasion of the host immune system (**Bliska et al. 1992, and Pierson et al. 1993**) and adhesion and invasion (**Miller et al. 1988**).

In *Listeria monocytogenes*, a Gram positive pathogen, ClpC mutants promote early escape from vacuoles after phagocytosis by macrophages (**Rouquette et al. 1998**) and has been shown to be necessary for both adhesion and invasion (**Nair et al. 2000**). It therefore is quite likely that ClpB might play a similar role in pathogenesis in *A. actinomycetemcomitans*. The heat shock proteins revealed from the *A. actinomycetemcomitans* genomic sequence are listed in table 3.39.

Protein	Function	Homology
hslU	Heat shock protein	*H. influenzae*
htrA	Heat shock protein (Periplasmic serine protease)	*H. influenzae*
htrB	Heat shock protein B (Lipid A biosynthesis)	*H. influenzae*
dnaJ	Chaperon for DnaK, heat shock protein	*H. influenzae*
dnaK	Chaperon Hsp70; DNA biosynthesis; autoregulated heat shock protein	*E. coli*
hscA	Member of Hsp70 protein family	*H. influenzae*
htpG	Chaperon Hsp90; heat shock protein. associated with cell surface.	*H. influenzae*
htpX	Heat shock protein	*H. influenzae*
hsp33	KDA chaperonin	*H. influenzae*
hsp15	Heat shock protein 15 homolog	*H. influenzae*
groEL	Heat shock protein HSP60	*Pasteurella multocida*
grpE	heat shock protein, protein repair	*H. influenzae*
clpB	Hsp100 Heat shock protein	*H. influenzae*
rpoE	RNA polymerase, sigma-E factor; heat shock and oxidative stress	*H. influenzae*

Table 3.39. Heat shock protein genes encoded in the *A. actinomycetemcomitans* genome.

3.4.7 Proteases

The extracellular matrix is a network of proteins and polysaccharides that are produced by epithelial and endothelial cells. Pathogenic organisms exploit this network by binding component(s) of the extracellular matrix as part of their strategy for primary colonization at an infection site. *A. actinomycetemcomitans* has been shown to bind the extracellular matrix proteins collagen and fibronectin (**Mintz et al. 1999**). Consistent

with the observation that a prominent feature of periodontal diseases is the reduction of the gingival collagen associated with loss of teeth (**Page et al. 1973**), a collagenolytic activity was observed in *A. actinomycetemcomitans* in the early eighties by Robertson and colleagues (**Robertson et al. 1982**). Analysis of the sequence of the *A. actinomycetemcomitans* genome reveals the gene for a protease homolog which is similar to a number of collagenases and proteases in Gram-negative bacteria. Figure 3.70 illustrates the alignment of these proteins using ClustalW 1.8 (**Thompson et al. 1994**).

Another protease gene found encoded in the *A. actinomycetemcomitans* genome is a secretory IgA-specific metalloendopeptidase which is believed to be a component of the first line of immune system evasion, since sIgA is associated with the protection of mucosal surfaces (**Salyers et al. 1994**). In addition, sIgA down regulates the through inhibition of IgG, and IgM activities. Hence, destruction of sIgA by metalloendopeptidase produced by colonyzing *A. actinomycetemcomitans*, leads to increased inflammation and consequently destruction of the connective tissues in the gingiva (**Gronbaek et al. 1999**).

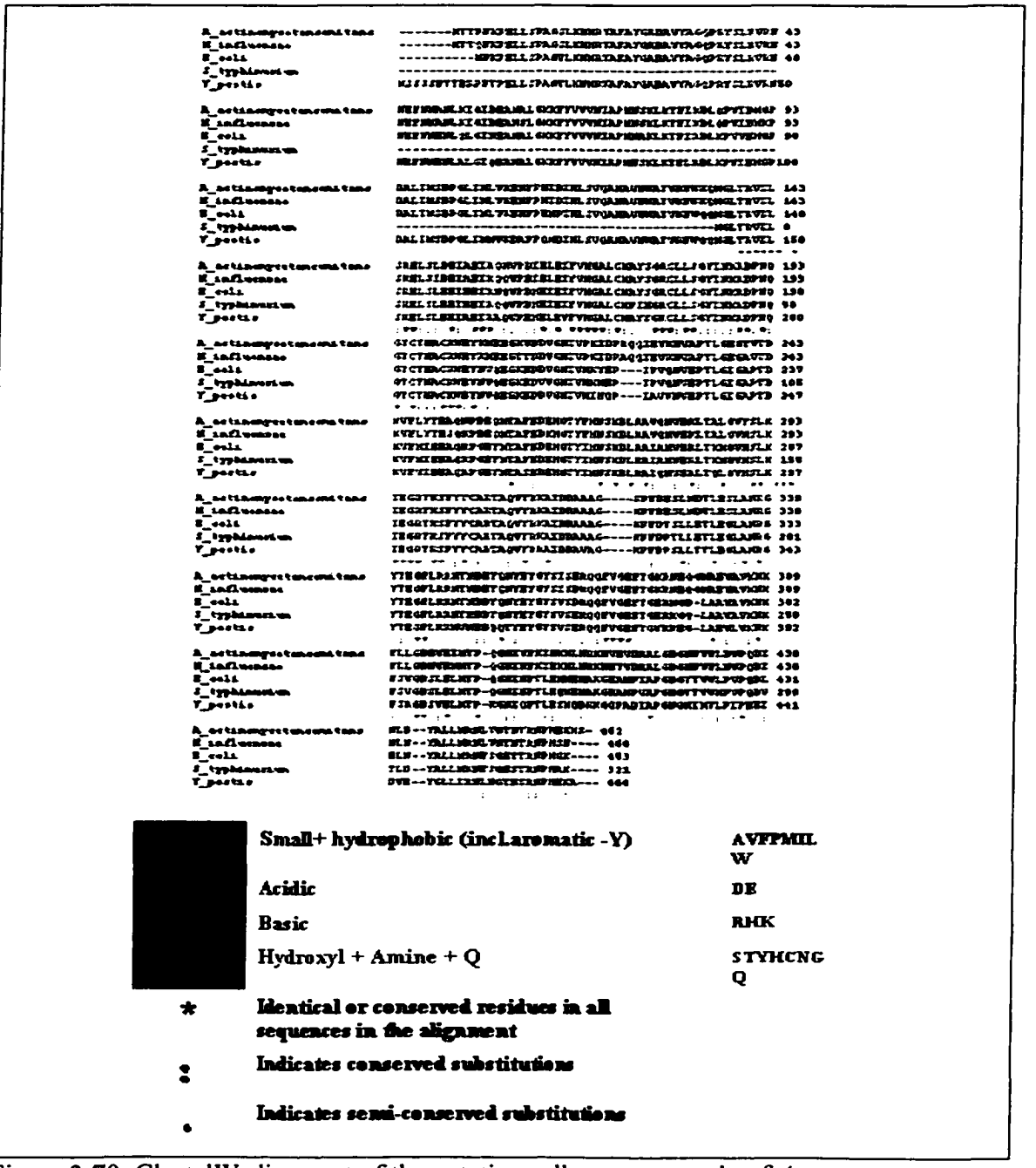


Figure 3.70. ClustalW alignment of the putative collagenase protein of *A. actinomycetemcomitans* with *H. influenzae, E. coli, Salmonella typhimurium,* and *Yersinia pestis* homologs. The alignment was run under the following conditions: KTUP=5; window size=5; Pairgap (gap penalty)=200; Gap open=25; Gap Ext=2.5; Gap Dist=6.

3.4.8. Antibiotic Resistance

In addition to the above mentioned virulence factors, *A. actinomycetemcomitans* appears to have developed a capacity to withstand a range of antibiotics. Generally, these antibiotic resistance proteins act by blocking antibiotic uptake into the cell or by modifying them (**Tillotson et al, 2001**). During the analysis of *A. actinomycetemcomitans* genomic sequence, several antibiotic resistance genes were discovered encoded in the genome as listed in table 3.40. In addition, the EmrA/B multidrug efflux pump also was observed encoded in this genome as an operon that is similar in gene organization to that found in *E. coli* and *H. influenzae*.

Virulence Factor	Genes
Colonization and invasion	
Adhesin / Invasion	Tad locus
	dTDP-L-rhamanose
	IalA homologue
	DksA: intercellular spread
	ClpC protein
	YadA protein
Antibiotic resistance	AcrB: Acriflavine resistance protein
	BcR: Bicyclomycin resistance protein
	CmlA: Resistance to chloramphenicol
	EmrA/B: Multidrug resistance efflux pump
Iron acquisition	Fec system
	BfeA Enterobactin receptor
Host cells destruction	
Collagenase	PrtC: collagenase

Table 3.40. Putative virulence factors discovered in *A. actinomycetemcomitans* genomic analysis

Chapter IV

Conclusion

The sequence of the bacterium *A. actinomycetemcomitans*, an opportunistic pathogen, which colonizes the oral cavity, reveals much new information about how this organism survives in this environment (see Taylor et al., 2000 for a comprehensive review of *A. actinomycetemcomitans* virulence proteins). Since a typical oral cavity contains more than 100 different bacterial species, including more than a dozen pathogens, *A. actinomycetemcomitans* has evolved a selective advantage by acquiring the ability to uptake the many of the nutrients it needs to survive, in addition to encoding many of their biosynthetic pathways. For example, as shown in figure 4.1, its need to import amino acids for protein synthesis and other pathways can be met both by its genome encoded biosynthetic enzymes as well as the Opp and the Sap systems for peptide uptake.

A. actinomycetemcomitans also encodes the genes for a large array of carbohydrate uptake proteins that import mono- and disaccharides, that give it the ability to survive on many different carbohydrate sources. The presence of all the genes needed to encode the enzymes for the entire pentose phosphate pathway indicates that *A. actinomycetemcomitans* can generate the NADPH needed for anabolic pathways and erythrosephosphate for aromatic amino acid synthesis. However, the citric acid cycle is not complete as it only encodes the enzymes for the conversion of α-ketoglutarate to oxaloacetate in the reducing half of this cycle. Analysis of the *A. actinomycetemcomitans* genomic sequence also reveals that although the genes are present to synthesize almost

every amino acid, the biosynthetic pathways for amino acids lysine, arginine, histidine, methionine, leucine, valine, and isoleucine, seem to be missing one or more enzyme from each pathway, a dehydratase for valine and leucine, and a transaminase for lysine, and several enzymes required for the biosynthetic pathways of arginine, histidine, and methionine. Putative alternative pathways for the synthesis of arginine and methionine were elucidated. A recent, further analysis of the *A. actinomycetemcomitans* genomic sequence using a more defined data set and a lower stringency BlastP, confirms the above observations that dihydroxy-acid dehydratase needed for valine and isolucine biosynthesis indeed was not present. However, a series of putative transaminases were found, one of which likely could be the missing enzyme in the lysine biosynthetic pathway. Therefore, to compensate for these metabolic deficiencies, *A. actinomycetemcomitans* possesses a large number of amino acid transport systems with apparently broad specificity in addition to a branched-chain-specific amino acid transport (Brn system). In the case of the histidine pathway, four of the biosynthetic enzymes are missing from the *A. actinomycetemcomitans* genome, in contrast to *H. influenzae* which possesses all 10 of the enzymes in this pathway. Here too, *A. actinomycetemcomitans* likely compensates by importing histidine from the nutrient rich oral cavity. Thus, *A. actinomycetemcomitans* likely can synthesize at least sixteen out of the twenty amino acids, and for those four (valine, isolucine, leucine, and histidine), it cannot synthesize, it utilizes peptide transporters for their uptake from the oral cavity.

A. actinomycetemcomitans can produce ATP by oxidizing sugars through mixed-acid fermentation as well as by electron-transport and occurs for oxidative phosphorylation. Interestingly, because *A. actinomycetemcomitans* possesses different

dehydrogenases, it also can use nitrate, nitrite, DMSO or fumarate as terminal electron acceptor.

Since the genes for the β-oxidation pathway are completely missing from *A. actinomycetemcomitans* genome, it clearly cannot utilize fatty acids as a carbon source. However, the genes for the synthesis of a number of fatty acids, phospholipids, peptidoglycan, and lipopolysaccharides which are needed for its membrane structure are present.

A. actinomycetemcomitans also encodes the complete set of the enzymes and protein factors needed for its replication, transcription, and translation machinery, as well as for the elaborate DNA repair pathways similar to those present in *H. influenzae*.

A. actinomycetemcomitans also is capable of synthesizing several vitamins and cofactors as its genome contains the full complement of genes needed for the biosynthesis of folate, biotin, and riboflavin. However, other vitamins must be imported either directly from the oral cavity or as derivatives that can be processed to their respective active forms. For example, since pantothenate, the coenzyme A precursor, is not be synthesized in *A. actinomycetemcomitans*, it must be imported from the environment and then metabolized to CoA. Thiamine also appears to be imported from the environment because its biosynthetic pathway is not complete.

Although the genes for most of the enzymes in the purine and pyrimidine *de novo* pathways are absent from the *A. actinomycetemcomitans* genome, all of the purine and pyrimidine salvage pathway enzymes and transporters are present.

The *A. actinomycetemcomitans* genome also encodes for the enzymes of both a typeII, general secretory pathway (GSP) and the twin-arginine pathway for protein secretion.

Figure 4.1 below summarizes the membrane proteins and metabolic pathways in *A. actinomycetemcomitans* as reconstructed from the genomic sequence data.

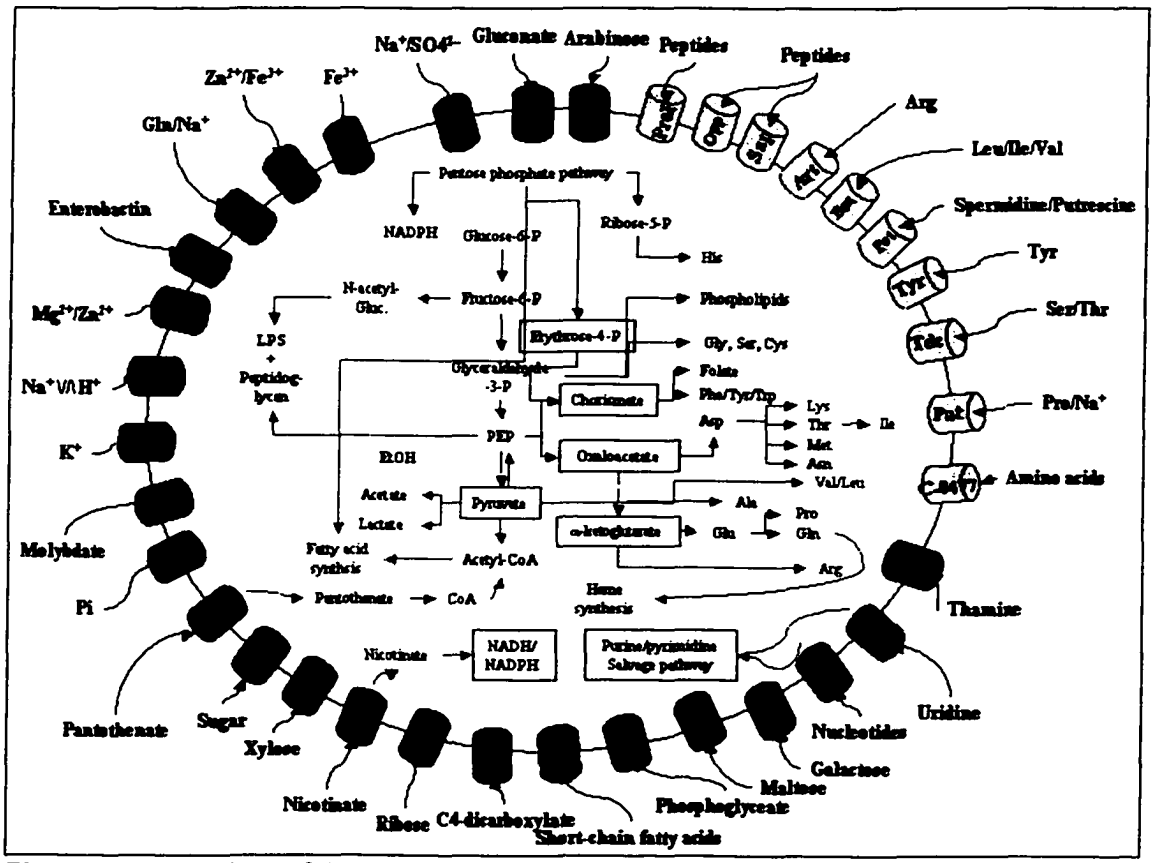


Figure 4.1. Overview of the metabolic scheme utilized by *A. actinomycetemcomitans* showing transport proteins and metabolic pathways.

A number of virulence factors were deduced from genes observed in the *A. actinomycetemcomitans* genomic sequence. The presence of leukotoxin and cytoleathal distending toxins, virulence factor involved in inducing leukocyte cell death, were

observed and characterized in late eighties and nineties (**Comayras et al., 1997, and Lally et al., 1989**), and their genomic environment now has been revealed. *A. actinomycetemcomitans* also has at least three modes for iron acquisition which have been implicated as virulence factors (**Schryvers et al., 1999**). These modes include direct iron uptake through iron transport systems, from hemoglobin binding, and possibly by scavenging through siderophore receptors. Additionally, LPS recently has been linked to iron acquisition (**Grenier et al., 1997**) and the complete pathway of this virulent factor was found in the *A. actinomycetemcomitans* genome. *A. actinomycetemcomitans* also can evade the host immune system using an encoded immunoglobulin protease (EC 3.4.24.13). Although there is still much to be known about *A. actinomycetemcomitans*, the genomic sequence data definitely has led to the discovery of new mechanisms of pathogenesis such as that of the Tad system (**Kachalaney et al. 2000**) which appears to play a role in tight nonspecific adhesion and will undoubtedly focus additional biochemical experiments aimed at confirming them.

Already, as a result of this dissertation research, several new discoveries have been made that shed light on the physiology of *A. actinomycetemcomitans* and its virulence. Indeed, as the genomic sequence of *A. actinomycetemcomitans* was made immediately publicly available as it was being collected, much new research on the pathogenesis of this organism began based on this genomic sequence data. Among the new potential virulence factors detected in *A. actinomycetemcomitans* genome was Hsf protein homologue, an adhesion factor also utilized in both *H. influenzae* and *S. typhimurium* (**St.Geme III et al. 1996**) as well as in *N. mengitidis* (**Peak et al. 2000**). Another new adhesin protein revealed is the YadA protein homologue. This protein,

which was discovered in *Y. pestis* (**Kapperud et al. 1987**), believed to be part of the colonization process in pathogens, is a surface protein that might be responsible for attaching the pathogen to the extracellular matrix as a prelude to colonization (**Hoiczyk et al. 2000**) and plays a role in shielding the pathogen from host complements (**Tertti et al. 1992**).

Among the invasion-associated genes encoded by *A. actinomycetemcomitans* genome is an InvA homologue. This protein originally was discovered in *Bartonella bacilliformis*, the only pathogen known to invade red blood cells (**Conyers et al. 1999**). *B. bacilliformis* causes bacilliary endocarditis (**Mitchelle et al. 1995**), a disease that *A. actinomycetemcomitans* is remotely associated with (**Fives-Taylor et al. 1999**), this protein may perform a similar function in *A. actinomycewtemcomitans* also by facilitating alteration of signal transduction (**Conyers et al. 1999**).

The gene encoding the DksA homologue was found in *A. actinomycetemcomitans* genome. This protein is responsible for intercellular spread in *Shigella flexneri* (**Mogull et al. 2001**). Mutants can still invade epithelial cells, but can not spread (**Mogull et al. 2001**). Additionally, genes encoding the enzymes for the DTDP-L-rhamanose pathway are encoded in the *A. actinomycetemcomitans* genomic sequence. This sugar plays a role in invasiveness that is not yet understood (**Ma et al. 2001**). Figure 4.2 below hypothesizes a putative virulence scenario based on sequence analysis of the *A. actinomycetemcomitans* genome.

The sequencing project results suggest that *A. actinomycetemcomitans* might indeed belong to the genus Haemophilus based on the significant homology to the *H. influenzae* over other bacterial genera. Although there are slight differences in the

metabolic lifestyle, possibly reflecting the different environmental niches of both species, it is clear that over 60% of the genes in *A. actinomycetemcomitans* and *H. influenzae* are at least 28% identical and that both organisms share significantly similar metabolic pathways.

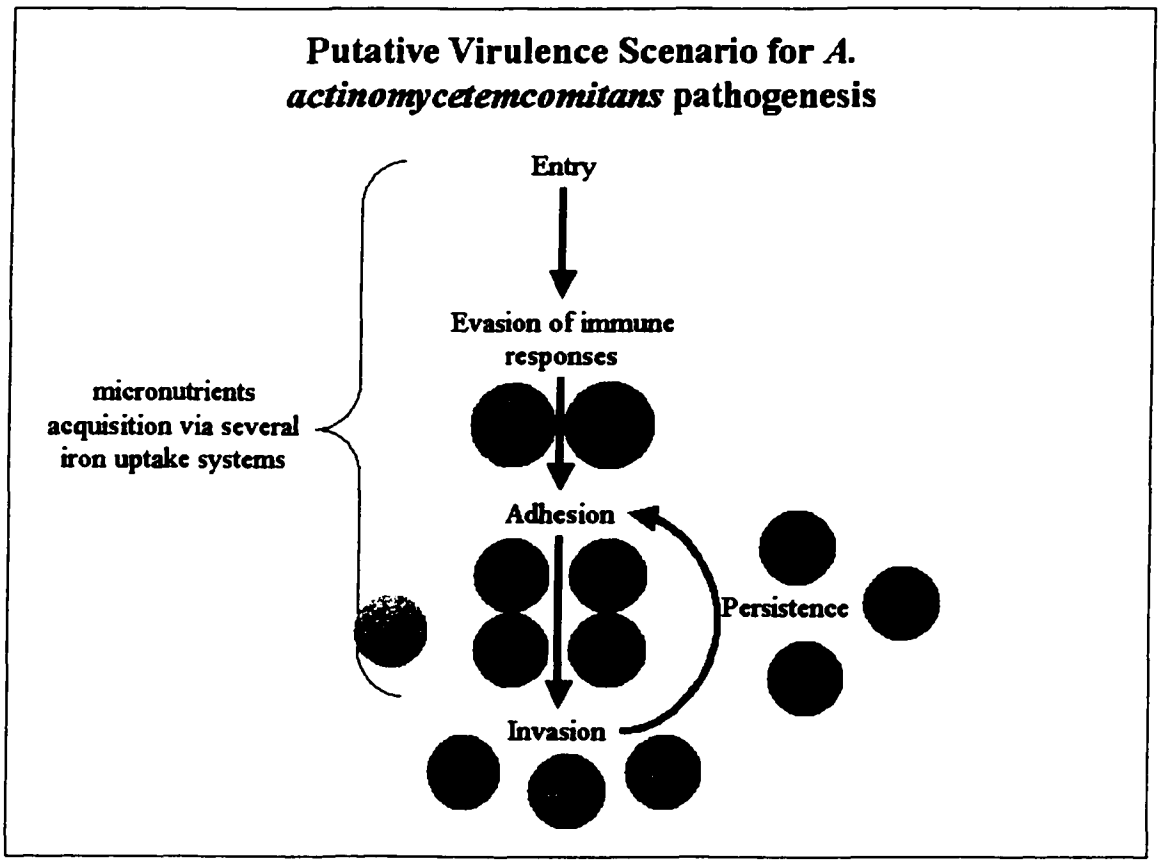


Figure 4.2. Virulence of *A. actinomycetemcomitans*.

The genomics era provides a new foundation for almost every biology-related discipline. Along with the massive amount of sequence information, new technologies and disciplines have been developed to aid in extracting and analyzing these data. The usefulness of genomics in guiding biochemical research in many areas is now well established.

The promise of genomic sequencing and bioinformatics is far-reaching and expectations are high of forthcoming medical breakthroughs in the form of vaccines and drugs that will rid society of many diseases. The availability of genome sequences for a number of pathogens is only the first step in a long process of discoveries. Mastering the technical ability to interpret the complex information generated from genomic sequencing projects will force the integration and development of novel methods for research.

With the advent of recent developments in DNA sequencing technology, an average bacteria can be sequenced to 6 to10 fold coverage within a few months. The major difficulty then, is to analyze the data generated from such projects by not only to assigning functions to predicted ORFs, but also to reconstruct the entire metabolic picture of the organism. This, in turn, will allow developing new testable hypothesis for confirmation of the functional predictions.

Comparative genomics has shown that the number of genes involved in most biological processes, such as replication and transcription, are quite similar in almost every organisms regardless of its genome size (**Fraser et al. 2000**), suggesting that these processes are both ancient and conserved as they are in *A. actinomycetemcomitans*. In contrast, other biological processes such as energy metabolism and transport systems vary from one microorganism to another, suggesting that these processes contribute to the different life styles evolved and biological niches occupied. As seen in Figure 4.3, processes such as intermediary metabolism and DNA metabolism have similar number of genes while for transport, energy metabolism, and cell envelope components, and the number of genes involved vary greatly between organisms. These unique features also

are likely to contribute in establishing their niche as pathogens in a specialized environment.

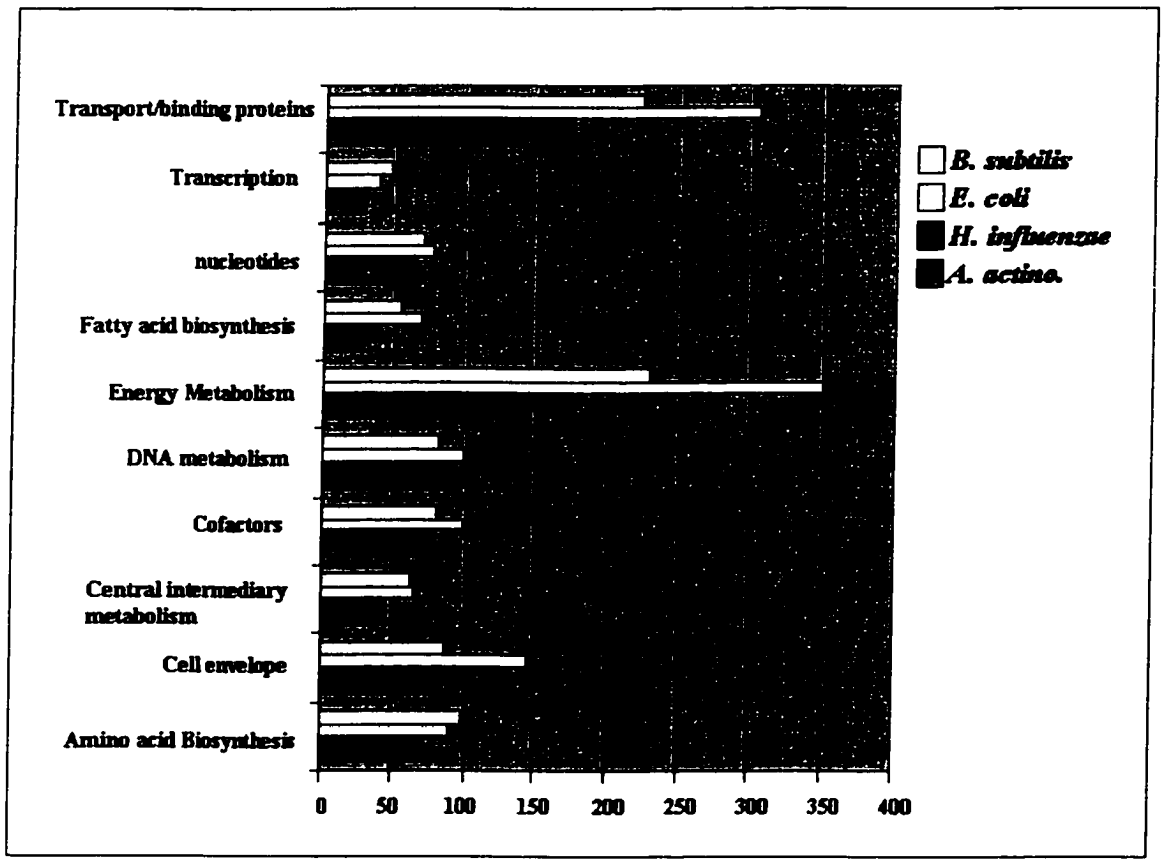


Figure 4.3. Comparison of number of genes in selected organisms involved in the major cellular processes.

As studies move from the era of genes to the era of genomes, new approaches will be needed to insure that genomics fulfills its potential. These approaches likely will require an increased collaboration between informatitions and biologists who can develop algorithms to aid in genomic data analysis and biologists/biochemists who then can design the appropriate experiments to confirm these observations. As the human genome project reaches its goal, we can look forward to even more powerful methods to evaluate

how pathogens react with their hosts within the context of the proteins encoded by both genomes.

Chapter V

References

Adhikari, P., S. A. Berish, A. J. Nowalk, K. L. Veraldi, S. A. Morse, and T. A. Mietzner. 1996. The fbpABC locus of *Neisseria gonorrhoeae* functions in the periplasm-to-cytosol transport of iron. *J. Bacteriol.* **178:** 2145-9.

Allard, S. T., M. F. Giraud, C. Whitfield, M. Graninger, P. Messner, and J. H. Naismith. 2001. The crystal structure of dTDP-D-glucose 4,6-dehydratase (RmlB) from *Salmonella enterica* serovar *Typhimurium*, the second enzyme in the dTDP-l-rhamnose pathway. *J. Mol. Biol.* **307:** 283-95.

Allugupali, K. R. and S. Kalfas. 1996. Degradation of lactoferrin by periodontitis-associated bacteria. *FEMS Microbiol. Lett.* **145:** 209-214.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403-410.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389-3402.

Andersen, P. S., D. Frees, R. Fast, and B. Mygind. 1995. Uracil uptake in *Escherichia coli* K-12: isolation of *uraA* mutants and cloning of the gene. *J. Bacteriol.* **177:** 2008-2013.

Andrade, M. A., N. P. Brown, C. Leroy, S. Hoersch, A. de Daruvar, C. Reich, A. Franchin, J. Tamames, A. Valencia, C. Ouzounis, and C. Sander. 1999. Automated genome sequence analysis and annotation. *Bioinformatics* **15:** 391-412.

Ansorge, W., B. Sproat, J. Stegemann, C. Schwager, and M. Zenke. 1987. Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res.* **15:** 4593-4602.

Ames, G. F., and A. K. Joshi. 1990. Energy coupling in Bacterial periplasmic permeases. *J. Bacteriol.* **172:** 4133-4137.

Bacher, A., S. Eberhardt, W. Eisenreich, M. Fischer, S. Herz, B. Illarionov, K. Kis, and G. Richter. 2001. Biosynthesis of riboflavin. *Vitam. Horm.* **61:** 1-49.

Badger, J. L., B. M. Young, A. J. Darwin, and V. L. Miller. 2000. *Yersinia entercolitica* ClpB affects levels of invasin and motility. *J. Bacteriol.* **182:** 5563-5571.

Baltimore, J. 1970. Central dogma reversed. *Nature.* **27:**1198-1199.

Bankier, A. T., K. M. Weston, and B. G. Barrel. 1987. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Meth. Enzymol.* **155:** 51-93.

Bartold, P. M. and S. J. Millar. 1988. Effect of lipopolysaccharide on proteoglycan synthesis by adult human gingival fibroblasts *in vitro*. *Infect. Immun.* **56:** 2149-2155.

Bateman, A., E. Birney, R. Durbin, S. R. Eddy, R. D. Finn, and E. L. Sonnhammer. 1999. Pfam 3.1: 1313 multiple alignments match the majority of proteins. *Nucleic Acids Res.* **27:** 260-262.

Bencher, M. and P. F. Sparling. 1995. Cloning, sequencing, and characterization of the gene encoding FrpB, a major iron-regulated, outer membrane protein of *Neisseria gonorrhoeae*. *J. Bacteriol.* **177:** 2041-2049.

Beall, B. and G. N. Sanden. 1995. A *Bordetella pertussis* fep*A* homologue required for utilization of exogenous ferric enterobactin. *Microbiology* **141:** 3193-3205.

Bear, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. G. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Sequin, P. S. Tuffnell, and B. G. Barrell. 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* **310:** 207-211.

Bearden, S. W., T. M. Staggs, and R. D. Perry. 1998. An ABC transporter system of *Yersinia pestis* allows utilization of chelated iron by *Escherichia coli* SAB11. *J. Bacteriol.* **180(5):** 1135-47.

Beaucage, S. L. and M. H. Caruthers. 1981. Deoxynucleoside phosphoramidites - a new class of key intermediates for deoxypolynucleotide synthesis . *Tetrahedron Lett.* **22:** 1859-1862.

Behnke, D., H. Malke, M. Hartmann, and F. Walter. 1979. Post-transformational rearrangement of an *in vitro* reconstructed group-A streptococcal erythromycin resistance. *Plasmid* **2:** 605-16.

Berks, B. C. 1996. A common export pathway for proteins binding complex redox cofactors. *Mol. Microbiol.* **22:** 393–404.

Berks, B. C., F. Sargent, E. D. Leeuw, A. P. Hinsley, N. R. Stanley, R. L. Jack, G. Buchanan, and T. Palmer. 2000. A novel protein transport system involved in the biogenesis of the electron transfer chains. *Biochim. Biophys. Acta* **1459:** 325-330.

Bhakdi, S., N. Mackman, J. M. Nicaud, and I. B. Holland. 1986. *Escherichia coli* hemolysin may damage target cell membranes by generating transmembrane pores. *Infect. Immun.* **52:** 63-69.

Binet, R., S. Letoffe, J. M. Ghigo, P. Delepelaire, and C. Wandersman. 1997. Protein secretion by Gram negative bacterial ABC exporters - a review. *Gene* **192:** 7-11.

Bjork, G. R. 1996. Stable RNA Modification. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 861-886. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Bliska, J. and S. Falkow. 1992. Bacterial resistance to complement killing mediated by the Ail protein of *Yersinia entercolitica. Proc. Natl. Acad. Sci. (USA)* **89:** 3561-3565.

Bodenteich, A., S. Chissoe, Y. F. Wang, and B. A. Roe. 1993. Shotgun cloning as a strategy of choice to generate templates for high-throughput dideoxynucleotide sequencing. In: *Automated DNA sequencing and analysis techniques.* J. C. Venter (Ed.). Academic Press, London.

Bock, A. and G. Sawer. 1996. Fermentation. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 262-282. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Bogsch, E. G., F. Sargent, N. R. Stanley, B. C. Berks, C. Robinson, and T. Palmer. 1998. An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. *J. Biol. Chem.* **273:** 18003–18006.

Bono, H., H. Ogata, S. Goto, and M. Kanehisa. 1998. Reconstruction of amino acid biosynthesis pathways from the complete genome sequence. *Genome Res.* **8:** 203-210.

Boos, W. and J. M. Lucht. 1996. Periplasmic Binding Protein-Dependent ABC Transporters. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 1175-1209. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Bosse, J. T., H. D. Gilmour, and J. I. MacInnes. 2001. Novel genes affecting urease acivity in *Actinobacillus pleuropneumoniae. J. Bacteriol.* **183:** 1242-1247.

Brogan J. M., E. T. Lally, K. Poulsen, M. Kilian, and D. R. Demouth. 1994. Regulation of *Actinobacillus actinomycetemcomitans* leukotoxin expression: Analysis of the regions of leukotoxic and minimally leukotoxic strains. *Infect. Immun.* **62:** 501-508.

Bucher, P. and A. Bairoch. 1994. A Generalized Profile Syntax for Biomolecular Sequences Motifs and its Function in Automatic Sequence Interpretation. In: *ISMB-94;*

Proceedings 2nd International Conference on Intelligent Systems for Molecular Biology. pp. 53-61. Altman, R., D. Brutlag, P. Karp, R. Lathrop, and D. Searls (Eds.). AAAIPress, Menlo Park.

Bukau, B. and A. L. Horwich. 1998. The Hsp70 and Hsp60 chaperone machines. *Cell* **92**: 351-66.

Chamberlan, J. S., R. A. Gibbs, J. E. Ranier and C. T. Caskey. 1990. Multiplex PCR for the Diagnosis of Duchenne Muscular Dystrophy. In: *PCR Protocols.* pp. 272-281. Innis, M. A., D. H. Gelfand, J. J. Sninsky and T. J. White (Eds.). Academic Press, NY.

Chin, N., J. Frey, C. F. Chang, and Y. F. Chang. 1996. Identification of a locus involved in the utilization of iron by *Actinobacillus pleuropneumoniae. FEMS Microbiol. Lett.* **143:** 1-6.

Christersson, L. A., B. Albini, J. J. Zambon, U. M. E. Wikesjo, and R. J. Genco. 1987. Tissue localization of *Actinobacillus actinomycetemcomitans* in human periodontitis. I. Light immunofluorescence and culture techniques. *J. Periodontol.* **58:** 528-539.

Claustres, M., P. Kjellberg, M. Desgeorges, H. Bellet, P. Sarda, H. Bonnet, and C. Boileau. 1989. Detection of deletions by the amplification of exons (multiplex PCR) in Duchenne muscular dystrophy. *J. Genet. Hum.* **37(3):** 251-7.

Collazo, C. M. and J. E. Galan. 1997. The invasion-associated type-III protein secretion system in *Salmonella* - a review. *Gene* **192:** 51-59.

Connel, C., S. Fung, W. Mordanm, M. Raff, M. Recknor, L. Smith, J. Springer, S. Woo, and M. Hunkapiller. 1989. Automated DNA sequence analysis. *Biotechniques* **5:** 342-347.

Comayras, C., C. Tasca, S. Y. Peres, B. Ducommun, E. Oswald, and J. DeRyke. 1997. *Escherichia coli* cytolethal distending toxin blocks the HeLa cell cycle at the G2/M transition by preventing Cdc2 protein kinase dephosphorylation and activation. *Infect. Immun.* **65:** 5088-5095.

Conyers, G. B. and M. J. Bessman. 1999. The gene, *ial*A, associated with the invasion of human erythrocytes by *Bartonella bacilliformis*, designates a nudix hydrolase active on dinucleoside 5*-polyphosphates*. *J. Biol. Chem.* **274:** 1203-1206.

Cope, L. D., S. Lumbley, J. L. Latimer, J. Klesney-Tait, M. K. Stevens, L. S. Johnson, M. Purven, R. S. Munson JR., T. Lagergard, J. D. Radolf, and E. J. Hansen. 1997. A diffusible cytotoxin of *Haemophilus ducreyi. Proc. Natl. Acad. Sci. (USA)* **94:** 4056-4061.

Cordwell, S. J. 1999. Microbial genomes and "missing" enzymes: redefining biochemical pathways. *Arch. Microbiol.* **172:** 269-279.

Corpet, F., J. Gouzy, and D. Kahn. 1999. Recent improvements of the ProDom database of protein domain families. *Nucleic Acids Res.* **27:** 263-267.

Crick, F. 1970. Central dogma of molecular biology. *Nature.* **227:** 561-563.

Cronan,J. E. and C. O. Rock. 1996. Biosynthesis of Membrane Lipids. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 612-636. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Curnow, A. W., K. Hong, R. Yuan, S. Kim, O. Martins, W. Winkler, T. M. Henkin, and D. Söll. 1997. Glu-tRNAGln amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl. Acad. Sci. (USA)* **94:** 11819-11826.

Cynamon, M. H., T. B. Sorg, and A. Patapow. 1988. Utilization and metabolism of NAD by *Haemophilus parainfluenzae. J. Gen. Microbiol.* **134:** 2789-2799.

Danchin, A., P. Guerdoux-Jamet, I. Moszer, and P. Nitschke. 2000. Mapping the bacterial cell architecture into the chromosome. *Phil. Trans. R. Soc. Lond. B* **355:** 179-190.

De Ley, J., W. Mannheim, R. Mutters, K. Piechulla, R. Tytgat, P. Segers, M. Bisgaard, W. Frederiksen, K. H. Hinz, and M. Vanhoucke. 1990. Inter- and intrafamilial similarities of rRNA cistrons of the *Pasteurellaceae. Int. J. Syst. Bacteriol.* **40:** 126-137.

Dean, C. R. and K. Poole. 1993. Cloning and characterization of the ferric enterobactin receptor gene (*pfeA*) of *Pseudomonas aeruginosa. J. Bacteriol.* **175:** 317-324.

Dear, S., and R. Staden. 1991. A sequence assembly and editing program for efficient management of large projects. *NAR* **19:** 3907-3911.

Deininger, P. L. 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal. Biochem.* **129:** 216-223.

Dicks, J. 2000. Graphical tools for comparative genome analysis. *Yeast* **17:** 6-15.

Dierick, H., M. Stul, W. De Kelver, P. Marynen, J. J. Cassiman. 1993. Incorporation of dITP or 7-deaza dGTP during PCR improves sequencing of the product. *NAR.* **21:** 4427-4428.

DiRienzo, J. M., J. Slots, M. Sixou, M. A. Sol, R. Harmon, and T. L. McKay. 1994. Specific genetic variants of *Actinobacillus actinomycetemcomitans* correlate with disease and health in a regional population of families with localized juvenile periodontitis. *Infect. Immun.* **62:** 3058-3065.

Eddy, S.R. and R. Durbin. 1994. RNA sequence analysis using covariance models. *NAR* **22:** 2079-2088.

Emanuelsson, O., H. Nielsen, S. Brunak, and G. von Heijne. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300:** 1005-1016.

Engelberg-Kulka, H. and R. Schoulaker-Schwarz. 1996. Suppression of Termination Codons. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 909-921. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Enz, S., S. Mahren, U. H. Stroeher, and V. Braun. 2000. Surface Signaling in Ferric Citrate Transport Gene Induction: Interaction of the FecA, FecR, and FecI Regulatory Proteins. *J. Bacterio.* **182:** 637-646.

Ewing, B., L. Hillier, C. M. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using Phred I. Accuracy assessment. *Genome Research* **8:** 175-185.

Ewing, B. and P. Green. 1998. Base-calling of automated sequencer traces using Phred II. Error probabilities. *Genome Research* **8:** 186-194.

Fabrega, C., M. A. Farrow, B. Mukhopadhyay, V. deCrecy-Lagard, A. R. Ortiz, and P. Schimmel. 2001. An aminoacyl tRNA synthetase whose sequence fits into neither of the two known classes. *Nature* **411:** 110-114.

Falkow, S. 1991. Bacterial entry into eukaryotic cells. *Cell* **65:** 1099-1102.

Falkow, S. 1996. The Evolution of Pathogenicity in *Escherichia*, *Shigella*, and *Salmonella*. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 2. pp. 2723-2729. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Felmlee, T. and R. A. Welch. 1988. Alteration of amino acid repeats in *Escherichia coli* hemolysin affect cytolytic activity and secretion. *Proc. Natl. Acad. Sci. (USA)* **85:** 5279-5273.

Fichant, G.A. and C. Burks. 1991. Identifying potential tRNA genes in genomic DNA sequences. *J. Mol. Biol.* **220:** 659-671.

Field, D., D. Hood, and R. Moxon. 1999. Contribution of genomics to bacterial pathogenesis. *Curr. Opin. Genet. Develop.* **9:** 700-703.

Fine, D. H., D. Furgang, H. C. Schreiner, P. Goncharoff, J. Charlesworth, G. Ghazwan, P. Fitzgerald-Bocarsly, and D. H. Figurski. 1999. Phenotypic variation in *Actinobacillus actinomycetemcomitans* during laboratory growth: implications for virulence. *Microbiology* **145:** 1335-1347.

Fitzgerald, M. C., P. Skowron, J. L. Van Etten, L. M. Smith, and D. A. Mead. 1992. Rapid shotgun cloning utilizing the two base recognition endonuclease *CviJI. Nucleic Acids Res.* **20:** 2753-2762.

Fives-Taylor, P. M., D. H. Meyer, K. P. Mintz, and C. Brissette. 1999. Virulence factors of *Actinobacillus actinomycetemcomitans. Periodontology 2000.* **20:** 136-167.

Fleischmann,R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. G. Sutton, W. FitzHugh, C. A. Fields, J. D. Gocayne, J. D. Scott, R. Shirley, L. I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496-512.

Foster, J. and A. G. Moat. 1980. Nicotinamide adenine dinucleotide biosynthsis and pyridine nucleotide cycle metabolism in microbial systems. *Microbiol. Rev.* **44:** 83-105.

Fraser, C. M., S. J. Norris, G. M. Weinstock, O. White, G. G. Sutton, R. Dodson, M. Gwinn, E. K. Hickey, R. Clayton, K. A. Ketchum, E. Sodergren, J. M. Hardham, M. P. McLeod, S. Salzberg, J. Peterson, H. Khalak, D. Richardson, J. K. Howell, M. Chidambaram, T. Utterback, L. McDonald, P. Artiach, C. Bowman, M. D. Cotton, J. C. Venter et al. 1998. Complete genome sequence of *Treponema pallidum*, the Syphilis spirochete. *Science* **281:** 375-88.

Fraser, C. M., J. Eisen, R. D. Fleischmann, K. A. Ketchum, and S. Peterson. 2000. Comparative genomics and understanding of microbial biology. *Emerg. Infect. Diseases* **6**: 505-512.

Friedberg, E. C., G. C. Walker, and W. Siede. 1995. DNA Repair and Mutagenesis *(ASM Press).*

Galan, J. E. and R. Curtiss. 1989. Cloning and molecular characterization of genes whose products allow *Salmonella typhimurium* to penetrate tissue culture cells. *Proc. Natl. Acad. Sci. (USA)* **86:** 6383-6387.

Geme III, J. W., D. Cutter, and S. J. Barenkamp. 1996. Characterization of the genetic locus encoding *Haemophilus influenzae* type b fibril. *J. Bacteriol.* **178:** 6281-6287.

Gennis, R. B. and V. Stewart. 1996. Respiration. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 217-261. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Gillespie, S. H., H. McWhinney, S. Patel, J. G. Raynes, K. P. McAdam, R. A. Whiley, and J. Hardie. 1993. Species of alpha-hemolytic *Streptococci* posessing a C-polysaccharide phosphorylcholine-containing antigen. *Infect. Immun.* **61:** 3076-3077.

Ginocchio, C., J. Pace, and J. E. Galan. 1992. Identification and molecular characterization of a *Salmonella typhimurium* gene involved in triggering the internalization of *Salmonellae* into cultured epithelial cells. *Proc. Natl. Acad. Sci. (USA)* **89**: 5976-5980.

Gordon, D., C. Abajian, and P. Green. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8:** 195-202.

Gottesman, S., C. Squires, E. Pichersky, M. Carrington, M. Hobbs, J. S. Mattick, B. Dalrymple, H. Kuramitsu, T. Foster, T. Shiroza, W. T. Clark, B. Ross, C. L. Goulhen, F. A. Hafezi, V.-J. Uitto, D. Hinode, R. Nakamura, D. Grenier, and D. Mayrand. 1998. Subcellular localization and cytotoxic activity of the GroEL-like protein isolated from *Actinobacillus actinomycetemcomitans. Infect. Immun.* **66:** 5307-5313.

Graber, K. R., L. M. Smoot, and L. A. Actis. 1998 Expression of iron binding proteins and hemin binding activity in the dental pathogen *Actinobacillus actinomycetemcomitans. FEMS Microbiol. Lett.* **163:** 135-142.

Graber, K. R., L. M. Smoot, and L. A. Actis. 1998. Expression of iron binding proteins and hemin binding activity in the dental pathogen *Actinobacillus actinomycetemcomitans. FEMS Micobiol. Lett.* **163:** 135-142.

Green, J. M., B. P. Nicholas, and R. G. Matthews. 1996. Folate Biosynthesis, Reduction, and Polyglutamylation. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 665-673. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Grenier, D., A. Leduc, and D. Mayrand. 1997. Interaction between *Actinobacillus actinomycetemcomitans* lipopolysaccharides and human hemoglobin. *FEMS Microbiol. Lett.* **151:** 77-81.

Gronbaek, F. E. V. 1999. Bacterial degradation of immunoglobulin A1 in relation to periodontal diseases. *APMIS Suppl.* **87:** 1-54.

Haase, E. M., J. L. Zmuda, and F. A. Scannapieco. 1999. Identification and molecular analysis of rough-colony-specific outer membrane proteins of *Actinobacillus acintomycetemcomitans. Infect. Immun.* **67:** 2901-2908.

Hamann, C. S., K. R. Sowers, R. S. A. Lipman, and Y.-M. Hou. 1999. An archaeal aminoacyl-tRNA synthetase missing from genomic analysis. *J. Bacteriol.* **181:** 5880–5884.

Hara, H., H. Ohta, T. Inoue, T. Ohashi, S. Takashiba, Y. Murayama, and K. Fukui. 2000. Cell surface-associated enolase in *Actinobacillus actinomycetemcomitans. Microbiol. Immunol.* **44**: 349-56.

Hardie, K. R., J.-P. Issartel, E. Koronakis, C. Hughes, and V. Koronakis. 1991. *In vitro* activation of *Escherichia coli* prohaemolysin to the mature membrane targeted toxin requires hlyC and a low molecular weight cytosolic polypeptide. *Mol. Microbiol.* **5:** 1669-1679.

Haubeck, D., K. Poulsen, S. Asikainen, and M. Kilian. 1995. Evidence for absence in northern Europe of especially virulent clonal type of *Actinobacillus actinomycetemcomitans. J. Clin. Microbiol.* **33**: 395-401.

Heinrichs, D. E., J. A. Yethon, and C. Whitfield. 1998. Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica. Mol. Microbiol.* **30:** 221-32.

Hejtmancik, J. F., S. G. Harris, C. C. Tsao, P. A. Ward, and C. T. Caskey. 1986. Carrier diagnosis of Duchenne muscular dystrophy using restriction fragment length polymorphisms. *Neurology* **36:** 1553-1562.

Henikoff, J. G., S. Henikoff, and S. Pietrokovski. 1999. New features of the Blocks Database servers. *Nucleic Acids Res.* **27:** 226-228.

Henikoff, S., J. G. Henikoff, and S. Pietrokovski. 1999. Blocks+: A non-redundant database of protein alignment blocks derived from multiple compilations. *Bioinformatics* **15:** 471-479.

Hofmann, K., P. Bucher, L. Falquet, and A. Bairoch. 1999. The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27:** 215-219.

Hoiczyk, E., A. Roggenkamp, M. Reichenbecher, A. Lupas, and J. Heesemann. 2000. Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. *EMBO J.* **19:** 5989-5999.

Hove-Jensen B. and P. Nygaard. 1989. Role of guanosine kinase in the utilization of guanosine for nucleotide synthesis in *Escherichia coli. J. Gen. Microbiol. 135:* 1263-1273.

Huynen, M. A., T. Dandekar, and P. Bork. 1999. Variation and evolution of the citric acid cycle: a genomic perspective. *Trends Microbiol.* **291:** 281-291.

Ibba, M., A. W. Curnow, and D. Soll. 1997. Aminoacyl-tRNA synthesis: divergent routes to a common goal. *Trends. Biochem. Sci.* **22:** 39-42.

Ibba, M., S. Morgan, A. W. Curnow, D. R. Pridmore, U. C. Vothknecht, W. Gardner, W. Lin, C. R. Woese, and D. So. 1997. A euryarchaeal lysyl–tRNA synthetase: resemblance to class I synthetases. *Science* **278:** 1119-1122.

Innis, M. A., K. B. Myambo, D. H. Gelfand, and M. A. D. Brow. 1988. DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci. (USA)* **85:** 9436-9440.

Inoue, T., I. Tanimoto, H. Ohta, K. Kayo, Y. Murayama, and K. Fukui. 1998. Molecular characterization of low-molecular-weight component protein, Flp, in *Actinobacillus actinomycetemcomitans* fimbriae. *Microbiol. Immunol.* **42:** 253-258.

Issartel, J.-P., V. Koronakis, and C. Hughes. 1991. Activation of *Escherichia coli* prohaemolysin to the mature toxin by acyl carrier protein dependent fatty acylation. *Nature* (London) **351:** 759-761.

Isola N. R., S. L. Allman, V. V. Golovlov, and C. H. Chen. 1999. Chemical cleavage sequencing of DNA using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Chem.* **71:** 2266-2269.

Jachowsky S. 1996. Biosynthesis of Pantothenic Acid and Coenzyme A. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 687-694. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Jackowski, S. and J. H. Alix. 1990. Cloning, sequence, and expression of the pantothenate permease (panF) gene of *Escherichia coli. J. Bacteriol.* **172:** 3842-3848.

Jensen, R. A., and D. H. Calhoun. 1981. Intracellular roles of microbial aminotransferases: overlap enzymes across different biochemical pathways. *Crit. Rev. Microbiol.* **8:** 229-66

Jin, H., Z. Ren, J. M. Pozsgay, C. Elkins, P. W. Whitby, D. J. Morton, and T. L. Stull. 1996. Cloning of a DNA fragment encoding a heme-repressible hemoglobin-binding outer membrane protein from *Haemophilus influenzae*. *Infect. Immun.* **64:** 3134-3141.

Johnson, K., I. Charles, G. Dougan, D. Pickard, P. O'Gaora, G. Costa, T. Ali, I. Miller, and C. Hormaeche. 1991. The role of stress-response protein in *Salmonella typhimurium* virulence. *Mol. Microbiol.* **5:** 401-407.

Johnson, W. M. and H. Lior. 1987. Response of Chinese hamster ovary cells to a cytolethal distending toxin (CDT) of *Escherishia coli* and possible misinterpretation as heat-labile (LT) enterotoxin. *FEMS Micribiol. Lett.* **43:** 19-23.

Ju J., C. Ruan, C. W. Fuller, A. N. Glazer, and R. A. Mathies. 1995. Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc. Natl. Acad. Sci. (USA)* **92:** 4347-4351.

Kachlany, S. C., D. H. Fine, and D. H. Figurski. 2000. Secretion of RTX leukotoxin by *Actinobacillus actinomycetemcomitans*. *Infect. Immun.* **68**: 6094-6100.

Kachlany, S. C., P. J. Planet, M. K. Bhattacharjee, E. Kollia, R. DeSalle, D. H. Fine, and D. H. Figurski. 2000. Nonspecific adherence by *Actinobacillus actinomycetemcomitans* requires genes widespread in bacteria and archaea. *J. Bacteriol.* **182:** 6169-6176.

Kahler, C. M., R. W. Carlson, M. M. Rhman, L. E. Martin, and D. S. Stephens. 1996. Inner core biosynthesis of lipooligosaccharide (LOS) in *Neisseria meningitidis* serogroup B: identification and role in LOS assembly of the 1,2 N-acetylglucosamine transferase (RfaK). *J. Bacteriol.* **178:** 1265–1273.

Kanehisa, M. and S. Goto. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28:** 27-30.

Kang, P. J. and E. A. Craig. 1990. Identification and characterization of a new *Escherichia coli* gene that is a dosage-dependent suppressor of a *dnaK* deletion mutation. *J. Bacteriol.* **172:** 2055–2064.

Kapperud, G., E. Namork, M. Skurnik, and T. Nesbakken. 1987. Plasmid-mediated surface fibrillae of *Yersinia pseudotuberculosis* and *Yersinia enterocolitica*: relationship to the outer membrane protein YOP1 and possible importance for pathogenesis. *Infect. Immun.* **55:** 2247–2254.

Kasarov, L. B. and A. G. Moat. 1973. Biosynthesis of NAD in *Haemophilus haemoglobinophilus*. *Biochim. Biophys. Acta* **320:** 372-378

Kato, S., M. Muro, S. Akifusa, N. Hanada, I. Semba, T. Fujii, Y. Kowashi, and T. Nishihara. 1995. Evidence for apoptosis of murine macrophages by *Actinobacillus actinomycetemcomitans* infection. *Infect. Immun.* **63:** 3914-3919.

Keener, J. and M. Nomura. 1996. Regulation of Ribosome Synthesis. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 1417-1431. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Kiley, P. and S. C. Holt. 1980. Characterization of the lipopolysaccharide from *Actinobacillus actinomycetemcomitans* Y4 and N27. *Infect. Immun.* **30:** 862-873.

Kisker C., H. Schindelin, and D. C. Rees. 1997. Molybdenum-cofactor-containing enzymes: structure and mechanism. *Annu. Rev. Biochem.* **66**: 233-267.

Kirby, A. C., S. Meghji, S. P. Nair, P. White, K. Reddi, T. Nishihara, K. Nakashima, A. C. Willis, R. Sim, M. Wilson, and B. Henderson. 1995. The potent bone-resorbing mediator of *Actinobacillus actinomycetemcomitans* is homologous to the molecular chaperon GroEL. *J. Clin. Invest.* **96:** 1185-1194.

Klena, J. D., E. Pradel, and C. A. Schnaitman. 1992. Comparison of lipopolysaccharide biosynthesis genes *rfaK*, *rfaL*, *rfaY*, and *rfaZ* of *Escherichia coli* K-12 and *Salmonella typhimurium*. *J. Bacteriol.* **174**: 4746-4752.

Koonin, E. V. and M. Y. Galperin. 1997. Prokaryotic genomes: the emerging paradigm of genome-based microbiology. *Curr. Opin. Genet. Develop.* **7: 757–763.**

Koonin, E. V., A. R. Mushegian, M. Y. Galperin, and D. R. Walker. 1997. Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel novel functions and suggests a chimeric origins of archaea. *Mol. Microbiol.* **25:** 619-637.

Koonin, E. V., R. L. Tatusov, and M. Y. Galperin. 1998. Beyond complete genome: from sequence to structure and function. *Curr. Opin. Struc. Biol.* **8:** 355-363.

Lachica, R. V. and D. L. Zink. 1984. Plasmid-associated cell surface charge and hydrophobicity of *Yersinia enterocolitica*. *Infect. Immun.* **44:** 540–543.

Lally, E. T., E. E. Golub, I. R. Kieba, N. S. Taichman, J. Rosenbloom, J. C. Roesenbloom, C. W. Gibson, and D. R. Demuth. 1989. Analysis of the *Actinobacillus actinomycetemcomitans* leukotoxin gene. Delineation of unique features and comparison to homologous toxins. *J. Biol. Chem.* **264:** 15451-15456.

Lander, E. S. and M. S. Waterman. 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2:** 231-239.

Lear, J. D., U. G. Furblur, E. T. Lally, and J. C. Tanaka. 1995. *Actinobacillus actinomycetemcomitans* leukotoxin forms large conductance, voltage-gated ion channels when incorporated into planar lipid bilayers. *Biochim. Biophys. Acta.* **1238:** 34-41.

Lee, L. G., C. R. Connell, S. L. Woo, R. D. Cheng, B. F. McArdle, C. W. Fuller, N. D. Halloran, and R. K. Wilson. 1992. DNA sequencing with dye-labeled terminators and T7 DNA polymerase: effect of dyes and dNTPs on incorporation of dye-terminators and probability analysis of termination fragments. *Nucleic Acids Res.* **20:** 2471-83.

Li, Q. and P. R. Reeves. 2000. Genetic variation of dTDP-L-rhamnose pathway genes in *Salmonella enterica. Microbiology* **146:** 2291-2307.

Loomer, P. M., R. P. Ellen, and H. C. Tenenbaum. 1995. Characterization of inhibitory effects of suspected periodontopathogens on osteogenesis *in vitro. Infect. Immun.* **63:** 3287-3296.

Lowe, T. M. and S. R. Eddy. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence, *NAR.* **25:** 955-964.

Mardis, E. R. and B. A. Roe. 1989. Automated methods for single-stranded DNA isolation and dideoxynucleotide DNA sequencing reactions on a robotic workstation. *Biotechniques* **7:** 736-746.

Martin, P. R., R. J. Shea, and M. H. Mulks. 2001. Identification of a plasmid-encoded gene from *Haemophilus ducreyi* which confers NAD independence. *J. Bacteriol.* **183:** 1168-1174.

Mead, D. A., J. A. McClary, J. A. Luckey, A. J. Kostichka, F. A. Whitney, and L. M. Smith. 1991. *Bst* DNA polymerase permits rapid sequence analysis from nanogram amounts of template. *Biotechniques* **11:** 76-87.

Maxam A. M. and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. (USA)* **74:** 560-4.

Ma, Y., R. J. Stern, M. S. Scherman, V. D. Vissa, W. Yan, V. C. Jones, F. Zhang, S. G. Franzblau, W. H. Lewis, and M. R. McNeil. 2001. Drug targeting *Mycobacterium tuberculosis* cell wall synthesis: genetics of dTDP-rhamnose synthetic enzymes and development of a microtiter plate-based screen for inhibitors of conversion of dTDP-glucose to dTDP-rhamnose. *Antimicrob. Agents Chemother.* **45:** 1407-1416.

Marinus, M. 1996. Methylation of DNA. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 782-791. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Mehta, P. K., T. I. Hale, and P. Christen. 1989. Evolutionary relationships among aminotransferases. Tyrosine aminotransferase, histidinol-phosphate aminotransferase, and aspartate aminotransferase are homologous proteins. *Eur. J. Biochem.* **186:** 249-253.

Mendz, G. L, A. J. Shepley, S. L. Hazell, and M. A. Smith. 1997. Purine metabolism and the microaerophily of *Helicobacter pylori. Arch. Microbiol.* **168:** 448-456.

Meyer, D. H. and P. M. Fives-Taylor. 1993. Evidence that extracellular components function in adherence of *Actinobacillus actinomycetemcomitans* to epithelial cells. *Infect. Immun.* **61**: 4933-4936.

Meyer, D. H., J. E. Lippmann, and P. M. Fives-Taylor. 1996. Invasion of epithelial cells by *Actinobacillus actinomycetemcomitans* : a dynamic, multistep process. *Infect. Immun.* **64**: 2988-2997.

Meyer, D. H., J. E. Rose, J. E. Lippmann, and P. M. Fives-Taylor. 1999. Microtubules are associated with intracellular movement and spread of the periodontopathogen *Actinobacillus actinomycetemcomitans. Infect. Immun.* **67:** 6518-6525.

Meyer, D. H., P. K. Sreenivasan, and P. M. Fives-Taylor. 1991. Evidence of invasion of a human oral cell line by *Actinobacillus actinomycetemcomitans. Infect. Immun.* **59:** 2719-2726.

Mills, D. R. and F. R. Kramer. 1979. Structure-independent nucleotide sequence analysis. *Proc. Natl. Acad. Sci. (USA)* **76:** 2232–2235.

Miller, C. G. 1996. Protein Degradation and Proteolytic Modification. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 938-954. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Miller, V. L. and S. Falkow. 1988. Evidence for two genetic loci in *Yersinia entercolitica* that can promote invasion of epithelial cells. *Infect. Immun.* **56:** 1242-1248.

Mintz, K. P. and P. M. Fives-Taylor. 1994. Adhesion of *Actinobacillus actinomycetemcomitans* to a human oral cell line. *Infect. Immun.* **62:** 3672-3678.

Mintz, K. P. and P. M. Fives-Taylor. 1994. Identification of an immunoglobulin Fc receptor of *Actinobacillus actinomycetemcomitans. Infect. Immun.* **62:** 4500-4505.

Mintz, K. P. and P. M. Fives-Taylor. 1999. Binding of the periodontal pathogen *Actinobacillus actinomycetemcomitans* to extracellular matrix proteins. *Oral Microbiol. Immunol.* **14:** 109-116.

Mitchell, S. J. and M. F. Minnick. 1995. Characterization of a two-gene locus from *Bartonella bacilliformis* associated with the ability to invade human erythrocytes. *Infect. Immun.* **63:** 1552-1562.

Mogull, S. A., L. J. Runyen-Janecky, M. Hong, and S. M. Payne. 2001. *dksA* is required for intercellular spread of *Shigella flexneri* via an RpoS-independent mechanism. *Infect. Immun.* **69:** 5742-5751.

Monack, D. M., J. Mecsas, N. Ghori, and S. Falkow. 1997. Yersinia signals macrophages to undergo apoptosis and YopJ is necessary for this cell death. *Proc. Natl. Acad. Sci. (USA)* **94:** 10385-90.

Morton, D. J., P. W. Whitby, H. Jin, Z. Ren, and T. L. Stull. 1999. Effect of multiple mutations in the hemoglobin- and hemoglobin-haptoglobin-binding proteins, HgpA, HgpB, and HgpC, of *Haemophilus influenzae* type b. *Infect. Immun.* **67(6):** 2729-2739.

Motz, M, S. Paabo, C. Kilger. 2000. Improved cycle sequencing of GC-rich templates by a combination of nucleotide analogs. *Biotechniques.* **29:** 268-270.

Moxon, E. R., D. Hood, and J. Richards. 1998. Bacterial lipopolysaccharides: candidate vaccines to prevent *Neisseria meningitidis* and *Haemophilus influenzae* infections. *Adv. Exp. Med. Biol.* **435:** 237-43.

Mullis, K. B. and Faloona, F. A. 1987. Specific synthesis of DNA *in vitro* via a polymerase catalyzed chain reaction. *Meth. Enzymol.* **155**: 335-350.

Murphy, C. K. and J. Beckwith. 1996. Export of Proteins to the Cell Envelope in *Escherichia coli.* In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 967-978. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Mushegian, A. R. and E. V. Koonin. 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. (USA)* **93:** 10268-10273.

Naider, F. and J. M. Becker. 1975. Multiplicity of oligopeptide transport systems in *Escherichia coli. J. Bacteriol.* **122:** 1208-1215.

Nair, S., E. Milohanic, and P. Berche. 2000. ClpC ATPase is required for cell adhesion and invasion of *Listeria monocytogenes. Infect. Immun.* **68:** 7061-7068.

Neidhardt, F. C. 1999. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, Second Edition. ASM press.

Neuhard, J. and R. A. Kelln. 1996. Biosynthesis and Conversions of Pyrimidines. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 580-599. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10:** 1-6.

Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27:** 29-34.

Ohkubo, S. and K. Yamaguchi. 1997. A suppressor of mutations in the region adjacent to the iterons of pSC101 *ori*. *J. Bacteriol.* **179:** 2089–2091.

Ohta, H., K. Fukui, and K. Kato. 1989. Effect of bicarbonate on the growth of *Actinobacillus actinomycetemcomitans* in anaerobic fructose-limited chemostat culture. *J. Gen. Microbiol.* **135:** 3485-3495.

Ohta, H., K. Kato, S. Kokeguchi, H. Hara, K. Fukui, and Y. Murayama. 1991. Nuclease-sensitive binding of an *Actinobacillus actinomycetemcomitans* leukotoxin to the bacterial cell surface. *Infect. Immun.* **59:** 4599-605.

Oresnik, I. J., L. C. Ladner, and R. J. Turner. 2001. Identification of a twin-arginine leader-binding protein. *Molec. Microbiol.* **40:** 323-331.

Page, R. C. and H. E. Schroeder. 1973. Biochemical aspects of the connective tissue alterations in inflammatory gingival and peridontal disease. *Int. Dent. J.* **23:** 455-469.

Paju, S., F. Goulhen, S. Asikainen, D. Grenier, D. Mayrand, and V.-J., Uitto. 2000. Localization of heat shock proteins in clinical *Actinobacillus actinomycetemcomitans* strains and their effects on epithelial cell proliferation. *FEMS Microbiol. Lett.* **182:** 213-235.

Parsell, D. A., Y. Sanchez, J. D. Stitzel, and S. Lindquest. 1991. Hsp104 is a highly conserved protein with two essential nucleotide-binding sites. *Nature* **353:** 270-273.

Parsons, J. D. 1995. Improved tools for DNA comparison and clustering. *Comput. Applic. Biosci.* **11:** 603-613.

Parsons, J. D. 1995. Miropeats: graphical DNA sequence comparisons. *Comput. Applic. Biosci.* **11:** 615-619.

Parsons, L. M., R. J. Limberger, and M. Shayegeni. 1997. Alterations in levels of DnaK results in diminished survival and adherence of stressed *Haemophilus ducreyi*. *Infect. Immun.* **65**: 2413-2419.

Patzer, S. I. and K. Hantke. 1998. The ZnuABC high-affinity zinc uptake system and its regulator Zur in *Escherichia coli*. *Mol. Microbiol.* **28:** 1199-1210.

Peak, I. R., Y. Srikhanta, M. Dieckelmann, E. R. Moxon, and M. P. Jennings. 2000. Identification and characterisation of a novel conserved outer membrane protein from *Neisseria meningitidis*. *FEMS Immunol. Med. Microbiol.* **28:** 329-334.

Pederson, K. J., S. Carlson, and D. E. Pierson. 1997. The ClpP protein, a subunit of the Clp protease, modulates *ail* gene expression in *Yersinia entercolitica*. *Mol. Microbiol.* **26:** 99-107.

Pierson, D. E. and S. Falkow. 1993. The *ail* gene of *Yersinia entercolitica* has a role in the ability of this organism to survive serum killing. *Infect. Immun.* **61:** 1846-1852.

Postma, P. W., J. W. Lengeler, and G. R. Jacobson. 1996. Phosphoenolpyruvate: Carbohydrate Phosphotransferase Systems. In: *Escherichia coli and Salmonella Cellular and Molecular Biology*. Vol. 1. pp. 1149-1174. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Poulsen, K., E. Theilade, E. T. Lally, D. R. Demuth, and Kilian, M. 1994. Population structure of *Actinobacillus actinomycetemcomitans*: a framework for studies of disease-associated properties. *Microbiology*. **140:** 2049-2060.

Poyart, C., E. Pellegrini, O. Gaillot, C. Boumaila, M. Baptista, and P. Trieu-Cuot. 2000. Contribution of Mn-cofactored superoxide dismutase (SodA) to the virulence of *Streptococcus agalactiae*. *Infect. Immun.* **69:** 5098-5106.

Pugsley, A. P. 1993. The complete general secretory pathway in Gram negative bacteria. *Microb. Review.* **57:** 50-108.

Raetz, C. R. H. 1996. Structure and Biosynthesis of Lipid A. In: *Escherichia coli and Salmonella Cellular and Molecular Biology*. Vol. 1. pp. 498-503. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Raffaelli, N., T. Lorenzi, P. L. Mariani, M. Emanuelli, A. Amici, S. Ruggieri, and G. Magni. 1999. The *Escherichia coli* NadR regulator is endowed with nicotinamide mononucleotide adenyltransferase activity. *J. Bacteriol.* **181:** 5509-5511.

Rajagopalan, K. V. 1996. Biosynthesis of the Molybdenum Cofactor. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 674-679. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Redaschi, N. and T. A. Bickle. 1996. DNA Restriction and Modification Systems. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 773-781. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Reitzer, L. J. 1996. Ammonia Assimilation and the Biosynthsis of Glutamine, Glutamate, Aspartate, Asparagine, L-Alanine, and D-Alanine. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 391-407. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Ren, Z., H. Jin, P. W. Whitby, D. J. Morton, and T. L. Stull. 1999. Role of CCAA nucleotide repeats in regulation of hemoglobin and hemoglobin-haptoglobin binding protein genes of *Haemophilus influenzae. J. Bacteriol.* **181:** 5865–5870.

Rietschel, E. T., T. Kirikae, F. U. Schade, U. Mamat, G. Schmidt, H. Loppnow, A. J. Ulmer, U. Zahringer, U. Seydel, F. DiPadova, M. Schreir, and H. Brade. 1994. Bacterial enotoxin: molecular relationships of structure to activity and function. *FASEB J.* **8:** 217-225.

Riley, M. and B. Labedan. 1996. *E. coli* Gene Products. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 2118-2202. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Robertson, P. B., M. Lantz, P. T. Marucha, K. S. Kornman, C. L. Trummel, and S. C. Holt. 1982. Collagenolytic activity associated with Bacteroides species and *Actinobacillus actinomycetemcomitans. J. Periodontal Res.* **17:** 275-283.

Roggenkamp, A, K. Ruckdeschel, L. Leitritz, R. Schmitt, and J. Heesemann. 1996. Deletion of amino acids 29 to 81 in adhesion protein YadA of *Yersinia enterocolitica* serotype O:8 results in selective abrogation of adherence to neutrophils. *Infect. Immun.* **64:** 2506-14

Rosenblum, B. B., L. G. Lee, S. L. Spurgeon, S. H. Khan, S. M. Menchen, C. R. Heiner, and S. M. Chen. 1997. New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res.* **25:** 4500-4504.

Rouquette, C., C. de Chastellier, S. Nair, and P. Berche. 1998. The ClpC ATPase of *Listeria monocytogenes* is a general stress protein required for virulence and promoting early bacterial escape from the phagosome of macrophages. *Mol. Microbiol.* **27:** 1235-1245.

Ruffolo, C. G. , and B. Adler. 1996. Cloning, Sequencing, Expression, and Protective Capacity of the *oma87* Gene Encoding the *Pasteurella multocida* 87-Kilodalton Outer Membrane Antigen. *Infect. Immun.* **64:** 3161–3167

Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.-A. Rajandream and B. Barrell. 2000. Artemis: sequence visualisation and annotation. *Bioinformatics* **16:** 944-945.

Sanger, F. and A. R. Coulson. 1978. The use of thin acrylamide gels for DNA sequencing. *FEBS Lett.* **87:** 107-110.

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

Samuelson, J. C., M. Chen, F. Jiang, I. Moller, M. Wiedmann, A. Kuhn, G. J. Phillips, and R. E. Dalbey. 2000. YidC mediates membrane protein insertion in bacteria. *Nature* **406:** 637-641.

Salyers, A. A. and D. D. Whitt. 1994. Bacterial Pathogenesis, a Molecular Approach. ASM Press, Washington, D. C.

Selkov, E., M. Galimova, G. I. Goryanin, Y. Gretchkin, N. Ivanova, Y. Komarov, N. Maltse, N. Mikhailova, V. Nenashev 1, R. Overbeek, E. Panyushkina 1, L. Pronevitch and E. Selkov Jr. 1997. The metabolic pathway collection: an update. *NAR.* **25:** *37–38.*

Schriefer, L. A., B. K. Genauer, L. Q. Q. Qin, R. H. Waterson, and R. K. Wilson. 1990. Low pressure DNA shearing: a method for random DNA sequence analysis. *Nucleic Acids Res.* **18:** 7455-7464.

Schenkein, H. A., S. E. Barbour, C. R. Berry, B. Kipps, and J. G. Tew. 2000. Invasion of human vascular endothelial cells by *Actinobacillus actinomycetemcomitans* via the receptor for platelet-activating factor. *Infect. Immun.* **68:** 5416-5419.

Schryvers, A. B. and I. Stojiljkovic. 1999. Iron acquisition systems in the pathogenic *Neisseria. Molec. Microbiol.* **32(6):** 1117-1123.

Schulze-Koops, H., H. Burkhardt, J. Heesemann, K. von der Mark, and F. Emmrich. 1992. Plasmid-encoded outer membrane protein YadA mediates specific

binding of enteropathogenic yersiniae to various types of collagen. *Infect. Immun.* **60:** 2153–2159.

Schweder, T., K.-H. Lee, O. Lomovskaya, and A. Matin. 1996. Regulation of *Escherichia coli* starvation sigma factor (σ^S)by ClpXP protease. *J. Bacteriol.* **178:** 470–476.

Selkov, E. Jr., Y. Grechkin, N. Mikhailova, and E. Selkov. 1998. MPW: the metabolic pathways database. *NAR* **26**: 43-45.

Shao, Z., R. T. Lin, and E. B. Newman. 1994. Sequencing and characterization of the *sdaC* gene and identification of the *sdaCB* operon in *Escherichia coli* K12. *Eur. J. Biochem.* **222:** 901-907.

Silver, S. 1996. Transport of Inorganic Cations. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 217-261. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Sloth, J., H. S. Reynolds, and R. J. Genco. 1980. *Actinobacillus actinomycetemcomitans* in human periodontal disease: a cross-sectional microbiological investigation. *Infect. Immun.* **29:** 1013-1020.

Slots, J. 1982. Salient biochemical characters of *Actinobacillus actinomycetemcomitans. Arch. Microbiol.* **131:** 60-67.

Smith, L. M., J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connel, C. Heiner, S. B. H. Kent, and L. E. Hood. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* **321:** 674-679.

Smith, L. M., R. J. Kaiser, J. Z. Sanders, and L. E. Hood. 1987. The synthesis and use of fluorescent oligonucleotides in DNA sequence analysis. *Meth. Enzymol.* **155:** 260-301.

Soloaga, A., M. P. Veiga, L. M. Garcia-Segura, H. Ostolaza, R. Brasseur, and F. M. Goni. 1999. Insertion *of Escherichia coli* _-hemolysin in lipid bilayers as a non-transmembrane integral protein: prediction and experiment. *Mol. Microbiol.* **31:** 1013-1024.

Sonnhammer, E. L. and D. Kahn. 1994. Modular arrangement of proteins as inferred from analysis of homology. *Protein Sci.* **3:** 482-492.

Squires, and M. R. Maurizi. 1990. Conservation of the regulatory subunit for the Clp ATP-dependent protease in prokaryotes and eukaryotes. *Proc. Natl. Acad. Sci. (USA)* **87:** 3513-3517.

Squires, C. and C. L. Squires. 1992. The Clp proteins: proteolysis regulators or molecular chaperones? *J. Bacteriol.* **174:** 1081-1085.

Squires, C. L., S. Pedersen, B. M. Ross, and C. Squires. 1991. ClpB in *Escherichia coli* heat shock protein F84.1. *J. Bacteriol.* **173:** 4254-4262.

Stanley, P., V. Koronakis, K. Hardie, and C. Hughes. 1998. Acylation of *Esherichia coli* hemolysin: a unique protein lipidation mechanism underlying toxin function. *Microbiol. Mol. Biol. Rev.* **62:** 309-333.

Staudenmaier, H., B. Van hove, Z. Yaraghi, and V. Braun. 1989. Nucleotide sequences of the *fecBCDE* genes and locations of the proteins suggest a periplasmic-binding-protein-dependent transport mechanism for iron(III) dicitrate in *Escherichia coli*. *J. Bacteriol.* **171:** 2626-2633.

Stec, B., H. Yang, K. A. Johnson, L. Chen, and M. F. Roberts. 2000. MJ0109 is an enzyme that is both an inositol monophosphatase and the 'missing' archaeal fructose-1,6-bisphosphatase. *Nature Struct. Biol.* **7:** 1046-50.

Stenesh, J. and B. A. Roe. 1972. DNA polymerase from mesophilic and thermophilic bacteria. I. Purification and properties of DNA polymerase from *Bacillus licheniformis* and *Bacillus stearothermophilus*. *Biochim. Biophys. Acta* **272:** 156-166.

Sugai, M., T. Kawamoto, S. Y. Peres, Y. Ueno, H. Komatsuzawa, T. Fujiwara, H. Kurihara, H. Suginaka, and E. Oswald. 1998. The cell-specific growth-inhibitory factor produced by *Actinobacillus actinomycetemcomitans* is a cytolethal distending toxin. *Infect. Immun.* **66:** 5008-5019.

Tabor, S. and C. C. Richardson. 1987. DNA sequence anlaysis with a bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. (USA)* **84:** 4767-4771.

Tabor, S. and C. C. Richardson. 1995. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. (USA)* **92:** 6339-6343.

Tamm, A., A. M. Tarkkanen, T. K. Korhonen, P. Kuusela, P. Toivanen, and M. Skurnik. 1993. Hydrophobic domains affect the collagen-binding specificity and surface polymerization as well as the virulence potential of the YadA protein of *Yersinia enterocolitica*. *Mol. Microbiol.* **10:** 995–1011.

Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. *Science* **278:** 631-637.

Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova, and E. V. Koonin. 2001.

The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29:** 22-28.

Tatusova, T. A. and T. L. Madden. 1999. Blast 2 sequences - a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174:** 247-250.

Tertti, R., M. Skurnik, T. Vartio, and P. Kuusela. 1992. Adhesion protein YadA of *Yersinia* species mediates binding of bacteria to fibronectin. *Infect. Immun.* **60:** 3021–3024.

Tettelin, H., D. Radune, S. Kasif, H. Khouri, and S. L. Salzberg. 1999. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics* **15:** 500-507.

Thomas, D. C., M. Levy, and A. Sancar. 1985. Amplification and Purification of UvrA, UvrB, and UvrC proteins of *Escherichia coli*. *J. Biol. Chem.* **260:** 9875-9883.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673-4680.

Ting, M., A. Contreras, and J. Slots. 2000. Herpes virus in localized juvenile periodontitis. *J. Periodontal Res.* **35:** 17-25.

Tillotson, G. S., and S. J. Watson. 2001. Antimicrobial resistance mechanisms: what's hot and what's not in respiratory pathogens. *Semin. Respir. Infect.* **16:** 155-168.

Tjalsma, H., A. Bolhuis, J. D. H. Jongbloed, S. Bron, and J. M. VanDijl. 2000. Signal Peptide-Dependent Protein Transport in *Bacillus subtilis*: A genome-based survey of the secretome. *Microbiol. Mol. Biol. Rev.* **64:** 515-547.

Tomasz, A. 1967. Choline in the cell wall of bacterium: novel type of polymer-linked choline in *Pneumococcus*. *Science* **157:** 694-697.

Tomb, J.-F., O. White, A. R. Kerlavage, R. A. Clayton, G. G. Sutton, R. D. Fleischmann, K. A. Ketchum, H. P. Klenk, S. Gill, B. A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E. F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H. G. Khalak, A. Glodek, K. McKenney, L. M. Fitzegerald, N. Lee, M. D. Adams, E. K. Hickey, D. E. Berg, J. D. Gocayne, T. R. Utterback, J. D. Peterson, J. M. Kelley, M. D. Cotton, J. M. Weidman, C. Fujii, C. Bowman, L. Watthey, E. Wallin, W. S. Hayes, M. Borodovsky, P. D. Karp, H. O. Smith, C. M. Fraser, and J. C. Venter. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388:** 539–547.

Tsukioka, Y., Y. Yamashita, Y. Nakano, T. Oho, and T. Koga. 1997. Identification of a fourth gene involved in dTDP-rhamnose synthesis in *Streptococcus mutans*. *J. Bacteriol.* **179:** 4411-4414

Turner, A. K., A. M. Lovell, S. D. Hulme, L. Zhang-Barber, and P. A. Barrow. 1998. Identification of *Salmonella typhimurium* genes required for colonization of the chicken alimentary tract and for virulence in newly hatched chicks. *Infect. Immun.* **66:** 2099-2106.

Umbarger, H. E. 1996. Biosynthesis of Branched-Chain Amino Acids. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 442-457. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Vallari, D. S. and C. O. Rock. 1985. Pantothenate transport in *Escherichia coli*. *J. Bacteriol.* **162:** 1156-1161.

Vallari, D. S. and C. O. Rock. 1985. Isolation and characterization of *Escherichia coli* pantothenate permease (panF) mutants. *J. Bacteriol.* **164:** 136-142.

Van Belkum, A., S. Scherer, L. Van Alphen, and H. Verbrugh. 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol. Molec. Biol. Rev.* **62:** 275-293.

Van Buul, C. P., J. J., and P. H. van Knippenberg. 1985. Nucleotide sequence of the *ksgA* gene of *Escherichia coli*: comparison of methyltransferases effecting dimethylation of adenosine in ribosomal RNA. *Gene* **38:** 65-72.

Van Heijenoort, J. 1996. Murein Synthesis. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 1025-1034. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Van Steenbergen, T. J., C. J. Bosch-Tijbof, A. J. Van Winkelhoff, R. Gmur, and J. de Graaff. 1994. Comparison of six typing methods for *Actinobacillus actinomycetemcomitans*. *J. Clin. Microbiol.* **32:** 2769-2774.

Vaara M. 1993. Outer membrane permeability barrier to azithromycin, clarithromycin, and roxithromycin in Gram negative enteric bacteria. *Antimicrob. Agents Chemother.* **37:** 354-356.

Visser, L. G., P. S. Hiemstra, M. T. Van Den Barselaar, P. A. Ballieux, and R. Van Furth. 1996. Role of YadA in resistance to killing of *Yersinia enterocolitica* by antimicrobial polypeptides of human granulocytes. *Infect. Immun.* **64:** 1653–1658.

Von Heijne, G. 1984. Analysis of the distribution of charged residues in the N-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells. *EMBO. J.* **3:** 2315-2318.

Von Heijne, G. 1984. How signal sequences maintain cleavage specificity. *J. Mol. Biol.* **173:** 243-351.

Voss, H., S. Wiemann, D. Grothues, C. Sensen, J. Zimmermann, C. Schwager, J. Stegemann, H. Erfle, T. Rupp, and W. Ansorge. 1993. Automated low-redundancy large scale DNA sequencing by primer walking. *Biotechniques* **15:** 714-721.

Wagegg, W., and V. Braun. 1981. Ferric citrate transport in *Escherichia coli* requires outer membrane receptor protein *fecA*. *J. Bacteriol.* **145:** 156-163.

Wandersman, C. 1996. Secretion across the Bacterial Outer Membrane. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 955-966. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Wang, P.-L., S. Shirasu, M. Shinohara, M. Daito, T. Fujii, Y. Kowashi, and K. Ohura. 1999. Purification and characterization of a trypsin-like protease from the culture supernatant of *Actinobacillus actinomycetemcomitans* Y4. *Eur. J. Oral Sci.* **107:** 147-153

Watson, J. D. and F. H. Crick. 1953. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171:** 737-738.

Webb, E., K. Claas, and D. Downs. 1997. thiBPQ encodes an ABC transporter required for transport of thiamine and thiamine pyrophosphate in *Salmonella typhimurium*. *J. Biol. Chem.* **273:** 8946-8950.

Welch, R. A. 1991. Pore-forming cytolysins of Gram negative bacteria. *Molec. Microbiol.* **5:** 521-528

Weiss, A. A. and E. L. Hewlett. 1986. Virulence factors Of *Bordetella pertussis*. *Ann. Rev. Microbiol.* **40**: 661-686.

West, D., K. Reddin, M. Matheson, R. Heath, S. Funnell, M. Hudson, A. Robinson, and A. Gorringe. 2001. Recombinant *Neisseria meningitidis* transferrin binding protein A protects against experimental meningococcal infection. *Infect. Immun.* **69:** 1561-1567.

White, D. 2000. The Physiology and Biochemistry of Prokaryotes. Oxford University Press, New York, NY.

White, P. A., S. P. Nair, M-J. Kim, M. Wilson, and B. Henderson. 1998. Molecular characterization of an outer membrane protein of *Actinobacillus actinomycetemcomitans* belonging to the OmpA family. *Infect. Immun.* **66:** 369-372.

Weiser, J. N., N. Pan, K. L. McGowan, D. Musher, A. Martin, and J. Richards. 1998. Phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* contributes to persistence in the respiratory tract and sensitivity to serum killing mediated by C-reactive protein. *J. Exp. Med.* **187:** 631–640.

Welz, D., and V. Braun. 1998. Ferric Citrate Transport of *Escherichia coli*: Functional Regions of the FecR Transmembrane Regulatory Protein. *J Bacteriol.* **180:** 2387-2394.

White, P. A., S. P. Nair, M.-J. Kim, M. Wilson, and B. Henderson. 1998. Molecular characterization of an outer membrane protein of *Actinobacillus actinomycetemcomitans* belonging to the OmpA family. *Infect. Immun.* **66:** 369-372.

White, R. L. and I. D. Spenser. 1996. Biosynthesis of Thiamin. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 680-686. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Wilde, C., S. Bachellier, M. Hofnung, and J. M. Clement. 2001. Transposition of IS1397 in the family *Enterobacteriaceae* and first characterization of ISKpn1, a new insertion sequence associated with *Klebsiella pneumoniae* palindromic units. *J. Bacteriol.* **183:** 4395-4404.

Wissenbach, U., S. Six, J. Bongaerts, D. Ternes, S. Steinwachs, and G. Unden. 1995. A third periplasmic transport system for L-arginine in *Escherichia coli*: molecular characterization of the artPIQMJ genes, arginine binding and transport. *Mol. Microbiol.* **17:** 675-686.

Wiemann, S., H. Voss, C. Schwager, T. Rupp, J. Stegemann, J. Zimmermann, D. Grothues, C. Sensen, H. Erfle, A. Banrevi, and W. Ansorge. 1993. Sequencing and analysis of 51.6 kilobases on the left arm of chromosome XI from *Saccharomyces cerevisiae* reveals 23 open reading frames including the fas1 gene. *Yeast* **9:** 1343-1348.

Wilcox, M. and M. Nirenberg. 1968. Transfer RNA as a cofactor coupling amino acid synthesis with that of protein. *Proc. Natl. Acad. Sci. (USA)* **61:** 229-236.

Willemsen, P. T. J., I. Vulto, M. Boxem, and J. DeGraaff. 1997. Characterization of periplasmic protein involved in iron utilization of *Actinobacillus actinomycetemcomitans*. *J. Bacteriol.* **179:** 4949-4952.

Yamada, R. H. and Y. Furukawa. 1980. Apparent pyridoxine transport mutants of *Escherichia coli* with pyridoxal kinase deficiency. *Biochim. Biophys. Acta* **600:** 581-584.

Zalkin, H. and P. Nygaard. 1996. Biosynthesis of Purine Nucleotides. In: *Escherichia coli and Salmonella Cellular and Molecular Biology.* Vol. 1. pp. 561-579. Neidhardt, F. C., R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (Eds.). American Society for Microbiology, Washington, D. C.

Zambon, J. J. 1985. *Actinobacillus actinomycetemcomitans* in human periodontal disease. *J. Clin. Periodontol.* **12:** 1-20.

Zambon, J. J., J. Sloth, and R. J. Genco. 1983. Serology of oral *Actinobacillus actinomycetemcomitans* and serotype distribution in human periodontal disease. *Infect. Immun.* **41:** 19-27.

Zhang, J. R., I. Idanpaan-Heikkila, and E. I. Tuomanen. 1999. Pneumococcal *lic*D2 gene is involved in phosphorylcholine metabolism. *Molec. Microbiol.* **31:** 1477-1488.

Zhang, L., P. Toivanen, and M. Skurnik. 1996. The gene cluster directing O-antigen biosynthesis in *Yersinia enterocolitica* serotype 0:8: identification of the genes for mannose and galactose biosynthesis and the gene for the O-antigen polymerase. *Microbiology* **142:** 277-88.

Zhou, D., W.-D. Hardt, and J. E. Galan. 1999. *Salmonella typhimurium* encodes a putative iron transport system within the centisome 63 pathogenicity island. *Infect. Immun.* **67:** 1974–1981.

Appendix A: *A. actionomycetemcomitans* ORFs

1. Small-molecule Metabolism

1.1 Degradation

1.1.1 Carbon Compounds

EC 1.1.1.83	D-Malate dehydrogenase (decarboxylating) (malS)
EC 1.2.1.27	Methylmalonate-semialdehyde dehydrogenase (acylating)
EC 3.5.1.27	N-Formylmethionylaminoacyl-tRNA deformylase
EC 3.1.3.18	Phosphoglycolate phosphatase
EC 2.4.1.-	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase (murG)
EC 4.2.1.-	(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase (fabZ, sefA)
EC 1.1.1.22	UDP glucose 6-dehydrogenase
EC 4.1.1.3	Oxaloacetate decarboxylase - alpha chain (Oxaloacetate beta-decarboxylase)
EC 4.1.1.3	Oxaloacetate decarboxylase - beta chain (Oxaloacetate beta-decarboxylase)
EC 1.1.1.21	Aldehyde reductase (Aldose reductase) (Polyol dehydrogenase(NADP+))
EC 5.3.1.5	Xylose isomerase
EC 5.1.3.1	Ribulose-phosphate 3-epimerase (Phosphoribulose epimerase) (Erythrose-4-phosphate epimerase)
EC 2.7.1.53	L-Xylulokinase (lyx)
EC 5.1.3.4	L-Ribulosephosphate 4-epimerase
EC 4.1.2.17	L-Fuculose-1-phosphate aldolase
EC 5.1.3.2	UDP-glucose 4-epimerase (galE)
EC 2.7.1.6	Galactokinase
galR	Repressor for galETK operon
EC 2.7.7.12	Galactose-1-phosphate uridylyltransferase
EC 2.7.7.9	Glucose-1-phosphate uridyltransferase
EC 1.1.1.-	Galactitol-1-phosphate dehydrogenase (gatD)
gatR	Galactitol utilization operon repressor
gat	Galacticol-permease IIB component) (phosphotransferase enzyme II, B component)

EC 2.7.1.12	Gluconokinase, thermosensitive	
EC 2.7.1.45	2-keto-3-deoxygluconokinase	
EC 2.3.1.18	Thiogalactoside acetyltransferase	
lacI	Repressor for lac operon	Possible secretory protein
malM	Periplasmic protein for mal regulon	
EC 2.4.1.1	Maltodextrin phosphorylase	
malT	Positive regulator for mal regulon	
EC 3.2.1.20	Maltodextrin glucosidase	
EC 5.3.1.8	Mannose-6-phosphate isomerase	
EC 2.3.1.8	Phosphotransacetylase activity;phosphate acetyltransferase	Possible secretory protein
EC 2.7.1.15	Ribokinase (rbsK)	
EC 5.3.1.12	Altronate isomerase	
EC 4.2.1.8	D-Mannonate hydrolase	
uxuR	Regulator of uxuAB operon	
EC 2.7.1.17	Xylulokinase (xylB)	
xylR	Regulator of xyl operon, putative	
EC 6.2.1.22	Citrate lyase ligase (citC)	
citG	CitG protein	
glcK	Glucose kinase	
EC 3.2.1.-	beta-hexosaminidase (exoII)	
dxs	1-deoxyxylulose-5-phosphate synthase	
ydbE	C4-dicarboxylate transport system	Probable secretory protein
dctQ	C4-dicarboxylate permease	
YgiK	C4-dicarboxylate transport system (permease large protein)	
EC 3.1.1.-	para-nitrobenzyl esterase; Chain A, Thermophylic Pnb Esterase	
EC 3.1.1.-	type B carboxylesterase; para-nitrobenzyl esterase	
sgbH	Hexulose-6-phosphate synthase	
sgbU	Hexulose-6-phosphate isomerase	
I.1.2 Amino acids and amines		
EC 3.5.1.1	periplasmic L-asparaginaseII	
EC 5.1.1.1	Alanine racemase	
EC 4.2.1.13	Serine deaminase, L-SD2	
EC 4.2.1.16	Threonine dehydratase, catabolic	

EC 2.7.2.2	Carbamate kinase
EC 4.1.99.2	Tyrosine phenol-lyase(beta-Tyrosinase)
EC 1.4.7.1	Glutamate synthase (ferredoxin)(Ferredoxin-dependent glutamate synthase)
EC 4.4.1.1	Cystathionine gamma-lyase(Homoserine deaminase)
1.1.3 Fatty acids	
EC 2.8.3.9	Butyrate--acetoacetate CoA-transferase large chain (atoA)
EC 2.8.3.9	Butyrate--acetoacetate CoA-transferase (atoA)
EC 2.3.1.9	Acetyl-CoA acetyl transferase (atoB)
atoE	Short-chain fatty acids transporter
EC 6.2.1.3	Acyl-CoA synthetase
1.1.4 Phosphorus Compounds	
EC 3.1.1.5	Lysophospholipase L(2)
EC 3.6.1.11	Exopolyphosphatase (metaphosphatase)
1.2 Energy Metabolism	
1.2.1 Glycolysis	
EC 4.2.1.11	Enolase
EC 4.1.2.13	Fructose-bophsphate aldolase, classII
EC 2.7.1.56	Fructose-1-phosphate kinase
fruR	Repressor of fru operon and others
EC 1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase A
EC 5.4.2.1	Phosphoglyceromutase
EC 5.4.2.1	Phosphoglyceromutase
EC 2.7.1.11	6-Phosphofructokinase I
EC 5.3.1.9	Glucosephosphate isomerase
EC 2.7.2.3	Phosphoglycerate kinase
EC 2.7.1.40	Pyruvate kinase II, glucose stimulated
EC 5.3.1.1	Triosphosphate isomerase
1.2.2 Pyruvate DH.	
EC 1.2.4.1	Pyruvate dehydrogenase (decarboxylase component)
EC 2.3.1.12	Pyruvate dehydrogenase (ddihydrolipoyltransacetylase component)
EC 1.8.1.4	Lipoamide dehydrogenase (NADH)
1.2.3 TCA cycle	
EC 4.2.1.2	Fumarase C (fumarate hydratase class II, isozyme)

EC 1.1.1.37	Malate dehydrogenase
EC 1.3.99.1	Succinate dehydrogenase, flavoprotein subunit
EC 1.3.99.1	Succinate dehydrogenase, iron sulfur protein
EC 1.3.99.1	Succinate dehydrogenase, cytochrome b556
EC 1.3.99.1	Succinate dehydrogenase, hydrophobic subunit
EC 2.3.1.61	Dihydrolipoamide S-succinyltransferase, dihydrolipoyltranssuccinate component
EC 6.2.1.5	Succinyl-CoA synthetase, beta subunit (succinate thiokinase)
EC 6.2.1.5	Succinyl-CoA synthetase, alpha subunit
EC 4.1.3.6	Citrate lyase (CITRATE COA-TRANSFERASE SUBUNIT); alpha subunit
EC 4.1.3.6	Citrate lyase (citryl-CoA lyase subunit); beta subunit
EC 1.2.4.2	Oxoglutarate dehydrogenase (lipoamide)(Oxoglutarate decarboxylase)
1.2.4 Pentose phosphate pathway	
EC 1.1.1.44	Gluconate-6-phosphate dehydrogenase, decarboxylase
EC 1.1.1.49	Glucose-6-phosphate dehydrogenase
EC 1.1.1.43	Phosphogluconate 2-dehydrogenase(6-Phosphogluconic dehydrogenase)
EC 3.1.1.31	6-Phosphogluconolactonase
EC 5.3.1.6	Ribosephosphate isomerase A
EC 2.2.1.2	Transaldolase B
EC 2.2.1.1	Transketolase
EC 2.2.1.1	Transketolase isozyme
1.2.5 Enter-Doudoroff pathway	
EC 4.1.2.14	2-keto-3-deoxygluconate 6-phosphate aldolase (KDPG aldolase)
gntR	Regulator of edd; transport and phosphorylation of gluconate
1.2.6 Respiration	
1.2.6.a Aerobic	
EC 1.10.3.-	Cytochrome o ubiquinol oxidase subunit II -- Found in ETC
EC 1.10.3.-	Cytochrome o ubiquinol oxidase subunit I -- Found in ETC
EC 1.1.99.5	sn-Glycerol-3-phosphate dehydrogenase (acrobic)
EC 1.18.99.1	Hydrogenase-1, small subunit, FORMATE HYDROGENLYASE SUBUNIT 2 (FHL SUBUNIT 2)
EC 1.6.99.3	Respiratory NADH dehydrogenase
EC 1.6.5.3	NADH dehydrogenase I chain B (nqrB)
EC 1.6.5.3	NADH dehydrogenase I chain E

nifC	nitrogen fixation protein; nitrogenase C	
nifR3	Nitrogen fixation protein	
rnfC	Nitrogen fixation protein	
rnfD	Nitrogen fixation protein. Possible membrane protein	
rnfG	Nitrogen fixation protein	
1.2.6.b Anaerobic		
EC 1.8.99.-	Anaerobic dimethyl sulfoxide reductase subunit A	Possible membrane protein
EC 1.8.99.-	dimethyl sulphoxide reductasesubunit C	Possible membrane protein
EC 1.8.99.-	ANAEROBIC DIMETHYL SULFOXIDEREDUCTASE CHAIN B (DMSO REDUCTASE IRON-SULFUR SUBUNIT)	
fdhD	Affects formate dehydrogenase-N	
fdhE	Affects formate dehydrogenase-N	
EC 1.2.1.2	Selenopolypeptide subunit of formate dehydrogenase H	Probable membrane protein
EC 1.2.1.2	Formate dehydrogenase-N, nitrate inducible, major subunit	Probable membrane protein
EC 1.2.1.2	Formate dehydrogenase-N, nitrate inducible, iron-sufur subunit	Probable membrane protein
EC 1.2.1.2	Formate dehydrogenase-N, nitrate inducible, cytochrome B556 (Fdn) subunit	Probable membrane protein
EC 1.3.99.1	Fumarate reductase (succinate dehydrogenase), anaerobic, iron-sulfur subunit -- Found in "TCA Cycle"	
EC 1.3.99.1	Fumarate reductase (succinate dehydrogenase), anaerobic,membrane anchor polypeptide -- Found in "TCA cycle"	
EC 1.3.99.1	Fumarate reductase (succinate dehydrogenase), anaerobic,membrane anchor polypeptide -- Found in "TCA cycle"	
EC 1.1.99.5	Glycerol-3-phosphate dehydrogenase (anaerobic) -- Found in "Aerobic"	
glpE	Protein of glp regulon	Probable secretory protein
glpG	Protein of glp regulon. Rhomboid family	Probable membrane protein
glpR	Repressor of the glp operon	
glpX	GlpX prortein	
hypA	pleotropic effect of three hydrogenaseisozymes	
hypD	pleotropic effect of three hydrogenaseisozymes	
hypE	Structural role in hydrogenase synthesis	
hypF	HYDROGENASE MATURATION PROTEIN	
EC 1.7.99.4	Nitrate reductase, alpha subunit	
narP	Nitrate/nitrite response regulator	
EC 2.7.3.-	narQ Sensor for nitrate reductase system, protein histidine kinase	

napB	Periplasmic nitrate reductase (napB)	
napH	Ferredoxin-type protein	
napG	Ferredoxin-type protein (napG)	
napC	Cytochrome C-type protein (nirT)	
nrfA	Formate-dependent nitrite reductase, tetra heme cytochrome c552	Probable membrane protein
nrfB	Formate-dependent nitrite reductase, a penta heme cytochrome c	
nrfC	Formate-dependent nitrite reductase,Fe-S center	
nrfD	Formate-dependent nitrite reductase complex,transmembrane protein	Probable membrane protein
nrfE	Formate-dependent nitrite reductase,assembly function?	Probable membrane protein
nrfF	Part of Formate-dependent nitrate reductase complex	
torC	Trimethylamine N-oxide reductase, cytochrome subunit	
dcuC	Anaerobic carrier for c4, dicarboxylates	Probable membrane protein
1.2.6.c Electron transport		
EC 2.7.2.1	Acetate kinase; (ackA)	
EC 1.10.3.-	Cytochrome d terminal oxidase, subunit I (cydA)	Probable membrane protein
EC 1.10.3.-	Cytochrome d terminal oxidase, subunit II (cydB)	Probable membrane protein
fldA	Flavodoxin	
EC 1.18.1.2	Ferredoxin-NADP reductase (fpr)	
EC 1.6.6.-	NAD(P)H-flavin oxidoreductase	
EC 3.6.1.1	Inorganic pyrophosphatase	
EC 1.6.5.3	NADH dehydrogenase (ubiquinone)(Ubiquinone reductase)(Type I dehydrogenase)(Complex I dehydrogenase) -- Found in "Aerobic"	
EC 1.-.-.-	Na+-translocating NADH-ubiquinoneoxidoreductase (beta chain)(nqr6)	Probable secretory protein
nqrA	Na(+)-translocatingNADH-ubiquinone reductase subunit A	
nqrE	Na(+)-translocating NADH-quinonereductase subunit E (Na-nqr complex subunit 5)	Probable membrane protein
nqrD	Na(+)-translocatingNADH-ubiquinone reductase subunit D	Probable membrane protein
nqrC	Na(+)-translocatingNADH-ubiquinone reductase subunit C	Probable membrane protein
1.2.7 Fermentation		
EC 4.1.1.49	Phosphoenolpyruvate carboxykinase (ATP) (Phosphopyruvate carboxylase (ATP)) -- Found in "Gluconeogenbesis"	
EC 1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+) ('Malic' enzyme)	
EC 1.1.99.6	D-2-Hydroxy-acid dehydrogenase.	

EC 4.2.99.11	Methylglyoxal synthase
EC 3.1.2.6	Hydroxyacylglutathione hydrolase; (GLYOXALASE II)
EC 1.1.1.1	Alcohol dehydrogenase
EC 1.2.1.10	Acetaldehyde-CoA dehydrogenase(acd, adhE, ana)
EC 1.97.1.4	Pyruvate formate-lyase 1-activating enzyme (act)
EC 1.1.1.1	Alcohol dehydrogenase class III
EC 1.18.99.1	Probable small subunit of hydrogenase-3, Fe-S protein. Part of FHL complex (hycB) -- Found in "Aerobic"
hycC	Membrane-spanning protein of hydrogenase-3 (FHL complex)(hevC)
hycI	Hydrogenase 3 maturation protease
EC 1.18.99.1	Membrane-spanning protein of hydrogenase-3 (FHL complex)(formate hydrogenlyase subunit 4)
hyfE	Hydrogenase 4 membrane subunit
hyfH	HYDROGENASE-4 COMPONENT I
EC 1.18.99.1	Large subunit of hydrogenase-3 (FHL complex) (hycE)
EC 1.18.99.1	Probable [Fe-S] protein of hydrogenase-3 (FHL complex) (hycF)
hycG	Hydrogenase activity (hevG)
EC 1.18.99.1	Processing of HycE subunit of hydrogenase-3 (hycH)
hypB	Guanine neucleotide binding protein, Ni+ donor for HycE in hydrogenase-3
EC 2.3.1.54	Pyruvate formate-lyase (pfl)
EC 1.19.6.-	NADPH-flavin oxidoreductase (flavinreductase P)
1.2.8 ATP-proton motive force interconversion	
EC 3.6.1.34	Membrane-bound ATP synthase, F1 sector, alpha-subunit (atpA, uncA)
EC 3.6.1.34	Membrane-bound ATP synthase, Fo sector, subunit a (atpB, uncB)
EC 3.6.1.34	Membrane-bound ATP synthase, F1 sector, Epsilon-subunit (atpC, uncC)
EC 3.6.1.34	Membrane-bound ATP synthase, F1 sector, beta-subunit (atpD. uncD)
EC 3.6.1.34	Membrane-bound ATP synthase, F1 sector, subunit b (atpF, uncF)
EC 3.6.1.34	Membrane-bound ATP synthase, F1 sector, gamma-subunit (atpG, uncG)
EC 3.6.1.34	Membrane-bound ATP synthase, F1 sector, delta-subunit (atpH, uncH)
1.3 Central Intermediary metabolism	
1.3.1 General	
EC 3.1.3.2	Acid phosphatase, phosphoanhydride phosphohydrolase, 6-phytase (appA)

EC 4.3.1.1	Aspartate ammonia-lyase (aspartase)
EC 5.4.2.8	Phosphomannomutase (mrsA) (rfbK)
EC 4.2.1.1	Carbonic anhydrase
EC 5.1.3.3	Galactose-1-epimerase (mutarotase) (aldose 1-epimerase)
gcvA	Positive regulator for gcv; glycine cleavage system transcriptional activator
EC 3.1.4.46	Glycerophosphodiester phosphodiesterase, glycerophosphoryl diester phosphodiesterase, periplasmic
lrp	Regulator for leucine (or lrp) regulon and high-affinity branched-chain amino acid transport system
EC 1.7.99.5	5,10-Methylenetetrahydrofolate reductase
EC 2.5.1.6	Methionine adenosyltransferase 2 (AdoMet synthase) (metK)
EC 1.6.1.1	Pyridine nucleotide transhydrogenase, alpha subunit (nicotinamide nucleotide transhydrogenase subunit alpha) (pntA). Possible membrane protein
EC 1.6.1.1	Pyridine nucleotide transhydrogenase, beta subunit (pntB). Possible membrane protein
pqq	Rdox cofactor, for apo glucose dehydrogenase; cryptic in K12
lacD	Tagatose-1,6-diphosphate aldolase
yeeZ	enzyme of sugar metabolism
Sugar kinase	
patB	Aminotransferase
1.3.2 Gluconeogenesis	
EC 3.1.3.11	Fructose biphosphatase (D-fructose-1,6-bisphosphate 1- phosphohydrolase) (FBPase) (fbp, fdp)
EC 4.1.1.49	Phosphoenolpyruvate carboxykinase (pckA)
1.3.3 Sugar-nucleotide biosynthesis, conversions	
rfbJ	CDP-abequose synthase
EC 2.7.7.23	N-acetylglucoseamine-1-phosphate uridyltransferase (UDP-N-acetylglucosamine pyrophosphorylase) (glmU)
EC 2.7.7.24	TDP-glucose pyrophosphorylase (DTDP-glucose synthase) (glucose-1-phosphate thymidylyltransferase) (rfbA,rmlA)
EC 4.2.1.46	DTDP-glucose-4,6 dehydratase (rfbB, rmlB)
EC 5.1.3.13	DTDP-4-dehydrorhamnose-3,5-epimerase (rfbC, rmlC)
EC 1.1.1.133	DTDP-4-dehydrorhamnose reductase (dtdp-4-keto-L-rhamnose reductase) (dtdp-6-deoxy-L-mannose dehydrogenase) (dtdp-L-rhamnose synthetase)

	(rfbD, rmlD)	
EC 3.1.3.5/EC 3.6.1.45	5'-nucleotidase; 5'-nucleotidase NucA precursor(ushA)	
1.3.4 Amino sugars		
EC 2.6.1.16	L-Glutamine:D-fructose-6-phosphate aminotransferase(glmS)	
EC 3.5.1.25	N-Acetylglucoseamine-6-phosphate deacetylase (nagA)	
EC 5.3.1.10	Glucoseamine-6-phosphate deaminase (glucosamine-6-phosphate isomerase) (nagB, glmD)	
nagC	Repressor for the nag operon	
EC 3.2.1.96	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase	Probable secretory protein
EC 3.1.3.-	Alpha-ribazole-5'-phosphate phosphatase (cobC, phpB)	
1.3.5 Sulfur metabolism		
cysQ	Affects pool of 3'-phosphoadenosine-5'-phosphosulfate in sulfite synthesis pathway (amtA)	
yccK	Putative sulfite reductase	
EC 2.3.1.31	Homoserine O-acetyltransferase	
1.4 Amino acid Biosynthesis		
1.4.1 Glutamate family		
EC 1.4.1.4	NADP-specific glutamate dehydrogenase (NADP-gdh)	
EC 6.3.1.2	Glutamine synthetase (glutamate--ammonia ligase)	
glnB	Regulatory proteinP-II for glutamine synthetase	
EC 2.7.7.59	Uridylyltransferase acts on glnA regulator (uridylyl removing enzyme) (UTASE) (glnD)	
EC 2.1.3.3	Ornithine carbamoyltransferase 2 (otcase-2)	
EC 6.3.4.5	Argininosuccinate synthetase (citrulline--aspartate ligase)	
EC 4.3.2.1	Argininosuccinate lyase (argH)	
argR	Repressor for arg regulon; cer-mediated site specific recombination	
EC 1.2.1.41	g-Glutamylphosphate reductase (GPR) (glutamate-5-semialdehyde dehydrogenase) (glutamyl-gamma-semialdehyde dehydrogenase)	
EC 2.7.2.11	g-Glutamylphosphate kinase glutamate 5-kinase (GK)	
EC 2.6.1.2	Alanine transaminase (Glutamic-pyruvic transaminase)	
EC 1.5.1.2	Pyrroline-5-carboxylate reductase (P5CR) (p5c reductase)	
1.4.2 Aspartate family		
EC 2.6.1.1	Aspartate aminotransferase (transaminase A) (ASPAT) (aspC)	

EC 6.3.1.1	Asparagine synthetase A (aspartate:ammonia ligase) (asnA)
asnC	Regulator for asnA, asnC, and gidA
EC 1.2.1.11	Asparagine synthetase B (glutamine-hydrolyzing) (asnB)
EC 4.2.1.52	Dihydrodipicolinate synthase (DHDPS) (dapA)
EC 1.3.1.26	Dihydrodipicolinate reductase (dapB)
EC 2.3.1.117	Tetrahydrapicolinate N-succinyltransferase (thp succinyltransferase)(dapD)
EC 3.5.1.18	Succinyl-diaminopimelate desuccinylase
EC 5.1.1.7	Diaminopimelate epimerase(dapF)
EC 4.1.1.20	Diaminopimelate decarboxylase (dap decarboxylase)(lysA)
lysR	Positive regulator for lys
EC 1.1.1.3	Homoserine dehydrogenase/Aspartokinase I-homoserine dehydrogenase
EC 2.7.2.4	Aspartokinase I-homoserine dehydrogenase/Homoserine dehydrogenase
EC 2.7.1.39	Homoserine kinase (thrB)
EC 4.2.99.2	Threonine synthase (thrC)
EC 2.1.1.14	Tetrahydropteroyltriglutamate methyltransferase (5-methyltetrahydropteroyltriglutamate:homocysteine methyltransferase (metE)
metJ	Repressor for all met genes but metF
metR	Regulator for metE and metH
I.4.3 Serine family	
EC 2.1.2.1	Serine hydroxymethyltransferase (serine methylase) (SHMT)(glyA)
EC 1.1.1.95	D-3-phosphoglycerate dehydrogenase (PGDH)(serA)
EC 3.1.3.3	3-Phosphoserine phosphatase (PSP) (O-phosphoserine phosphohydrolase)(serB)
EC 2.6.1.52	3-Phosphoserine aminotransferase(serC)
cysB	Positive regulator for cysteine regulon
EC 2.3.1.30	Serine acetyltransferase(cysE)
EC 4.2.99.8	Cysteine synthase A, O-acetylserine sulfhydrolase A (O-acetylserine (THIOL)-lyase) (CSASE)(cysK)
I.4.5 Cysteine metabolism	
EC 1.1.1.27	L-Lactate dehydrogenase (Lactic acid dehydrogenase) (lctD)
I.4.6 Valine, leucine and isoleucine metabolism	
EC 1.1.1.31	3-Hydroxyisobutyrate dehydrogenase
I.4.7 Lysine metabolism	

EC 1.2.4.2	Oxoglutarate dehydrogenase (lipoamide) (sucA) -- Found in "TCA Cycle"
1.4.8 Arginine and proline metabolism	
EC 4.1.1.3	Oxaloacetate decarboxylase (Oxaloacetate beta-decarboxylase)beta chain -- Found in "Carbon Compounds"
EC 4.1.1.3	Oxaloacetate decarboxylase (Oxaloacetate beta-decarboxylase) alpha chain -- Found in Carbon Compounds"
1.4.9 Aromatic Amino acid family	
EC 2.5.1.19	5-Enolpyruvylshikimate-3-phosphate synthetase (3-phosphoshikimate 1-carboxyvinyltransferase) (EPSP synthase)(aroA)
EC 4.6.1.3	Dehydroquinate synthase(aroB)
EC 4.6.1.4	Chorismate synthase (5-enolpyruvylshikimate-3-phosphate phospholyase)(aroC)
EC 4.2.1.10	5-Dehydroquinate dehydratase (3-dehydroquinase) (aroD)
EC 1.1.1.25	Dehydroshikimate reductase, shikimate 5-dehydrogenase(aroE)
EC 4.1.2.15	Phospho-2-dehydro-3-deoxyheptonate aldolase (DAHP synthetase, phenylalanine repressible)(aroG)
EC 2.7.1.71	Shikimate kinase I (aroK)
EC 4.2.1.51	Chorismate mutase
EC 5.4.99.5	Prephenate dehydratase
EC 1.3.1.12/EC 5.4.99.5	Chorismate mutase T and prephenate dehydrogenase
EC 4.2.1.20	Tryptophan synthase, A protein(trpA)
EC 4.2.1.20	Tryptophan synthase, B protein(trpB)
EC 4.1.1.48/EC 5.3.1.24	N-(5-Phosphoribosyl)anthranilate isomerase and indole-3-glycerolphosphate synthetase(trpC)
EC 4.1.3.27	Glutamine amidotransferase and phosphoribosylanthranilate transferase(trpD) anthranilate synthase component II (trpG)
EC 4.1.3.27	Anthranilate synthase component I(trpE)
trpR	Regulator for trp operon and aroH; trp aporepressor
EC 2.1.1.-	site-specific DNA-methyltransferase ; modification methylase, type III R/M system
EC 1.2.1.-	glutamyl-tRNA reductase (hemA)
EC 2.4.2.18	Anthranilate phosphoribosyltransferase (Phosphoribosyl-anthranilate pyrophosphorylase)

1.4.17 Glutathione metabolism

EC 3.4.11.4	Tripeptide aminopeptidase (Aminotripeptidase) (Lymphopeptidase) (Imidoendopeptidase) (Peptidase B)
EC 2.5.1.18	Glutathione S-transferase (Glutathione S-aryltransferase)
EC 6.3.1.8	Glutathionylspermidine synthase (GSP synthetase)

1.4.18 Histidine

EC 2.6.1.9	Histidinol-phosphate transaminase (hisC)
EC 2.4.2.-	Amidotransferase (hisH)
EC 3.5.4.19	Phosphoribosyl-AMP cyclohydrolase
EC 3.6.1.31	phosphoribosyl-ATP pyrophosphohydrolase (hisIE, hisI)

1.4.19 Pyruvate Family

EC 5.1.1.1	Alanine racemase, biosynthetic (alr) -- Found in "Amino acids and amines"

1.4.20 Branched-chain family

EC 4.1.3.18	Acetolactate synthase, acetohydroxy acid synthase II
EC 1.1.1.86	ketol-acid reductoisomerase (acetohydroxy-acid isomeroreductase)(ilvC)
EC 2.6.1.42	Branched-chain amino acid aminotransferase (transaminase B) (BCAT) (ilvE)

1.5 Purines, Pyrimidines, nucleosides, and nucleotides

1.5.1 Purine ribonucleotide biosynthesis

EC 2.7.4.3	Adenylate kinase; pleiotropic effects on glycerol-3-phosphate acyltransferase activity(adk)
EC 2.7.4.8	Guanylate kinase (gmk)
EC 6.3.5.2	GMP synthase (glutamine-hydrolyzing) (glutamine amidotransferase)(guaA)
EC 1.1.1.205	Inosine-5'-monophosphate dehydrogenase (IMP dehydrogenase) (IMPDH) (IMPD) (guaB)
EC 2.7.4.6	Nucleoside diphosphate kinase (NDK) (ndp kinase)(ndk)
EC 2.7.6.1	Ribose-phosphate pyrophosphokinase (phosphoribosyl pyrophosphate synthetase)(prs, prsA)
EC 6.3.4.4	Adenylosuccinate synthetase (IMP--aspartate ligase)(purA)
EC 4.3.2.2	Adenylosuccinate lyase (adenylosuccinase) (ASL)(purB)
EC 4.1.1.21	Phosphoribosylaminoimidazole carboxylase catalytic subunit (purE)
EC 2.4.2.14	Amidophosphoribosyltransferase (glutamine phosphoribosylpyrophosphate amidotransferase) (ATASE) (GPATase)(purF)

EC 6.3.3.1	Phosphoribosylformylglycinamidine cyclo-ligase (AIRS) (phosphoribosyl-aminoimidazole synthetase) (air synthase)(purM)
EC 2.1.2.2	Phosphoribosylglycinamide formyltransferase (GART) (GAR transformylase) (5'-phosphoribosylglycinamide transformylase)(purN)
purR	Repressor for pur regulon, glyA, glyB, prsA, and speA
EC 3.5.1.10	Formyltetrahydrofolate deformylase (formyl-fh(4) hydrolase)(purU)
EC 2.7.4.14	Cytidylate kinase (Deoxycytidylate kinase)
EC 3.5.4.1	Cytosine deaminase
1.5.2 Pyrimidine riboneucleotide biosynthesis	
EC 1.3.3.1	Dihydroorotate oxidase (pyrD)
EC 2.4.2.10	Orotate phosphoribosyltransferase (OPRT) (OPRTase) (pyrE)
EC 4.1.1.23	Orotidine 5'-phosphate decarboxylase (OMP decarboxylase) (pyrF)
EC 1.1.1.22	UDPglucose 6-dehydrogenase -- Found in "Carbon Compunds"
EC 3.2.1.-	membrane-bound lytic murein transglycosylase D precursor (murein hydrolase D) (regulatory protein dnir) (mltD, dniR)
1.5.3 2'Deoxyriboneucleotide metabolism	
EC 3.5.4.13	2'-deoxycytidine 5'-triphosphate deaminase (deoxycytidine triphosphate deaminase) (dCTP deaminase)(dcd)
EC 3.6.1.23	Deoxyuridine 5'-triphosphate nucleotidohydrolase (DUTPase) (dutp pyrophosphatase)(dut)
EC 3.6.1.-	(Deoxy)neucleoside triphosphatase (7,8-dihydro-8-oxoguanine-triphosphatase) (8- oxo-dgtpase) (dgtp pyrophosphohydrolase). Causes AT-GC transversion, pyrophosphohydrolase). Causes AT-GC transversion (mutT)
EC 1.17.4.1	Ribonucleoside diphosphate reductase, subunit B1 (alpha chain)(nrdA)
EC 1.17.4.1	Ribonucleoside diphosphate reductase, subunit B2 (beta chain) (nrdB)
EC 1.97.1.-	Anaerobic ribonucleoside-triphosphate reductase activatingprotein (nrdG)
EC 1.17.4.2	Anaerobic ribonucleoside-triphosphate reductase(nrdD)
EC 2.1.1.45	Thymidylate synthase (TS) (thyA)
EC 2.7.4.9	Thymidylate kinase (dTMP kinase) (tmk)
EC 1.6.4.5	Thioredoxin reductase (trxB)
EC 1.11.1.-	thioredoxin peroxidase (tpx)

I.5.4 Salvage of nucleosides and neucleotides	
EC 3.6.1.41	Bis(5'-nucleosyl)-tetraphosphatase (symmetrical) (diadenosine tetraphosphatase)(apaH)
EC 2.4.2.7	Adenine phosphoribosyltransferase(apt)
EC 3.5.4.5	Cytidine deaminase (cytidine aminohydrolase) (CDA)(cdd)
EC 3.1.4.17	3',5'-cyclic-nucleotide phosphodiesterase (cpdA)
EC 3.1.4.16	2',3'-cyclic-nucleotide 2'-phosphodiesterase precursor(cpdB)
EC 4.1.2.4	Deoxyribose-phosphate aldolase (phosphodeoxyriboaldolase) (deoC)
EC 2.4.2.1	Purine nucleoside phosphorylase (inosine phosphorylase) (PNP)(deoD)
deoR	Regulator for deo operon, tsk, nupG
EC 2.4.2.8	Hypoxanthine phosphoribosyltransferase (HPRT)(hpt)
EC 2.7.1.21	Thymidine kinase(tdk)
EC 2.4.2.3	Uridine phosphorylase (UDRPase) (udp)
EC 2.4.2.9	Uracil phosphoribosyltransferase (UMP pyrophosphorylase) (UPRTase)(upp)
I.5.5 Miscellaneous	
EC 3.1.5.1	Deoxyguanosinetriphosphate triphosphohydrolase (DGTPase)(dgt)
lepA	GTP-binding membrane protein
mrp	Putative ATPase
EC 6.3.4.2	CTP synthase (UTP--ammonia ligase) (CTP synthetase) (pyrG)
EC 2.7.4.-	UMP kinase (pyrH)
EC 2.7.1.48	Uridine/cytidine kinase (pyrimidine ribonucleoside kinase)(udk)
oapA	Opacity associated protein A
oapB	Opacity associated protein B
ostA/imp	Organic solvent tolerance protein
pfs	MTA/SAH nucleosidase; includes:5'-methylthioadenosine nucleosidase ;s-adenosylhomocysteine nucleosidase; pfs protein
EC 3.1.7.2	guanosine 3',5'-bis(diphosphate) 3'-pyrophosphatase
I.6 Biosynthesis of cofactors, prosthetic groups, and carriers	
I.6.1 Biotin	
EC 2.6.1.62	Adenosylmethionine--8-amino-7-oxononanoate transaminase (DAPA AMINOTRANSFERASE)

EC 2.8.1.6	Biotin synthetase; Biotin synthesis protein, prior to pimeloyl-CoA reaction (bioB, bioH)
bioC	Biotin synthesis protein, prio to pimeloyl-CoA reaction
EC 6.3.3.3	Dethiobiotin synthetase (dethiobiotin synthase) (dtb synthetase) (DTBS)(bioD)
bioD-2	Dethiobiotin synthase2
EC 2.3.1.47	8-amino-7-oxononanoate synthase, (7-keto-8-amino-pelargonic acid synthetase) (7-kapsynthetase) (l-alanine--pimelyl coa ligase)
EC 6.3.4.15	Bifunctional protein (biotin operon repressor) (biotin-- [acetyl-CoA-carboxylase] synthetase) (biotin--protein ligase)(birA)
EC 1.-.-.-	Biotin sulfoxide reductase 1 (bds reductase 1) (bso reductase 1) (bisC)
1.6.2 Folic acid	
EC 1.5.1.3	Dihydrofolate reductase, trimethoprim resistance(folA, tmrA)
EC 6.3.2.12	Folylpolyglutamate synthase (folylpoly-gamma-glutamate synthetase) (FPGS)
EC 6.3.2.17	dihydrofolate synthase (folC, dedC)
EC 1.5.1.5	Methylenetetrahydrofolate dehydrogenase
EC 3.5.4.9	methenyltetrahydrofolate cyclohydrolase (folD, ads)
EC 3.5.4.16	GTP cyclohydrolase I(folE)
EC 2.7.6.3	2-Amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase (7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase) (HPPK) (6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase) (PPPK) (folK)
EC 2.5.1.15	Dihydropteroate synthase (DHPS) (dihydropteroate pyrophosphorylase)(folP)
EC 4.1.3.-	para-aminobenzoate synthase component I (adc synthase) (pabB)
EC 4.6.1.10	6-Pyruvoyl tetrahydrobiopterin synthase
EC 4.1.2.25	Dihydroneopterin aldolase
1.6.3 Lipoate	
lipA	Protein A of lipoic acid synthetase
lipB	Lipoate-protein ligase B (lipoate biosynthesis protein B)
1.6.4 Molybdopterin	
moaA	Molybdenum cofactor biosynthesis protein A (chlA1, chlA, narA, bisA)
moaC	Molybdenum cofactor biosynthesis protein C (chlA3)
moaD	Molybdopterin (MPT) converting factor, subunit 1
moaE	Molybdopterin (MPT) converting factor, subunit 2 (chlA5)
mobB	Molybdopterin-guanine dinucleotide biosynthesis protein B
moeA	Molybdopterin biosynthesis moeA protein (chlE, bisB, narE)

moeB	Molybdopterin biosynthesis moeB protein
mog	Required for efficient incorporation of molybdate in molybdoprotein (chlG)
1.6.5 Pantothenate	
EC 2.7.1.33	Pantothenate kinase (coaA, rts, panK)
EC 3.1.4.14	[Acyl-carrier-protein] phosphodiesterase
EC 2.7.1.24	Dephospho-CoA kinase
EC 2.7.7.3	Pantetheine-phosphate adenylyltransferase
1.6.6 Pyridoxine	
EC 1.1.1.-	1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXP reductoisomerase) (dxr)
EC 1.4.3.5	Pyridoxamine 5'-phosphate oxidase (pnp/PMP oxidase) (pdxH)
nadR	Transcriptional regulator (nadI)
1.6.8 Thiamine	
tbpA	Thiamin ABC transporter, periplasmic-binding protein
thiQ	Thiamin ABC transporter,ATP-binding protein
thiP	Thiamin ABC transporter, permease protein
thiI	Thiamine biosynthesis protein
EC 2.7.4.16	Thiamin-monophosphate kinase (thiamin-phosphate kinase) (thiL)
1.6.9 Riboflavin	
EC 3.5.4.25	GTP cyclohydrolase II (ribA)
EC 3.5.4.25	3,4-dihydroxy-2-butanone 4-phosphate synthase (dhbp synthase) (ribB)
EC 2.5.1.9	Riboflavin synthase alpha chain (ribC)
EC 2.5.1.9	Riboflavin synthase beta chain (ribH)
EC 3.5.4.-	Riboflavin-specific deaminase (ribD)
EC 1.1.1.193	5-Amino-6-(5-phosphoribosylamino)uracil reductase V
EC 3.5.4.26	Diaminohydroxyphosphoribosylaminopyrimidine deaminase A
ribF	Riboflavin kinase / FMN adenylyltransferase
1.6.10 Thioredoxin, glutaredoxin, and glutathione	
EC 2.3.2.2	gamma-glutamyltranspeptidase precursor (ggt)
EC 1.6.4.2	Glutathione reductase (NADPH), (GR), (GRASE)
EC 6.3.2.2	Glutamate--cysteine ligase (gamma-glutamylcysteine synthetase) (gamma-ecs) (GCS) (gshA)
trxA	Thioredoxin 1 (TRX1) (TRX) (tsnC, fipA, trxM)

1.6.11 Vitamin B6 metabolism		
EC 2.7.1.35	Pyridoxal kinase	
1.6.12 Nicotinate and nicotinamide metabolism		
EC 2.4.2.12	Nicotinamide phosphoribosyltransferase	
EC 2.7.7.1	Nicotinamide-nucleotide adenylyltransferase. Identified as a regulator nadR as well	
ttk	Transcriptional regulator	
1.6.13 Menaquinone, and ubiquinone		
EC 2.5.1.10	Geranyltranstransferase (farnesyl-diphosphate synthase) (fpp synthase) (ispA)	
EC 2.5.1.-	1,4-dihydroxy-2-naphthoate octaprenyltransferase (menA)	Probable membrane protein
EC 4.1.3.36	Naphthoate synthase (dihydroxynaphthoic acid synthetase) (dhna synthetase) (menB)	
EC 6.2.1.26	O-succinylbenzoate-CoA synthase (osb synthase) (4-(2'- carboxyphenyl)-4-oxybutyric acid synthase) (menC)	
EC 4.1.1.71	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase (shchc synthase) / 2-oxoglutarate decarboxylase (alpha-ketoglutarate decarboxylase) (KDC) (menD)	
EC 6.2.1.26	O-succinylbenzoic acid--CoA ligase (osb-CoA synthetase) (O-succinylbenzoate-CoA synthase) (menE)	
ubiC	chorismate--pyruvate lyase (4-hydroxybenzoate synthetase)	
EC 1.14.13.-	UbiH protein , 2-octaprenyl-6-methoxyphenol hydroxylase; 2-octaprenyl-6 -methoxyphenol4-monoxygenase(ubiH, visB)	
1.6.14 Heme and porphyrin		
ccmA	Heme exporter ATP-binding protein A (ABC transporter)	
ccmB	Heme exporter protein B	
ccmC	Heme exporter protein C	
ccmE	Cytochrome C-type biogenesis protein	Possible membrane protein
ccmF	Cytochrome C-type biogenesis protein	Possible membrane protein
ccmH	Cytochrome C-type biogenesis protein	Possible membrane protein
EC 2.1.1.107	siroheme synthase (hemX) (contains: uroporphyrin-III C-methyltransferase (urogen III methylase) (SUMT) (uroporphyrinogen III methylase) (UROM) / precorrin-2 oxidase / ferrochelatase) (cysG)	
EC 4.2.1.24	delta-aminolevulinic acid dehydratase (porphobilinogen synthase) (ALAD)	

	(ALADH) (hemB)	
EC 4.3.1.8	porphobilinogen deaminase (PBG) (hydroxymethylbilane synthase) (HMBS) (pre-uroporphyrinogen synthase) (hemC)	
EC 4.2.1.75	uroporphyrinogen-III synthase (UROS) (uroporphyrinogen-III cosynthetase) (hydroxymethylbilane hydrolyase (cyclizing)) (hemD)	
EC 4.1.1.37	uroporphyrinogen decarboxylase (UPD) (hemE)	Possible membrane protein
EC 1.3.3.3	coproporphyrinogen III oxidase, aerobic (coproporphyrinogenase) (coprogen oxidase) (hemF)	
EC 1.3.3.4	protoporphyrinogen oxidase (PPO) (hemK, hemG, hemY)	Probable membrane associated protein
EC 5.4.3.8	glutamate-1-semialdehyde 2,1-aminomutase (GSA) (glutamate-1-semialdehyde aminotransferase) (gsa-at) (hemL)	
lolB	outer membrane lipoprotein LolB precursor (hemM)	
EC 4.99.1.1	ferrochelatase (protoheme ferro-lyase) (hemE synthetase) (visA)	
EC 4.4.1.17	Holocytochrome-c synthase (Cytochrome c heme-lyase); cytochrome C-type biogenesis protein	Probable membrane protein
EC 2.7.1.-	Tagatose 6-phosphatekinase (gatZ)	
EC 1.6.99.2	Hypothetical NAD(P)H oxidoreductase (ycaK)	
hmsH	Involved in hemin binding and autoagglutination	
hmsF	Involved in hemin binding and autoagglutination. (Polysaccharide deacetylase domain)	Probable membrane protein
hmsR	Involved the regulation of the hms locus hemin binding and autoagglutination	Probable membrane protein
1.6.15 Enterochelin		
EC 5.4.99.6	sochorismate synthase (entC)	
1.6.17 Iron-Sulfur assembly		
iscU	Part of iscSUA-hscBA-fdx cluster	
nifS	Iron-sulfur cofactor synthesis protein	
tipB	Cytochrome C biogenesis protein	
1.7 Fatty acid biosynthesis		
EC 6.2.1.20	2-acylglycerophosphoethanolamine acyltransferase / acyl-acyl carrier protein synthetase (2-acyl-gpe acyltransferase / acyl-ACP synthetase)(aas)	Possible membrane protein
EC 6.4.1.2	acetyl-CoA carboxylase carboxyl transferase subunit alpha(accA)	
EC 6.4.1.2	acetyl-CoA carboxylase biotin carboxyl carrier protein (accB, fabE)	
EC 6.3.4.14	biotin carboxylase (A subunit of acetyl-CoA carboxylase) (accC)	
EC 6.4.1.2	acetyl-CoA carboxylase (EC 6.4.1.2) carboxyltransferase beta chain (accD)	

EC 2.7.7.41	phosphatidate cytidylyltransferase (CDP-diglyceride synthetase) (CDP-diglyceride pyrophosphorylase) (CDP-diacylglycerol synthase) (cdsA)	Probable membrane protein
EC 2.7.1.107	diacylglycerol kinase (DAGK) (diglyceride kinase) (DGK)(dgkA)	
EC 4.2.1.60	3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase (beta- hydroxydecanoyl thioester dehydrase) (fabA)	
EC 2.3.1.41	3-oxoacyl-[acyl-carrier-protein] synthase (fabB)	
EC 2.3.1.39	malonyl CoA-acyl carrier protein transacylase (MCT) (fabD)	
EC 2.3.1.41	3-oxoacyl-[acyl-carrier-protein] synthase II (fabF)	
EC 1.1.1.100	3-oxoacyl-[acyl-carrier protein] reductase (3-ketoacyl-acyl carrier protein reductase) (fabG)	
EC 2.3.1.41	3-oxoacyl-[acyl-carrier-protein] synthase III (fabH)	
EC 1.3.1.9	enoyl-(acyl-carrier-protein) reductase (NADH) (fabI)	
lktC	Acyl carrier protein for processing of prohemolysin (Toxin activation protein C)	
EC 3.1.2.-	acyl-CoA thioesterase II (tesB)	
1.8 Broad regulatory functions		
arcA	aerobic respiration control protein arcA homolog	
arcB	aerobic respiration control sensor protein arcB	Probable membrane protein
cpxA	sensor protein acting on arcA, F-pillin formation, and energy coupling (His kinase)(ecfB, ssd, eup)	Probable membrane protein
cpxR	Transcription regulator	
crp	cdc2-like serine/threonine-protein kinase CRP	
tldD	Suppresses inhibitory activity of CsrA	
EC 4.6.1.1	adenylate cyclase (cyaA)	
era	GTP-binding protein(1st module)	
fnr	Regulatory gene for oxidoreductases and others	
fur	Fe III uptake; negative regulator	
EC 3.6.1.11	Exopolyphosphatase (metaphosphatase) -- Found in "Phosphorus"	
lexA	Regulator of SOS regulon	
EC 3.4.21.53	ATP-dependent protease La (lon, capR, deg, muc, lopA)	
lytB	Control of stringent response; involved in penicillin tolerance	
oxyR	Activator, hydrogen peroxide-iducible genes	
EC 2.7.6.5	Regulation of RNA synthesis; ppGpp synthetase (relA)	
sigG	RNA POLYMERASE SIGMA FACTOR	

rpoD	RNA polymerase , sigma(70) factor; regulation of proteins induced at high temperatures
rseB	Sigma-E factor regulatory protein
mclA	Sigma-E factor negative regulatory protein
rpoE	RNA polymerase, sigma-E factor; heat shock and oxidative stress
hslU	Heat shock protein
rpoH	RNA polymerase, sigma(32) factor; regulation of proteins induced at high temperatures
rpoN	RNA polymerase, sigma(54 or 60) factor; nitrogen and fermentation regulation
sspA	stringent starvation protein A
sspB	stringent starvation protein B
sorC	Putative transcriptional regulator
II. Macromolecules	
II.1 Synthesis and modification	
II.1.1 Ribosomal and stable RNAs	
rRNA (adenine-N6)-methyltransferase	
rsuA	16s pseudouridylate 516 synthase
II.1.2 Ribosomal protein synthesis and modification	
EC 2.1.1.-	ribosomal protein L11 methyltransferase
rplA	50S ribosomal protein L1, regulates synthesis of L1 and L11
rplB	50S ribosomal protein L2
rplC	50S ribosomal protein L3
rplD	50S ribosomal protein L4, regulates expression of S10 operon
rplE	50S ribosomal protein L5
rplF	50S ribosomal protein L6
rplI	50S ribosomal protein L9
rplJ	50S ribosomal protein L10
rplK	50S ribosomal protein L11
rplL	50S ribosomal protein L7/L12
rplM	50S ribosomal protein L13

rplN	50S ribosomal protein L14
rplO	50S ribosomal protein L15
rplP	50S ribosomal protein L16
rplQ	50S ribosomal protein L17
rplR	50S ribosomal protein L18
rplS	50S ribosomal protein L19
rplT	50S ribosomal protein L20, and regulator
rplV	50S ribosomal protein L22
rplW	50S ribosomal protein L23
rplX	50S ribosomal protein L24
rpS1	30S ribosomal protein S1
rpS2	30S ribosomal protein S2
rpS3	30S ribosomal protein S3
rpS4	30S ribosomal protein S4
rpS5	30S ribosomal protein S5
rpS7	30S ribosomal protein S7
rpS8	30S ribosomal protein S8
rpS9	30S ribosomal protein S9
rpS10	30S ribosomal protein S10
rpS11	30S ribosomal protein S11
rps12	30S ribosomal protein S12 (strA)
rpS13	30S ribosomal protein S13
rpS14	30S ribosomal protein S14
II.1.3 Ribosome maturation and modification	
riml	Modification of 30S ribosomal subunit protein S18; acetylation of N-terminal alanine
rimK	Modification of ribosomal protein S6
rbfA	Ribosome binding factor A
II.1.4 tRNAs	
alaT	alanine tRNA 1B
argQ	arginine tRNA 2
argU	arginine tRNA 4
asnT	Asparagine tRNA

aspT	Aspartate tRNA1
glnU	Glutamine tRNA 1
glnV	Glutamine tRNA 2
glyT	Glycine tRNA 2
glyU	Glycine tRNA 1
ileT	Isoleucine tRNA 1
leuP	Leucine tRNA 1
leuQ	Leucine tRNA 1
leuT	Leucine tRNA 1
leuU	Leucine tRNA 2
leuV	Leucine tRNA 1
lysT	Lysine tRNA
lysV	Lysine tRNA
lysV	Lysine tRNA
lysW	Lysine tRNA
metT	Methionine tRNA
metU	Methionine tRNA
metV	Methionine tRNA-fMet2
metW	Methionine tRNA-fMet1
	Cysteinyl-tRNA
	Trptophane tRNA
pheR	Phenylalanine tRNA
proK	Proline tRNA 1
serT	Serine tRNA 1
serU	Serine tRNA 2
thrT	Threonine tRNA 3
tyrT	Tyrosine tRNA 1
valT	Valine tRNA 1
valU	Valine tRNA 1
valV	Valine tRNA 2A
valW	Valine tRNA 2B

II.2.6 Aminoacyl tRNA synthesis and modification

EC 6.1.1.7	Alanyl-tRNA synthetase (alaS)
EC 6.1.1.19	Arginyl-tRNA synthetase (argS)
EC 6.1.1.22	Asparaginyl-tRNA synthetase (asnS)
EC 6.1.1.12	Aspartyl-tRNA synthetase (aspS)
EC 2.7.7.25	tRNA nucleotidyltransferase (cca)
EC 6.1.1.16	Cysteinyl-tRNA synthetase (cysS)
EC 2.1.2.9	Methionyl-tRNA formyltransferase (fmt)
EC 6.1.1.18	Glutaminyl-tRNA synthetase (glnS)
EC 6.1.1.17	Glutamyl-tRNA synthetase (gltX)
EC 6.1.1.14	Glycyl-tRNA synthetase alpha chain (glyQ)
EC 6.1.1.14	Glycyl-tRNA synthetase beta chain (glyS)
EC 6.1.1.21	Histidyl-tRNA synthetase (hisS)
sfhB	PSEUDOURIDYLATE SYNTHASE)(URACIL HYDROLYASE)
EC 4.2.1.70	Pseudouridylate synthase I (pseudouridine synthase I) (uracil hydrolyase) (truA, hisT)
EC 5.4.99.12	tRNA-pseudouridine synthase I (truA)
EC 5.4.99.-	tRNA-pseudouridine synthase (truB)
EC 6.1.1.5	Isoleucyl-tRNA synthetase (ileS)
sun	Met-tRNAi formyltransferase, fmt
EC 6.1.1.4	Leucyl-tRNA synthetase (leuS)
EC 6.1.1.6	Lysyl-tRNA synthetase, constitutive, repressor of ColE1 mutation in primer RNA (lysS)
EC 6.1.1.6	Lysyl-tRNA synthetase, heat induciblelys (lysU)
EC 6.1.1.10	Methionyl-tRNA synthetase (metG)
EC 2.5.1.8	tRNA delta(2)-isopentenylpyrophosphate transferase (miaA) (trpX)
EC 6.1.1.20	Phenylalanyl-tRNA synthetase alpha chain (pheS)
EC 6.1.1.20	Phenylalanyl-tRNA synthetase beta chain (pheT)
EC 6.1.1.15	Prolyl-tRNA synthetase (proS)
EC 3.1.1.29	Peptidyl-tRNA hydrolase (pth)
queA	S-adenosylmethionine:tRNA ribosyltransferase-isomerase
EC 2.9.1.1	L-seryl-tRNA(SER) selenium transferase (cysteinyl-tRNA(SER) selenium transferase) (selenocysteine synthase) (selenocysteinyl-tRNA(SER) synthase) (selA)
EC 2.7.9.3	Selenide,water dikinase (selenophosphate synthetase) (selenium donor protein)

	(selD)	
EC 6.1.1.11	Seryl-tRNA synthetase (serS)	
EC 2.4.2.29	tRNA-guanine transglycosylase (tgt)	
EC 6.1.1.3	Threonyl-tRNA synthetase (thrS)	
EC 2.1.1.35	tRNA (uracil-5-)-methyltransferase (trmA)	
EC 2.1.1.31	tRNA (guanine-n1)-methyltransferase (m1g-methyltransferase) (trmD)	
EC 6.1.1.2	Tryptophanyl-tRNA synthetase (trpS)	
EC 6.1.1.1	Tyrosyl-tRNA synthetase (tyrS)	
EC 6.1.1.9	Valyl-tRNA synthetase (valS)	
ycfB	tRNA(5-methylaminomethyl-2-thiouridylate)-methyltransferase	
ygcA	RNA methyltransferase	
II.1.6 Protein translation and modification		
EC 3.5.1.31	Polypeptide deformylase (def)	
EC 5.3.4.1	Polypeptide deformylase (def)	
dsbD	Thiol:disulfide interchange protein; (C-type cytochrome biogenesis protein cycZ)	Probable membrane protein
dsbE	Thiol:disulfide interchange protein	Probable secretory protein
EC 5.3.4.1	Protein disulfide-isomerase (dsbC)	
frr	ribosome recycling factor (rrf)	
fusA	elongation factor EF-G	
EC 2.7.7.42	Glutamate-ammonia-ligase adenylyltransferase (glutamine-synthetase adenylyltransferase) (glnE)	
greA	transcription elongation factor	
EC 3.4.11.-	Aminotripeptidase; peptidase T (pepT)	
infB	translation initiation factor IF-2	
infC	translation initiation factor IF-3	
EC 3.4.11.18	methionine aminopeptidase (MAP) (peptidase M) (map)	
pmbA	Maturation of antibiotic MccB17	
EC 5.2.1.8	Peptidyl-prolyl cis-trans isomerase; survival protein-surA (ppiA)	
EC 5.2.1.8	Peptidyl-prolyl cis-trans isomerase (FkbP)	
EC 5.2.1.8	Peptidyl-prolyl cis-trans isomerase (fkpA)	
EC 5.2.1.8	Peptidyl-prolyl cis-trans isomerased (ppiase D) (rotamase D)	
EC 5.2.1.8	Peptidyl-prolyl cis-trans isomerase B (ppiase B) (rotamase B) (ppiB)	

prfA	peptide chain release factor RF-1	
prfB	peptide chain release factor RF-2	
prfC	peptide chain release factor RF-3	
selB	selenocysteine-specific elongation factor	
tsf	elongation factor EF-Ts	
tufA	elongation factor EF-Tu	
bipA	Tyrosine phosphorylatedprotein A; GTP-binding elongation factor family protein TypA/BipA	
Peptide synthase		
II.2 Metabolism of Complex Lipids		
II.2.1 Glycerolipid metabolism		
EC 3.1.1.3	Triacylglycerol lipase (Lipase) (Tributyrase)	
EC 6.3.2.4	D-Alanine-D-alanine ligase -- Found in "peptidoglycan"	
EC 1.1.1.8	Glycerol-3-phosphate dehydrogenase(NAD+)	
II.2.2 Inositol Phosphate Metabolism		
EC 3.1.3.25	myo-Inositol-1(or 4)-monophosphatase; participate in posttranscriptional control of gene expression (suhB)	
iolE	myo-inositol catabolism	
iolD	myo-inositol catabolism	
II.2.3 Phospholipid metabolism		
iolC	myo-inositol catabolism	
EC 2.7.8.-	Cardiolipin synthetase (cls)	Probable membrane protein
EC 2.4.99.-	Prolipoprotein diacylglyceryl transferase (lgt)	Probable membrane protein
hel	outer membrane protein P4; lipoprotein E	
EC 3.1.3.27	Phosphatidylglycerophosphatase A, membrane bound (pgpA)	
pgpB	phosphatidylglycerophosphatase B, membrane bound	
EC 2.7.8.5	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (phosphatidylglycerophosphate synthase) (pgp synthase) (pgsA)	
EC 2.3.1.15	Glycerol-3-phosphate acyltransferase (plsB)	
EC 2.3.1.51	1-acyl-sn-glycerol-3-phosphate acyltransferase (plsC). Phosphate acyltransferases; Function in phospholipid biosynthesis	Probable membrane protein

plsX	fatty acid/phospholipid synthesis protein, glycerolphosphate auxotrophy	
EC 4.1.1.65	Phosphatidylserine decarboxylase proenzyme (psd)	
lppC	LppC lipoprotein	
EC 2.7.8.8	CDP-diacylglycerol--serine O-phosphatidyltransferase (phosphatidylserine synthase) (pssA)	
lolA	outer membrane lipoproteins carrier protein	
II.3 Degradation of macromolecules		
II.3.1 RNA		
EC 3.1.13.1	Exoribonuclease II (ribonuclease II) (RNase II), mRNA degradation (rnb)	
EC 3.1.26.3	Ribonuclease III (RNase III), for ds-RNA (rnc)	
EC 3.1.26.3	Ribonuclease D (RNase D), transfer tRNA precursors	
EC 3.1.4.-	Ribonuclease E (RNase E), mRNA turnover, maturation of 5S RNA (rne)	
EC 3.1.26.4	Ribonuclease HII (RNase HII) (rnhB)	
EC 2.7.7.56	Ribonuclease ph (RNase ph) (tRNA nucleotidyltransferase)(rph)	
vacB	RIBONUCLEASE R (RNASE R)	
basR	transcriptional regulatory protein	
EC 2.7.7.-	ATP-dependent RNA helicase (deaD)	
greB	transcription elongation factor	
EC 1.8.4.6	Peptide methionine sulfoxide reductase (msrA). Upon exposure to antibiotics, msrA expression increases	Probable membrane protein
nusA	N utilization substance protein A, Transcriptional termination	
nusB	N utilization substance protein A, Transcriptional termination	
nusG	transcription antitermination protein	
EC 3.1.26.5	Ribonuclease P protein component (protein c5) (RNase P)RNase P RNA component, M1 RNA, processes tRNA, 4.5S RNA (rnpA)	
EC 2.7.7.19	Poly(A) polymerase (PAP) (plasmid copy number protein) (pcnB)	
EC 2.7.7.8	polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase) (PNPase) (pnp	
rho	transcription termination factor	
EC 2.7.7.6	DNA-directed RNA polymerase alpha chain (rpoA)	
EC 2.7.7.6	DNA-directed RNA polymerase beta chain (rpoB)	
EC 2.7.7.6	DNA-directed RNA polymerase beta' chain (rpoC)	
srmB	ATP-dependent RNA helicase (srmB)	

EC 3.1.21.5	Type III site-specific deoxyribonuclease
fis	factor-for-inversion stimulation protein
EC 3.2.2.23	Formamidopyrimidine-DNA glycosylase (fpg)
gidA	glucose inhibited division protein A, chromosome replication?
gidB	glucose inhibited division protein, chromosome replication?
EC 5.99.1.3	DNA gyrase subunit A, type II topoisomerase (gyrA)
EC 5.99.1.3	DNA gyrase subunit B (gyrB)
EC 3.6.1.-	helicase IV (75 kD helicase)
himD	integration host factor beta-subunit
EC 2.7.7.7	DNA polymerase III, delta subunit (holA)
EC 2.7.7.7	DNA polymerase III, delta' subunit (holB)
EC 2.7.7.7	DNA polymerase III, chi subunit (holC)
EC 2.7.7.7	DNA polymerase III, psi subunit (holD)
hsdS	Specificity determenant for hsdM and hsdR
EC 6.5.1.2	DNA ligase (lig)
mfd	transcription-repair coupling factor, Mutation frequency decline
mioC	Initiation of chromosome replication
rhlB	(mmrA) Postreplication repair
mutH	DNA mismatch repair protein (mutR, prv)
mutL	Methyl-directed mismatch repair
mutS	Methyl-directed mismatch repair
EC 3.2.2.-	A/G-specific adenine glycosylase (mutY)
EC 2.1.1.63	Methylated-DNA--protein-cysteine methyltransferase (6-O- methylguanine-DNA methyltransferase) (O-6-methylguanine-DNA- alkyltransferase)(ogt)
EC 5.99.1.-	topoisomerase IV subunit A (parC)
EC 5.99.1.-	topoisomerase IV subunit B (parE)
EC 2.7.7.7	DNA polymerase I (polA)
priA	primosomal protein N'
priB	primosomal replication protein N
recA	recombination protein, ATP-dependent coprotease
recX	Regulatory protein recX
recF	ssDNA and dsDNA binding protein
recG	DNA helicase, resolution of Holliday junctions

trmH	tRNA/rRNA methyltransferase	
rpn	ribonuclease BN	Probable membrane protein
tex	transcription accessory protein	
yfiF	rRNA methylase; rRNA/rRNA methyltransferase	
II.3.2 DNA		
EC 3.1.21.3	Type I restriction enzyme EcoKI R protein (hsdR)	
EC 3.1.21.3	Type I restriction enzymeEcoR124II R protein	
EC 4.2.99.18	Endonuclease III (DNA-(apurinic or apyrimidinic site) lyase); 8-oxoguanine DNA glycosylase (nth)	
EC 3.1.11.5	DNA helicase, exodeoxyribonuclease V beta chain, ssDNA endonulease, chi sequence recognition (recB)	
EC 3.1.11.5	DNA helicase, exodeoxyribonuclease V beta chain, ssDNA endonulease, chi sequence recognition (recC)	
EC 3.1.11.5	DNA helicase, exodeoxyribonuclease V beta chain, ssDNA endonulease, chi sequence recognition (recD)	
EC 3.1.-.-	Single-stranded-DNA-specific exonuclease (recJ)	
EC 3.1.11.1	Exodeoxyribonuclease I (exonuclease I) (DNA deoxyribophosphodiesterase) (DRPase) (sbcB)	
uvrB	excinuclease ABC subunit B	
uvrC	excinuclease ABC subunit C	
EC 3.1.11.6	Exodeoxyribonuclease large subunit (exonuclease VII large subunit) (xseA)	
EC 3.1.11.2	Exodeoxyribonuclease III (exonuclease III) (xthA)	
EC 2.1.1.72	DNA adenine methylase (deoxyadenosyl-methyltransferase)(dam)	
dfp	DNA/pantothenate metabolism flavoprotein	
dinG	probable ATP-dependent helicase dinG homolog	
dksA	DnaK suppressor protein	
dnaA	chromosomal replication initiator protein, can be transcription regulator	
EC 3.6.1.-	replicative DNA helicase (dnaB)	
ssb	ssDNA-binding DNA	
EC 3.6.1.-	replicative DNA helicase (recQ)	
EC 2.7.7.7	DNA polymerase III, alpha chain (dnaE, polC)	
EC 2.7.7.-	DNA primase	
EC 2.7.7.7	DNA polymerase III, beta chain (dnaN)	
EC 2.7.7.7	DNA elongation factor III, DNA polymerase III subunits gamma and tau (dnaX)	

recN	Recombination and DNA repair	
recO	Interaction with RecR and RecF	
recR	recombination protein recR	
EC 3.6.1.-	ATP-dependent DNA helicase (rep)	
rec2	Recombination protein	Probable membrane protein
ruvA	holliday junction DNA helicase	
ruvB	holliday junction DNA helicase	
ruvC	Holliday junction nuclease, resolution of structures	
seqA	Negative modulator of initiation of replication	
EC 5.99.1.2	DNA-topoisomerase type I, omega protein (topA)	
EC 5.99.1.2	DNA topoisomerase III (topB)	
EC 3.2.2.-	uracil-DNA glycosylase (UDG) (ung)	
EC 3.2.2.-	DNA-3-methyladenine glycosidase I (tagI)	
uvrA	Repair of UV damage to DNA (dinE)	
EC 3.6.1.-	DNA helicase II, DNA-dependent ATPase (uvrD)	
xerC	integrase/recombinase	
xerD	integrase/recombinase (xprB)	
dinP	DNA damage-induced mutagenesis	
IIaFI	Type III restriction-modification enzyme homologus to EcoPI	
dprA	DNA processing chain A; smf protein	
Snf2	ATP-dependent helicase (Snf2/Rad54 family)	
hrpA	ATP-dependent helicase	
II.3.3 Proteins and peptides		
EC 3.4.21.92	ATPase subunit of ATP-dependent protease (clpB)	
EC 3.4.21.-	ATP-dependent clp proteinase; proteolytic subunit (clpP)	
EC 3.4.21.-	ATP-dependent Clp protease, ATP-binding subunit (clpX) (ABC)	
EC 3.4.24.15	Oligopeptidase A; thimet oligopeptidase (prlC)	
EC 3.4.-.-	Protease-specific for phage lambda cII repressor (hflC)	
EC 3.4.21.-	Trypsin-like proteinase (degS). Possible secretory protein	
EC 3.4.99.-	Zinc protease (pqqL); (yddC)	
EC 3.4.24.57	O-sialoglycoprotein endopeptidase (gcp)	
EC 3.4.24.3	Collagenase (prtC)	
hflX	GTP-binding protein	

lktA	Leukotoxin protein	
EC 3.4.21.-	Periplasmic serine protease; heat shock protein (htrA)	
EC 3.4.11.1	Aminopeptidase A/I (pepA) (LEUCYL AMINOPEPTIDASE)(LAP)	
EC 3.4.13.3	Aminoacyl-histidine dipeptidase (xaa-His dipeptidase) (X-His dipeptidase) (beta-alanyl-histidine dipeptidase) (carnosinase) (peptidase D) (pepD)	
EC 3.4.-.-	peptidase E (alpha-aspartyl dipeptidase)	
EC 3.4.11.2	Aminopeptidase N (alpha-aminoacylpeptide hydrolase) (pepN)	
EC 3.4.11.9	aminopeptidase P II (pepP)	
EC 3.4.21.-	carboxyl-terminal protease for penicillin binding protein (prc)	
slpA	Integrase of P4-like prophage	
intX	Integrase of P4-like prophage	
sohB	Putative protease	Probable secretory protein
EC 3.4.-.-	Protease IV, a signal peptide peptidase (sppA)	Secretory protein
EC 2.7.8.6	Undecaprenyl-phosphate galactosephosphotransferase	
pmtl	Dolichyl-phosphatemannose synthase	
ykoT	Dolichol phosphate mannose synthase	
II.3.4 Polysaccharides		
EC 2.4.1.25	4-alpha-glucanotransferase (amylomaltase) (disproportionating enzyme) (D-enzyme) (malQ)	
EC 3.2.1.1	Alpha-amylase precursor (1,4-alpha-D-glucan glucanohydrolase) (malS)	
EC 2.4.1.21	Glycogen synthase (starch (bacterial glycogen) synthase) (glgA)	
EC 2.4.1.18	1,4-alpha-glucan branching enzyme (glycogen branching enzyme) (glgB)	
EC 2.7.7.27	Glucose-1-phosphate adenylyltransferase (ADP-glucose synthase) (ADP-glucose pyrophosphorylase) (glgC)	
EC 2.4.1.1	Glycogen phosphorylase (glgP)	
EC 3.2.1.-	(glgX) Probable part of glycogen operon	
II.4 Cell envelope		
II.4.1 Membranes, Lipoproteins, and Porins		
acrA	Lipoprotein mutants sensitive to drugs	
hlpA	outer membrane protein	
hlpB	Hlp protein	
nlpB	Lipoprotein-34	
nlpC	NlpC lipoprotein	

nlpD	Lipoprotein	
nlpI	NlpI lipoprotein homologue	Probable secretory protein
ompA	Outer membrane protein 3a (Omp29); Fc-binding protein	
plp4	Lipoprotein Plp4	
ompC	Outer membrane protein 1b	
ompW	Outer membrane protein W	
rlpB	minor lipoprotein	
smpB	Small protein B	
Oma87	Outer membrane antigen; protective surface antigen D15 in H. inf.	Membrane protein
apbE	Thiamine biosynthesis lipoprotein	
vacJ	Lipoprotein (vacJ)	
plpI	Lipoprotein1	
II.4.2 Surface Polysaccharides and Lipopolysaccharides		
EC 3.5.1.-	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase (enva protein)(lpxC, envA)	
EC 2.3.1.-	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase (fira protein) (rifampicin resistance protein)(lpxD, firA)	
EC 2.7.1.130	TETRAACYLDISACCHARIDE 4'-KINASE	
EC 2.3.1.-	Lipid A biosynthesis lauroyl acyltransferase;lauroylacyltransferase (heat shock protein B) (htrB)	
msbB	Lipid A biosynthesis (kdo)2-(lauroyl)-lipid IVa acyltransferase	
EC 2.7.7.38	3-deoxy-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase) (kdsB)	
EC 2.4.99.-	3-deoxy-D-manno-octulosonic-acid transferase (KDO transferase) (kdtA)	
kdkA	KDO kinase	
kdtB	lipopolysaccharide core biosynthesis	
EC 2.3.1.129	UDP-N-acetylglucosamine acyltransferase (lpxA)	
EC 2.4.1.182	Lipid-A-disaccharide synthase (lpxB)	
EC 4.1.3.3	N-acetylneuraminate lyase	
EC 5.1.3.20	ADP-L-glycero-D-mannoheptose-6-epimerase, permits growth at high temperature (rfaD)	
rfaE	ADP-heptose synthase homolog	
rfaF	ADP-heptose--lps heptosyltransferase II, lipopolysaccharide core biosynthesis	

rfaG	Glycosyl transferase	
rfaK	Lipooligosaccharide (LOS) core biosynthesis; probably hexose transferase	
EC 4.1.2.16	2-dehydro-3-deoxyphosphooctonate aldolase (ksdA)	
rfaQ	Lipopolysaccharide core biosynthesis	
EC 2.4.1.-	Synthesis of enterobacterial common antigen (ECA):UDP-GlcNAc:undecaprenylphosphate GlcNAc-1-phosphate transferase;(rfe)	
losA	LOS biosynthesis enzyme; putative LPS biosynthesis protein	
lgtF	Glycosyl transferase; lipopolysaccharide biosynthesis protein	
yibB	Involved in lipopolysaccharide biosynthesis	
lic2A	Involved in lipooligosaccharide biosynthesis	
gmhA	Phosphoheptose isomerase	
pssK	exopolysaccharidepolymerization protein	
II.4.3 Surface structures		
EC 5.3.4.1	protein disulfide-isomerase (dsbA)	
dsbB	Reoxidizes DsbA protein following formation of disulfide bond in P-ring flagella	Possible secretory protein
pilB	Protein transport protein	Probable membrane protein
pilC	Protein transport protein	Probable membrane protein
pilF		Probable membrane protein
pilG		Probable membrane protein
EC 3.4.-.-	Type 4 prepilin-like protein specific leader peptidase; PilD (hopD)	Probable membrane protein
fimA	Major type 1 subunit fimbrin	
pilA	Putative fimbrialsubunit PilA	
mlc	Making large colonies protein	
impA	Inner membrane protein, influences colonial morphology	
oprN	Outer membrane protein	
hsf	Surface protein; associated with fibrils in H. influenzae. It is specific for adesion	
ompP1	Outer membrane protein P1	
II.4.4 Peptidoglycan		
EC 3.5.1.28	N-acetylmuramoyl-L-alanine amidase (amiB)	
EC 3.4.16.4	Penicillin-binding protein 5 precursor (D-alanyl-D-alanine carboxypeptidase fraction A)	
EC 3.4.16.4	Penicillin-binding protein 7 precursor (PBP-7)	

EC 3.4.16.4	Penicillin-binding protein 4 precursor (D-alanyl-D-alanine carboxypeptidase fraction A)	
EC 6.3.2.4	D-alanine--D-alanine ligase B (D-alanylalanine synthetase)	
EC 3.4.99.-	Penicillin-insensitive murein endopeptidase precursor (D-alanyl-D- alanine-endopeptidase) (dd-endopeptidase)	
EC 3.2.1.-	Membrane-bound lytic murein transglycosylase A precursor (murein hydrolase A) (MLT38)(mltA)	
EC 3.2.1.-	MEMBRANE-BOUND LYTIC MUREINTRANSGLYCOSYLASE C PRECURSOR (MUREIN HYDROLASE C) (mltC)	Probable membrane protein
EC 2.7.8.13	phospho-N-acetylmuramoyl-pentapeptide-transferase	Probable membrane protein
mrcA	penicillin-binding protein 1a	Probable secretory protein
mrcB	penicillin-binding protein 1B	Probable secretory protein
ponA	PENICILLIN-BINDING PROTEIN 1A (PBP-1A)	
mrdB	Rod shape-determining protein, sensitivity to radiation and drugs (rodA)	Possible membrane protein
mreB	Rod shape-determining protein (envB, rodY)	
mreC	Rod shape-determining protein	
mreD	Rod shape-determining protein	Probable membrane protein
EC 1.1.1.158	UDP-N-acetylenolpyruvoylglucosamine reductase (UDP-N-acetylmuramate dehydrogenase	
EC 6.3.2.8	UDP-N-ACETYLMURAMATE--ALANINELIGASE (UDP-N-ACETYLMURAMOYL-L-ALANINE SYNTHETASE) (murC)	
EC 6.3.2.9	UDP-N-acetylmuramoylalanine--D-glutamate ligase (UDP-N-acetylmuranoyl-L-alanyl-D-glutamate synthetase) (murD)	
(UDP-N-acetylmuramyl-tripeptide synthetase) (murE)	(UDP-N-acetylmuramyl-tripeptide synthetase) (murE)	
EC 6.3.2.15	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl- D-alanyl ligase	
EC 5.1.1.3	glutamate racemase (murI, dga, glr)	
EC 2.5.1.7	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (enoylpyruvate transferase) (UDP-N-acetylglucosamine enolpyruvyl transferase)	
EC 3.2.1.-	soluble lytic murein transglycosylase (slt)	
ywhE	Monofunctional biosyntheticpeptidoglycan transglycosylase (monofunctional tgase); gene name from B. subtilis	Probable secretory protein
gcpE	Negative regulators of the aac(2')-Ia gene (density-dependent regulation)	

III. Processes		
III.1 Transport/binding proteins		
III.1.1 Amino acids and amine		
artI	arginine 3rd transport system periplasmic binding protein	
artM	arginine 3rd transport system permease protein	
artP	ABC superfamily (atp&memb), ATP-binding component of 3rd arginine transport system(2nd module)	
artQ	ABC superfamily (membrane), membrane component of 3rd arginine ABC transport system	
brnQ	branched chain; mutants valine and o-methylthreonine resistant, glyclyvaline sensitive; transport system I for Ile, Leu, and Val	
ala tmaporter	sodium/alanine symporter	
gltS	Sodium--glutamate symport carrier protein	
EC 2.7.1.69	PTS system,n-acetylglucosamine-specific IIABC component (EIIABC-Nag)	Probable membrane protein
EC 2.7.1.69	Unknown pentitol phosphotransferaseenzyme II, A COMPONENT	Probable membrane protein
EC 2.7.1.69	PTS system, mannose-specific IID component (IID-MAN)	Probable membrane protein
potA	Spermidine/putrrscine transport protein	
potB	Spermidine/putrrscine transport protein	Probable membrane protein
potC	Spermidine/putrrscine transport protein	Probable membrane protein
potD	Spermidine/putrrscine transport protein	
sdaC	Probable ser transport	Probable membrane protein
tnaB	Low-affinity trp permease	
tyrP	Tyr-specific transport system	Probable membrane protein
tyrR	Regulation of aroF, aroG, and tyrA and aromatic amino acid transport systems	
III.1.2 Cations		
PfeA	Ferric enterobactin receptor	
corA	Mg++ transport system	
EC 2.7.1.69	PTS system, nitrogen regulatory IIA component (ptsN)	Probable membrane protein
exbB	uptake of enterochelin; tonB-dependent uptake of B colicins	
exbD	uptake of enterochelin; tonB-dependent uptake of B colicins; biopolymer transport protein	
fecB	Citrate-dependent Fe transport, periplasmic	
fecC	Citrate-dependent Fe (III) transport, cytosolic	Probable membrane protein

fecD	Citrate-dependent Fe transport, membrane-bound protein	Probable membrane protein
fecE	Citrate-dependent Fe (III) transport, membrane bound (yfeB)	
nhaB	Na+/H antiporter, pH independent	Probable membrane protein
nhaC	Na+/H antiporter, pH independent	
nikA	periplasmic binding protein for Ni	Probable membrane protein
panF	Na/pantothenate symporter	Membrane protein
putP	Major Na/pro symporter	Probable membrane protein
rsgA	Ferritin-like protein	
tonB	Energy transducer; uptake of Fe and cyanocobalamin; sensitivity to phage, colicins	
trkA	Transport of K	
trkH	K uptake	
yfeA	Mn++ transport protein; iron (chelated) ABC transporter	
EC 3.6.1.-	K+/Cu++ transporting ATPase (copA)	
snf	Sodium-dependent transporter	
znuA	zinc ABC transporter	Probable membrane protein
znuB	zinc ABC transporter	
III.1.3 Anions		
Na_Sulfate	putative Na_Sulfate (or dicarboxilate) transporter	
cysZ	Cysteine synthetas; required for sulfate transport	Probable membrane protein
modA	Molybdate uptake (ABC)	
modE	Molybdenum transport protein	
modB	Molybdate uptake	
modC	Molybdate uptake (ABC) (chlD)	
molB	Molybdenum transport protein	
pitA	low-affinity inorganic phosphate transporter 1 (pit)	Probable membrane protein
eriC	putative chloride-channel protein	
III.1.4 Carbohydrates, organic alcohols, and acids		
araH	High-affinity L-arabinose transport system; membrane protein	
EC 2.7.1.69	PTS system, Glucose phosphotransferase enzyme III glucose-specific IIA component (EIIA-Glc) (crr, gsr, iex, tgs)	Probable membrane protein
pgtA	Phosphoglycerate transport system activator	
pgtB	phosphoglycerate transport regulatory protein	

pgtC	Transport regulatory protein	
dcuB	Anaerobic dicaboxilate transport	Membrane protein
EC 2.7.1.69	pts system, fructose-specific IIBC component (EIIBC-Fru) (fructose- permease IIBC component) (phosphotransferase enzyme II, bc component) (EII-Fru)	Probable membrane protein
EC 2.7.1.69	pts system, fructose-specific IIA/Hpr component (EIIA-Fru) (fructose-permease IIA/HPr component) (phosphotransferase enzyme II, A/HPr component) (phosphotransferase fpr protein) (pseudo-HPr) (EIII-Fru) (fructose pts diphosphoryl transfer protein)	Probable membrane protein
EC 2.7.1.69	pts system, fructose-like-1 IIA component (phosphotransferase enzyme II, A component)	Probable membrane protein
EC 2.7.1.69	Galactitol-specific enzyme IIA of phosphotransferase system	Probable membrane protein
gatC	PTS system. Galactitol-specific enzyme IIC of phosphotransferase system	Probable membrane protein
glpT	sn-glycerol-3-phosphate permease	Probable membrane protein
gntT	high-affinity gluconate transporter (gluconate permease) (gnt-I system)(usgA, gntM)	Membrane protein
lamB	Phage lambda receptor and maltose uptake system (malB)	
lctP	L-lactate permease	Probable membrane protein
malE	maltose binding protein	
malF	maltose transport permease	
malG	maltose transport permease protein, active transport of maltose and maltodextrin	
malK	maltose transport ATP-binding protein	
EC 2.7.1.69	PTS system, mannose-specific IIAB COMPONENT (EIIAB-MAN)	Probable membrane protein
EC 2.7.1.69	pts system, mannose-specific IIC component (EIIC-Man) (mannose- permease IIC component) (phosphotransferase enzyme II, C component) (ptsP, pel, manY)	Probable membrane protein
mglA	Methylgalactoside and D-galactose transport ATP-binding protein	
mglB	D-galactose ABC transporter, periplasmic-binding protein	
mglC	D-galactose transport permease protein (ABC)	
EC 2.7.1.69	pts system, glucose-specific IIABC component (EIIABC-Glc) (ptsG)	Probable membrane protein
EC 2.7.3.9	phosphoenolpyruvate-protein phosphotransferase	
rbsA	D-ribose transport (ABC) ATP-binding protein	Membrane protein
rbsB	D-ribose binding protein (rbsP, prlB)	Probable secretory protein
rbsC	D-ribose transport permease protein (ABC)	membrane protein
rbsD	D-ribose transport permease protein	

uhpC	regulatory protein/sugar phosphate permease (COG2271)	membrane protein
xylF	D-xylose binding protein (xylT)	
xylG	D-xylose transport ATP-binding protein	
xylH	D-xylose transport permease protein	
EC 5.4.2.6	beta-Phosphoglucomutase	
EC 3.6.1.-	datP pyrophosphohydrolase (ntpA)	
FocA	Putative formate transporter (mutation causes hypophosphite--resistantance)	Possible membrane protein
yflS	2-oxoglutarate/malate translocator homolog	Possible membrane protein
III.1.5 Nucleosides, purines, and pyrimidines		
nupC	nucleoside permease except guanosine (cru)	Probable membrane protein
uraA	Uracil transport, ABC transporter	
III.1.6 Other		
abc	Permease component of an uncharacterized ABC transporter (COG2011)	
yfeC	ABC-type Mn2+/Zn2+ transport systems, permease components (COG1108)	
yfeD	ABC-type Mn2+/Zn2+ transport systems, permease components (COG1108)	
uup	ATPase components of ABC transporters with duplicated ATPase domains (COG0488)	
COG1133	ABC-type long-chain fatty acid transport system, fused permease and ATPase components (sbmA)	
ArtI	ABC-type amino acid transport system, periplasmic component (COG0834)	
cbiO	ABC-type cobalt transport system, ATPase component (COG1122) (cobalt trasnport system)	
uup	ATPase components of ABC transporters with duplicated ATPase domains (COG0488)	
GlnQ	ABC-type polar amino acid transport system, ATPase component (COG1126)	
COG1879	Periplasmic sugar-binding proteinsn (RbsB)	
COG1132	ABC-type multidrug/protein/lipid transport system, ATPase component (mdlB)	
FecB	ABC-type Fe3+-siderophores transport systems, periplasmic components (COG0614)	
COG1132	ABC-type multidrug/protein/lipid transport system, ATPase component (mdlB)	
MglA	ABC-type sugar (aldose) transport system, ATPase component (COG1129)	
COG1174	ABC-type proline/glycine betaine transport systems, permease component	

	(ProW)--choline transporter	
ProV	ABC-type proline/glycine betaine transport systems, ATPase components (COG1125)--choline transporter	
afuC		
afuB	Ferric transport system permease protein	Membrane protein
PhnL	ABC-type transport systems, involved in lipoprotein release, ATPase components (COG1136)	
PhnK	Various ABC transport systems, ATPase components (COG1101)	
MalK	ABC-type sugar/spermidine/putrescine/iron/thiamine transport systems, ATPase component (COG1130)	
PotC	ABC-type spermidine/putrescine transport system, permease component II (COG1177)	
abc	Uncharacterized ABC-type transport system ATPase component (COG1135)	
uup	ATPase components of ABC transporters with duplicated ATPase domains (COG0488)	
BtuC	ABC-type cobalamin/Fe3+-siderophores transport systems, permease components (COG0609)	
FepC	ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components (COG1120)	
FepC	ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components (COG1120)	
BtuC	ABC-type cobalamin/Fe3+-siderophores transport systems, permease components (COG0609)	
CcmA	ABC-type multidrug transport system, ATPase component (COG1131)	
abc	ABC-type transport systems, involved in lipoprotein release, permease components (COG0577)	
PhnL	ABC-type transport systems, involved in lipoprotein release, ATPase components (COG1136)	
ThiP	ABC-type thiamine transport system, permease components (COG1178)	
MalK	ABC-type sugar/spermidine/putrescine/iron/thiamine transport systems, ATPase component (COG1130)	
TagH	ABC-type polysaccharide/polyol phosphate transport system, ATPase component (COG1134)	
znuC	ABC-type Mn/Zn transport systems, ATPase component (COG1121)	
FepC	ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components (COG1120)	

BtuC	ABC-type cobalamin/Fe3+-siderophores transport systems, permease components (COG0609)	
BtuC	ABC-type cobalamin/Fe3+-siderophores transport systems, permease components (COG0609)	
uup	ATPase components of ABC transporters with duplicated ATPase domains (COG0488)	
abc	ABC-type (unclassified) transport system, ATPase component (COG1137)	
abc	ABC-type (unclassified) transport system, ATPase component (COG1127)	
(COG0765)	ABC-type amino acid transport system, permease component(artM)	Membrane protein
RbsB	Periplasmic sugar-binding proteins (COG1879)	
abc	ABC-type transport systems, involved in lipoprotein release, permease components (COG0577)	
btuB	Vitamin b12 receptor; TonB-dependent outer membrane receptor	
cydD	ABC superfamily (atp&memb), ATP-binding and membrane components of cytochrome-related ABC transport, Zn sensitive(2nd module)	Membrane protein
modD	Molybdate uptake	
msbA	Probable ATP-binding transport protein; multicopy suppressor of htrB	Possible membrane protein
III.2 Chaperones		
cbpA	curved DNA-binding protein; functions closely related to DnaJ	
dnaJ	Chaperon for DnaK, heat shock protein	
djlA	DNAJ-LIKE PROTEIN DJLA	
dnaK	Chaperon Hsp70; DNA biosynthesis; autoregulated heat shock protein	Possible secretory protein
hscA	Member of Hsp70 protein family	
htpG	Chaperon Hsp90; heat shock protein C62.6	
htpX	Heat shock protein	
hsp33	KDA chaperonin (heat shockprotein 33 homolog)	
hsp15	Heat shock protein 15 homolog (HSP15)	
groEL	Heat shock protein HSP60	
III.3 Cell division		
cafA	Bundles of cytoplasmic filaments; cytoplasmic axial filamentprotein	
fic	Induced in stationary phase, recognized by rpoS, affects cell division; cell filamentationprotein	
sprT	Involved in bolA gene expression at the stationary phase	
ftsA	Cell division protein, complexes with FtsZ	

ftsE	Cell division membrane protein	
ftsH	Inner membrane protein essential for cell division, putative ATPase, chaperone	
ftsI	Septum formation, Penicillin-binding protein 3; peptidoglycan synthetase. Probable secretory protein	
pbp2	Penicillin-binding protein 2; cell elongation	
ftsJ	Cell division protein	
ftsN	Essential cell division protein	Probable secretory protein
ftsQ	Cell division protein, ingrowth of wall at septum	Probable secretory protein
ftsW	Cell division, involved in shape determination	
ftsX	Cell division membrane proteinbbbbb	
ftsY	Cell division membrane protein	
ftsZ	Cell division; forms circumferential ring; GTPase	
minC	Cell division inhibitor; inhibits ftsZ ring formation	
mukB	Cell division protein involved in chromosome partitioning	
sulA	suppressor of lon; inhibits cell division and ftsZ ring formation	
tig	Trigger factor, chaperon involved in cell division	
pemK	Probable growth inhibitor protein 1, PemK-like	
zipA	cell division protein	
ftsK	cell division protein FtsK-related protein	
surE	Stationary-phase survival protein	
parB	Partitioning protein (SpoOJ)	Possible membrane protein
mesJ	Cell cycle protein	
III.4 Chemotaxis and mobility		
motC	Energization and switching of the flagellar motor (ABC transporter)	Possible membrane protein
III.5 Protein and peptide secretion		
ffh	Protein transport	
lktB	ABC protein translocator, exports leukotoxin	
lktD	Translocator for hemolysin; Leukotoxin secretion protein	
iepB	Leader peptidase (signal peptidase I)	Probable membrane protein
oppA	Oligopeptide transport, periplasmic binding protein	Membrane protein
oppB	Oligopeptide transport (ABC)	Membrane protein
oppC	Oligopeptide transport (ABC)	Membrane protein
oppD	Oligopeptide transport (ABC)	Membrane protein

oppF	Oligopeptide transportm ATP hydrolysis (ABC)	
sapA	Peptide ABC transporter, periplasmic-binding protein	
sapB	peptide ABC transporter, permease protein	Possible membrane protein
sapC	peptide ABC transporter, permease protein	Possible membrane protein
sapD	peptide transport system ATP-binding protein	
sapF	Peptide transport, ABC family of transporters	
secA	Protein secretion, ATP hydrolysis	
secB	Protein export, molecular chaperon	
secD	Protein secretion, membrane protein	Membrane protein
secE	Inner membrane protein, protein secretion (with secY)	Membrane protein
secF	Membrane protein, protein secretion function	Membrane protein
secY	Protein secretion, protein secretion (with secE)	
yajC	Protein translocase	
yidC	mediates membrane protein insertion in bacteria	Probable membrane protein
EC 3.4.23.36	Lipoprotein signal peptidase	
hns	Histone-like protein HLP (HU,BH2,HD,NS); pleiotropic effect	
lgtD	Glycosyltransferase	
tatA	Sec-independent translocase protein	
tatB	Sec-independent translocase protein	
tatC	Sec-independent translocase protein	
tatD	SEC-independent protein translocase protein	
III.6 Osmotic adaptation		
proP	Influences osmotic activation of compatible solute transporter ProP in Escherichia coli K-12	
III.7 Detoxification		
EC 1.11.1.6	Catalase hydroperoxidase HPII(III)	
EC 1.15.1.1	Cu/Zn-superoxide dismutase. Mutants are susceptible to macrophage killing	Probable membrane protein
thdF	GTP-binding protein in thiophene and furan oxidation	
bcp	hioredoxin-dependent hydroperoxide peroxidase; a bacterioferritin comigratory protein	
EC 1.11.1.5	Cytochrome C peroxidase (cytC)	
III.8 Cell killing	\	
kicA	Killing protein	
kicB	Suppressor of killing protein	

IV. Other		
IV.1 Phage-related functions and prophages		
grpE	Phage lambda replication, host DNA synthsis, heat shock protein, protein repair	
hfq	Host factor 1 for bacteriophage Qb replication, a growth-related protein	
C1gb	repressor protein CI	
muA	MU-like prophage flumu transposase	
GP49	phage-related protein	
P43	phage-related protein	
put	Repressor protein - phage (CI)	
gp16	MU-like prophage Flumu protein	
fxsA	Resistance of toxisity of gene 10 of P1 in E. coli. Not essential for growth	
IV.2 Colicin-related functions		
cvaB	Colicin V secretion ATP-binding protein (ABC)	
tolA	Membrane-spanning protein, required for outer membrane integrity	Probable membrane protein
tolB	Tolerance of colicin E2, E3, and K; leakage of periplasmic proteins	Probable membrane protein
tolQ	Inner membrane protein, membrane spanning, maintains integrity of cell envelop; tolerance to group A colicins	Probable membrane protein
tolR	Inner membrane protein, maintains integrity of cell envelop; tolerance to group A colicins	Probable membrane protein
IV.3 Plasmid-related functions		
comA	Transformation competence-related protein comA	
comB	Transformation competence-related protein comB	
comC	Transformation competence-related protein comC	Probable membrane protein
comE	Competence protein E	
comEA	comEA protein-related protein sor competence	
comF	Transformation competence protein F	
comJ	Transformation competence-related protein	
comM	Transformation competence-related protein	
tfoX	DNA transformation protein	
orfG	Transformation locus protein HI0433	
parB	nvolved in plasmid partitioning during replication	Probable membrane protein

IV.4 Drug/analog sensitivity		
acrB	Acriflavine resistance protein	Probable membrane protein
maz G	beta-lactamase regulatory protein	
ampD	Regulates ampC	
bcr	Bicyclomycin resistance protein; transmembrane protein	
cmlA	Resistance or sensitivity to chloramphenicol	Probable membrane protein
emrA	Multidrug resistance efflux pump	Membrane protein
emrB	Multidrug resistance efflux pump	Membrane protein
ksgA	S-Adenosylmethionine-6-N', N'-adenosyl (rRNA) dimethyltransferase; kasuguamycin resistance	
tehB	Tellurite resistance	
sanA	Involved in nikkomycin biosynthesis in Strptomyces ansochromogenes	Probable membrane protein
merR	Mercuric resistance operon regulatory protein	
IV.6 Radiation sensitivity		
radA	Sensitivity to gamma and UV radiation and methyl methanesulfonate	
radC	Sensitivity to radiation	
IV.6 Adaptation and atypical conditions		
dps	Global regulator, starvation conditions	
mscL	Mechanosensitive channel	Probable membrane protein
V. Virulence associated proteins and Insertion elements		
V.1 Virulence associated proteins		
mviN	Virulence factor	
cdtC	Cytolethal distending toxin C	
cdtB	Cytolethal distending toxin B	
cdtA	Cytolethal distending toxin A	
AhpA	Affects the expression of the latent Escherichia coli K12 hemolysin, ShcA, under anaerobic conditions	Probable membrane protein
kpsF	Thermoregulation of region 1 operon in the pathogenicity island of neuroinvasive Escherichia coli K1. The operon is responsible for polysialic acid export; a virulence factor on the external leaflet of the outer membrane	

vapD	virulence-associated protein D	
hhuA	Hemoglobin and hemoglobin-haptoglobin binding protein	
EC 3.4.24.13	Immunoglobulin A1 protease; IgA-specific metalloendopeptidase (IgA1 protease)	Probable secretory protein
afuA	Iron(III) ABC transporter; afuA protein	
VirB4	Conjugal transferprotein	
hitA	Iron utilization protein A	
hitB	Iron utilization protein B	Membrane protein
hitC	Iron utilization protein C	
tbp1	Transferrin-binding protein 1	
tadA	Tight adhesion	
tadB	Tight adhesion	Possible membrane protein
tadC	Tight adhesion	Possible membrane protein
tadD	Tight adhesion	
tadE	Tight adhesion	Probable secretory protein
tadF	Tight adhesion	
tadG	Tight adhesion	Possible membrane protein
rcpA	Rough colony protein A	Probable secretory protein
rcpB	Rough colony protein B	
rcpC	Rough colony protein C	Probable secretory protein
invA	Homologus to an invasion protein in H. inf. And invA protein from *Rickettsia prowazekii (lalA)*	
irgA	TonB-dependent outer membrane receptor; iron-regulated virulence protein	
V.2 Insertion sequences		
IS150	transposase	
slr2036	transposase	
IS1223	transposase	
tniA	transposase	
tnsC	ATP-dependent transposition protein	
tniB	Part of an In0. an integrons of multiresistance plasmids and transposons of gram-negative bacteria	
yadA	Adhesin/invasin (yop1), (uspA1 from Moraxella)	
In21	transposase	
VI. Unknown		

VI.1 Haemophilus influenzae

ORFS_00009.table	COG0792	Predicted endonuclease distantly related to archaeal Holliday junction resolvase
ORFS_00010.table	COG0279	Phosphoheptose isomerase
ORFS_00011.table	COG0637	Predicted phosphatase/phosphohexomutase
ORFS_00013.table	COG1238	Uncharacterized membrane protein
ORFS_00017.table	Hypothetical	
ORFS_00018.table	COG3102	Uncharacterized BCR
ORFS_00023.table	COG2974	DNA recombination-dependent growth factor C
ORFS_00028.table	N/A	
ORFS_00048.table	COG1666	Uncharacterized BCR
ORFS_00063.table	N/A	
ORFS_00066.table	COG0344	Predicted membrane protein
ORFS_00370.table	COG2932	transcriptional regulator
ORFS_00068.table	COG0834	ABC-type amino acid transport system, periplasmic component; Prokaryotic transglycosylases signature. Contains LysM/invasin domains). Probable membrane protein
ORFS_00073.table	Hypothetical	Probable membrane protein
ORFS_00086.table	COG0824	Predicted thioesterase; 4-hydroxybenzoyl-CoA thioesterase family active site.
ORFS_00101.table	COG0237	Dephospho-CoA kinase (COENZYME A BIOSYNTHESIS); Uncharacterized protein family UPF0038 signature / ATP/GTP-binding site motif A (P-loop).
ORFS_00102.table	COG0327	Uncharacterized Ancient Conserved Region
ORFS_00107.table	COG0313	Predicted methyltransferases; Uncharacterized protein family UPF0011 signature

ORFS_00109.table	N/A	
ORFS_00110.table	COG0471	N-terminal: Di- and tricarboxylate transporters (Inorganic ion transport and metabolism), C-terminal: Triphosphoribosyl-dephospho-CoA synthetase (Coenzyme metabolism).
ORFS_00118.table	COG0596	Probable membrane protein Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
ORFS_00129.table	COG2377	Uncharacterized BCR
ORFS_00137.table	COG3076	Uncharacterized BCR
ORFS_00154.table	Hypothetical	Probable membrane protein
ORFS_00158.table	Hypothetical	Probable membrane protein
ORFS_00159.table	COG2908	Uncharacterized BCR
ORFS_00161.table		
ORFS_00165.table	COG0718	Uncharacterized BCR
ORFS_00172.table	COG0561	Predicted hydrolases of the HAD superfamily (glycosyltransferases); Hypothetical cof family signature 1.
ORFS_00176.table	COG0742	Probable membrane protein N6-adenine-specific methylase (DNA replication, recombination and repair); N-6 Adenine-specific DNA methylases signature (VFLDPPF)
ORFS_00181.table	COG3201	Nicotinamide mononucleotide transporter (Coenzyme metabolism).
ORFS_00182.table	COG0500	Probable membrane protein SAM-dependent methyltransferases (Secondary metabolites biosynthesis, transport and catabolism); N-6 Adenine-specific DNA methylases signature
ORFS_00192.table	COG0500/COG2230/COG2226	SAM-dependent methyltransferases / Cyclopropane fatty acid synthase and related methyltransferases (Cell

		envelope biogenesis, outer membrane) / Methylase involved in ubiquinone/menaquinone biosynthesis--UbiE
ORFS_00195.table	COG2823	Predicted periplasmic or secreted lipoprotein
ORFS_00199.table	COG0526	Thiol-disulfide isomerase and thioredoxins
ORFS_00200.table	N/A	
ORFS_00123.table	COG2963	Transposase (DNA replication, recombination and repair)
ORFS_00128.table	N/A	
ORFS_00218.table	COG2056	Predicted permease
ORFS_00231.table	N/A	
ORFS_00232.table	COG0471 / COG1055	Di- and tricarboxylate transporters / Integral membrane protein, possible transporter. Probable membrane protein
ORFS_00237.table	COG3079	Uncharacterized BCR
ORFS_00247.table	COG3110	Uncharacterized BCR
ORFS_00250.table	COG1289	Uncharacterized membrane protein
ORFS_00251.table	COG2840	Uncharacterized BCR
ORFS_00881.table	COG0494	NTP pyrophosphohydrolases including oxidative damage repair enzymes (mutT); Aminoacyl-transfer RNA synthetases class-II signature 2. / Nudix hydrolase signature.
ORFS_00256.table	COG0277 / COG0247	N-terminal: FAD/FMN-containing dehydrogenases / Fe-S oxidoreductases; 4Fe-4S ferredoxins, iron-sulfur binding region signature. / ATP/GTP-binding site motif A (P-loop).
ORFS_00257.table	COG2050	Uncharacterized protein, possibly involved in aromatic compounds catabolism

ORFS_00263.table	COG1737	Transcriptional regulators
ORFS_00265.table	COG3010	Putative N-acetylmannosamine-6-phosphate epimerase
ORFS_00269.table	COG0537	Diadenosine tetraphosphate (Ap4A) hydrolase and other HIT family hydrolases; HIT family signature.
ORFS_00270.table	N/A	
ORFS_00276.table	N/A	
ORFS_00277.table	COG3022	Uncharacterized BCR
ORFS_00288.table	COG0477	Permeases of the major facilitator superfamily (Amino acid transport and metabolism)
ORFS_00299.table	N/A	
ORFS_00302.table	COG2062	Phosphohistidine phosphatase SixA (Signal transduction mechanisms)
ORFS_00311.table	COG0115	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase (branced-chain amino acid synthesis) (IlvE)
ORFS_00318.table	N/A	Probable secretory protein
ORFS_00319.table	COG2956	Predicted N-acetylglucosaminyl transferase
ORFS_00321.table	Translation initiation factor SUI1 signature.	
ORFS_00322.table	COG3094	Uncharacterized BCR. Probable membrane protein
ORFS_00326.table	COG0596	Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily); Lipases, serine active site / ATP/GTP-binding site motif A (P-loop).
ORFS_00338.table	COG0500	SAM-dependent methyltransferases
ORFS_00341.table	COG1160	Predicted GTPases; ATP/GTP-binding site motif A (P-loop). (two identitcal)
ORFS_00362.table	COG1573	Uracil-DNA glycosylase (DNA replication, recombination and repair)
ORFS_00362.table	N/A	

ORFS_00363.table	COG0012	Predicted GTPase
ORFS_00373.table	N/A	
ORFS_00376.table	COG0471 / COG1055	Di- and tricarboxylate transporters (Inorganic ion transport and metabolism) / Na+/H+ antiporter NhaD and related arsenite permeases; Sodium:sulfate symporter family signature
ORFS_00377.table	COG3028	Uncharacterized BCR
ORFS_00380.table	COG1929	Glycerate kinase (Carbohydrate transport and metabolism)
ORFS_00381.table	COG2610	H+/gluconate symporter and related permeases / Eukaryotic and viral aspartyl proteases active site. (VLIDSGAANTIA)
ORFS_00382.table	COG2508	Regulator of polyketide synthase expression
ORFS_00387.table	COG2731	Uncharacterized BCR
ORFS_00398.table	N/A	Uncharacterized BCR
ORFS_00399.table	COG0217	Uncharacterized ACR
ORFS_00403.table	COG1755	Uncharacterized BCR. Probable membrane protein
ORFS_00404.table	COG0500	SAM-dependent methyltransferases (Secondary metabolites biosynthesis, transport and catabolism)
ORFS_00411.table	N/A	Uncharacterized BCR
ORFS_00414.table	COG1778	Uncharacterized proteins of HAD superfamily, CMP-Neu5Ac homologs
ORFS_00420.table	COG0037	Predicted ATPase of the PP-loop superfamily implicated in cell cycle control (Cell division and chromosome partitioning)
ORFS_00425.table	N/A	Uncharacterized BCR
ORFS_00426.table	COG3059	Uncharacterized membrane protein
ORFS_00427.table	COG1297	Uncharacterized membrane protein
ORFS_00428.table	N/A	Probable membrane protein

ORFS_00435.table	COG0729	Predicted outer membrane protein (Cell envelope biogenesis, outer membrane);Serine proteases, subtilase family, aspartic acid active site
ORFS_00436.table	COG2911	KDPG and KHG aldolases active site. Probable secretory protein
ORFS_00440.table	COG2816	NTP pyrophosphohydrolases containing a Zn-finger, probably nucleic-acid-binding (DNA replication, recombination and repair)
ORFS_00442.table	COG3068	Uncharacterized BCR
ORFS_00446.table	COG1399	Predicted metal-binding, possibly nucleic acid-binding protein
ORFS_00456.table	COG3610	Uncharacterized BCR. Probable membrane protein
ORFS_00458.table	COG1253	Uncharacterized CBS domain-containing proteins (Cystathionine beta-synthase (prototype for a family of repeats). Probable membrane protien
ORFS_00459.table	COG0260	Leucyl aminopeptidase (Amino acid transport and metabolism)
ORFS_00460.table	COG3084	Uncharacterized BCR
ORFS_00462.table	COG3085	Uncharacterized BCR
ORFS_00464.table	COG0500	SAM-dependent methyltransferases
ORFS_00469.table	COG3622	Hydroxypyruvate isomerase (Carbohydrate transport and metabolism)
ORFS_00471.table	COG3395	Uncharacterized BCR
ORFS_00475.table	COG1559	Predicted periplasmic solute-binding protein
ORFS_00489.table	N/A	
ORFS_00490.table	COG3120	Uncharacterized BCR
ORFS_00492.table	COG3036	Uncharacterized BCR
ORFS_00493.table	N/A	
ORFS_00494.table	N/A	

ORFS_00509.table	N/A	
ORFS_00518.table	COG0310	Cobalamin biosynthesis protein CbiM (COBALAMIN BIOSYNTHESIS, Coenzyme metabolism). Probable membrane protein
ORFS_00519.table	ATP/GTP-binding site motif A (P-loop).	Probable membrane protein
ORFS_00520.table	N/A	Probable membrane protein
ORFS_00542.table	N/A	
ORFS_00544.table	COG1739	Uncharacterized BCR; Uncharacterized protein family UPF0029 signature.
ORFS_00564.table	N/A	
ORFS_00567.table	COG2166	SufE protein probably involved in Fe-S center assembly
ORFS_00570.table	COG1380	Putative effector of murein hydrolase LrgA.
ORFS_00571.table	COG1346	Putative effector of murein hydrolase (Cell envelope biogenesis, outer membrane). Probable membrane protein
ORFS_00576.table	COG2990	Uncharacterized BCR
ORFS_00577.table	N/A	
ORFS_00580.table	N/A	azlD. Brancch-chain amino acid transport. Probable membrane protein
ORFS_00581.table	COG1296	redicted branched-chain amino acid permease (azaleucine resistance) (Amino acid transport and metabolism) (AzlC)
ORFS_00587.table	N/A	
ORFS_00588.table	N/A	
ORFS_00600.table	COG0564	Pseudouridylate synthases, 23S RNA-specific (Translation, ribosomal structure and biogenesis); Rlu family of pseudouridine synthase signature.
ORFS_00601.table	COG2431	Uncharacterized membrane protein
ORFS_00616.table	N/A	

ORFS_00620.table	COG2923	Uncharacterized protein involved in the oxidation of intracellular sulfur (Inorganic ion transport and metabolism)
ORFS_00621.table	COG1553	Uncharacterized ACR involved in intracellular sulfur reduction
ORFS_00622.table	COG2964	
ORFS_00624.table	COG0678 / COG0695	C-terminal: Peroxiredoxin / N-terminal: Glutaredoxin and related proteins; Glutaredoxin active site.
ORFS_00626.table	COG1309	Transcriptional regulator
ORFS_00633.table	COG0020	Undecaprenyl pyrophosphate synthase (Lipid metabolism)
ORFS_00635.table	COG0750	Predicted membrane-associated Zn-dependent proteases 1 (Cell envelope biogenesis, outer membrane); Neutral zinc metallopeptidases, zinc-binding region signature.
ORFS_00647.table	COG0127	Xanthosine triphosphate pyrophosphatase (Nucleotide transport and metabolism), Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) -- FATTY ACID BIOSYNTHESIS (fabG)
ORFS_00655.table	COG1028	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases). Probable secreted protein
ORFS_00673.table	COG0599	Uncharacterized ACR, homolog of gamma-carboxymuconolactone decarboxylase subunit
ORFS_00674.table	N/A	
ORFS_00683.table	COG0779	Uncharacterized BCR
ORFS_00715.table	COG0665	Glycine/D-amino acid oxidases (deaminating) (Amino acid transport and metabolism)—dadA

ORFS_00720.table	COG0061	Predicted kinase
ORFS_00723.table	N/A	
ORFS_00724.table	COG2917	Intracellular septation protein.
ORFS_00725.table	COG1607	Probable membrane protein
ORFS_00726.table	COG2350	Acyl-CoA hydrolase (Lipid metabolism); ATP/GTP-binding site motif A (P-loop).
ORFS_00730.table	N/A	Uncharacterized BCR
ORFS_00739.table	COG0325	Predicted enzyme with a TIM-barrel fold
ORFS_00745.table	COG2091	Phosphopantetheinyl transferase (Secondary metabolites biosynthesis, transport and catabolism)
ORFS_00750.table	COG1385	Uncharacterized BCR
ORFS_00751.table	COG1678	Putative transcriptional regulator
ORFS_00752.table	COG0816	Predicted endonuclease involved in recombination (possible Holliday junction resolvase in Mycoplasmas and B. subtilis)
ORFS_00763.table	COG1426	Uncharacterized BCR
ORFS_00766.table	COG2976	Uncharacterized BCR
ORFS_00768.table	COG0593	ATPase involved in DNA replication initiation; ATP/GTP-binding site motif A (P-loop).
ORFS_00773.table	COG2194	Predicted membrane-associated, metal-dependent hydrolase
ORFS_00808.table	COG2258	Uncharacterized BCR
ORFS_00809.table	COG3012	Uncharacterized BCR
ORFS_00811.table	COG1490	D-Tyr-tRNAtyr deacylase(Translation, ribosomal structure and biogenesis)
ORFS_00815.table	N/A	
ORFS_00816.table	N/A	
ORFS_00817.table	COG3083	Predicted hydrolase of alkaline phosphatase superfamily. Probable membrane protein

ORFS_00818.table	COG0402	Cytosine deaminase and related metal-dependent hydrolases (PYRIMIDINE SALVAGE)
ORFS_00827.table	N/A	
ORFS_00831.table	COG1959	Predicted transcriptional regulator
ORFS_00832.table	COG0565	rRNA methylase
ORFS_00840.table	COG1496	Uncharacterized ACR
ORFS_00841.table	COG1368	Phosphoglycerol transferase and related proteins, alkaline phosphatase superfamily (Cell envelope biogenesis, outer membrane)—mdoB.Turnover of phophatidylglycerol to synthesize membrane-derived oligosaccharides.
ORFS_00844.table	COG0762	Possible membrane protein
ORFS_00846.table	COG0500	Predicted integral membrane protein SAM-dependent methyltransferases (smtA), Secondary metabolites biosynthesis, transport and catabolism; ATP/GTP-binding site motif A (P-loop).
ORFS_00848.table	COG0845	Membrane-fusion protein (asrA)
ORFS_00849.table	COG1309	Transcriptional regulator; Bacterial regulatory proteins, tetR family signature.
ORFS_00855.table	Contains prokaryotic N-terminal methylation site	Probable secretory protein
ORFS_00856.table	N/A	
ORFS_00860.table	COG2813	16S RNA G1207 methylase RsmC; N-6 Adenine-specific DNA methylases signature.
ORFS_00877.table	COG0603	Predicted ATPase (PP-loop superfamily), confers aluminum resistance
ORFS_00882.table	COG1876	D-alanyl-D-alanine carboxypeptidase
ORFS_00884.table	COG1393	Arsenate reductase and related proteins, glutaredoxin family (ArsC)

ORFS_00889.table	COG2194	Predicted membrane-associated, metal-dependent hydrolase. putatively phase variable gene (dca)
ORFS_00893.table	N/A	Probable membrane protein
ORFS_00890.table	COG0477	Permeases of the major facilitator superfamily (ProP); ATP/GTP-binding site motif A (P-loop).
ORFS_00895.table	COG0802	Predicted ATPase or kinase
ORFS_00901.table	Lipocalin signature.	
ORFS_00902.table	COG3106	Predicted ATPase; ATP/GTP-binding site motif A (P-loop).
ORFS_00912.table	N/A	
ORFS_00916.table	COG1947	4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase (DEOXYXYLULOSE PATHWAY OF TERPENOID BIOSYNTHESIS) ispE
ORFS_00919.table	COG3103	SH3 domain protein
ORFS_00921.table	COG1392	Phosphate transport regulator (distant homolog of PhoU)
ORFS_00922.table	COG3025	Uncharacterized ACR
ORFS_00927.table	COG2256	Uncharacterized ATPase related to the helicase subunit of the Holliday junction resolvase; ATP/GTP-binding site motif A (P-loop).
ORFS_00946.table	COG1738	Uncharacterized ACR. Probable membrane protein
ORFS_00957.table	Ribonuclease T2 signiture	Probable secretory protein
ORFS_00960.table	COG3307	Lipid A core - O-antigen ligase and related enzymes. Probable membrane protein
ORFS_00962.table	COG0561	Predicted hydrolases of the HAD superfamily; Hypothetical cof family signature 1. / Hypothetical cof family signature 2.

ORFS_00970.table	COG1211	4-diphosphocytidyl-2-methyl-D-erithritol synthase (DEOXYXYLULOSE PATHWAY OF TERPENOID BIOSYNTHESIS) ispD
ORFS_00971.table	COG0245	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (DEOXYXYLULOSE PATHWAY OF TERPENOID BIOSYNTHESIS) ispF
ORFS_00972.table	COG0585	Uncharacterized ACR
ORFS_00978.table	COG3008	Paraquat-inducible protein B (PqiB); Shiga/ricin ribosomal inactivating toxins active site signature. Probable secretory protein
ORFS_00979.table	COG2995	Uncharacterized paraquat-inducible protein A (PqiA). Probable membrane protein
ORFS_00982.table	COG2989	Uncharacterized BCR; N-6 Adenine-specific DNA methylases signature.
ORFS_00983.table	COG3108	Uncharacterized BCR. Probable secretory protein
ORFS_00984.table	COG0491	Zn-dependent hydrolases, including glyoxylases (GloB)
ORFS_00999.table	COG0730	Predicted permeases. Probable membrane protein
ORFS_01002.table	COG0590	Cytidine and deoxycytidylate deaminases zinc-binding region signature. / Eukaryotic thiol (cysteine) proteases histidine active site.
ORFS_01004.table	N/A	Probable secretory protein
ORFS_01006.table	COG0454	Histone acetyltransferase HPA2 and related acetyltransferases (WecD)
ORFS_01009.table	N/A	
ORFS_01014.table	N/A	
ORFS_01015.table	COG3699	Predicted outer membrane protein; Serpins signature.

ORFS_01022.table	N/A	
ORFS_01024.table	COG0697	Permeases of the drug/metabolite transporter (DMT) superfamily (RhaT). Probable membrane protein
ORFS_01037.table	COG2259	Predicted membrane protein
ORFS_01039.table	N/A	
ORFS_01047.table	COG2860	Predicted membrane protein
ORFS_01048.table	N/A	Aldehyde dehydrogenases cysteine active site. Possible membrane protein
ORFS_01052.table	COG3099	Uncharacterized BCR
ORFS_01055.table	COG1187	16S rRNA uridine-516 pseudouridylate synthase and related pseudouridylate synthases (RsuA); Rsu family of pseudouridine synthase signature.
ORFS_01056.table	COG0009	Putative translation factor (SUA5)
ORFS_01059.table	COG0778	Nitroreductase (NfnB)
ORFS_01067.table	COG2707	Uncharacterized membrane protein
ORFS_01068.table	COG2363	Uncharacterized small membrane protein
ORFS_01083.table	COG0316	Uncharacterized ACR
ORFS_01091.table	COG2855	Uncharacterized membrane protein
ORFS_01099.table	COG2606	Uncharacterized ACR
ORFS_01100.table	COG0534	Na+-driven multidrug efflux pump. Possible membrane protein
ORFS_01102.table	COG3636	Predicted transcriptional regulator
ORFS_01103.table	COG1534	Predicted RNA-binding protein containing KH domain, possibly ribosomal protein
ORFS_01124.table	COG3092	Uncharacterized BCR
ORFS_01126.table	COG0220	Predicted S-adenosylmethionine-dependent methyltransferase
ORFS_01127.table	COG3171	Uncharacterized BCR
ORFS_01131.table	COG0042	Predicted TIM-barrel enzymes, possibly dehydrogenases, nifR3 family

ORFS_01135.table	COG1561	Uncharacterized stress-induced protein
ORFS_01136.table	COG0354	Predicted aminomethyltransferase related to GcvT
ORFS_01155.table	N/A	
ORFS_01158.table	N/A	
ORFS_01197.table	N/A	
ORFS_01206.table	N/A	
ORFS_01228.table	COG0319	Predicted metal-dependent hydrolase
ORFS_01245.table	N/A	Possible membrane protein
ORFS_01261.table	N/A	
ORFS_01282.table	N/A	
ORFS_01284.table	COG0009	Putative translation factor (SUA5)
ORFS_01285.table	COG1322	Uncharacterized BCR
ORFS_01286.table	COG2862	Uncharacterized membrane protein
ORFS_01288.table	COG3013	Uncharacterized ACR
ORFS_01291.table	COG0739	Membrane proteins related to metalloendopeptidases (NlpD)
ORFS_01292.table	COG2861	Uncharacterized BCR. Possible membrane protein
ORFS_01311.table	N/A	
ORFS_01312.table	N/A	
ORFS_01318.table	COG0241	Histidinol phosphatase and related phosphatases (HisB)
ORFS_01320.table	N/A	
ORFS_01323.table	N/A	
ORFS_01331.table	COG0251	Putative translation initiation inhibitor
ORFS_01332.table	Zinc carboxypeptidases, zinc-binding region 2 signature.	Possible membrane protein
ORFS_01333.table	COG0790	TPR repeat proteins (); ATP/GTP-binding site motif A (P-loop). Possible secretory protien
ORFS_01339.table	COG3683	Predicted periplasmic metal-binding protein
ORFS_01340.table	COG2215	Predicted membrane protein; ATP/GTP-binding site motif A (P-loop).

ORFS_01344.table	COG3055	Uncharacterized BCR. Probable secretory protein
ORFS_01345.table	COG1593	Integral membrane protein, possible transporter
ORFS_01346.table	COG1638	Dicarboxylate-binding periplasmic protein (DctP)
ORFS_01349.table	N/A	Possible membrane protein
ORFS_01364.table	N/A	
ORFS_01366.table	N/A	
ORFS_01367.table	COG0589	Universal stress protein UspA and related nucleotide-binding proteins
ORFS_01384.table	COG0084	Mg-dependent Dnase (TatD); TatD deoxyribonuclease family signature 1.
ORFS_01391.table	N/A	
ORFS_01394.table	COG1576	Uncharacterized ACR
ORFS_01395.table	COG0799	Uncharacterized ACR (homolog of plant lojap proteins)
ORFS_01416.table	COG2081	Predicted flavoproteins
ORFS_01438.table	COG0697	Permeases of the drug/metabolite transporter (DMT) superfamily (RhaT). Possible membrane protein
ORFS_01446.table	N/A	Possible membrane protein
ORFS_01454.table	COG2983	Uncharacterized ACR
ORFS_01458.table	COG1538	Outer membrane protein
ORFS_01464.table	COG0577	ABC-type transport systems, involved in lipoprotein release, permease components
ORFS_01466.table	COG0577	ABC-type transport systems, involved in lipoprotein release, permease components
ORFS_01475.table	COG1280	Putative threonine efflux protein (RhtB). Possible membrane protein
ORFS_01505.table	COG1944	Uncharacterized ACR
ORFS_01515.table	COG0477	Permeases of the major facilitator superfamily. Possible membrane protein

ORFS_01517.table	COG2252	Permeases (); ATP/GTP-binding site motif A (P-loop). Possible membrane protein
ORFS_01526.table	N/A	
ORFS_01528.table	N/A	
ORFS_01533.table	COG3065	Starvation-inducible outer membrane lipoprotein (Slp)
ORFS_01534.table	COG1214	Inactive homologs of metal-dependent proteases, putative molecular chaperones ()
ORFS_01536.table	COG0676	Uncharacterized enzymes related to aldose 1-epimerase ()
ORFS_01537.table	COG3118	Thioredoxin domain-containing protein
ORFS_01542.table	COG3642	Mn2+-dependent serine/threonine protein kinase
ORFS_01570.table	COG0463	Glycosyltransferases involved in cell wall biogenesis
ORFS_01577.table	COG1434	Uncharacterized ACR. Possible membrane protein
ORFS_01582.table	COG0795	Predicted permeases. Possible membrane protein
ORFS_01588.table	COG0471	Di- and tricarboxylate transporters (CitT); Sodium:sulfate symporter family signature. Possible membrane protein
ORFS_01592.table	COG1327	Predicted transcriptional regulator, consists of a Zn-ribbon and ATP-cone domains; Cytochrome c family heme-binding site signature.
ORFS_01593.table	COG2933	Predicted SAM-dependent methyltransferase
ORFS_01600.table	COG2209	Na+-transporting NADH:ubiquinone oxidoreductase subunit 5 (NqrE). Possible membrane protein
ORFS_01601.table	COG2878	Predicted alternative beta subunit of Na+-transporting NADH:ubiquinone

ORFS_01605.table	COG1347	oxidoreductase. Possible secretory protein
ORFS_01607.table	N/A	Na+-transporting NADH:ubiquinone oxidoreductase subunit 4 (NqrD). Possible membrane protein
ORFS_01610.table	COG0739	Membrane proteins related to metalloendopeptidases (NlpD)
ORFS_01621.table	N/A	
ORFS_01629.table	COG0251	Putative translation initiation inhibitor (TdcF)
ORFS_01640.table	COG1757	Na+/H+ antiporter (NhaC). Possible membrane protein
ORFS_01645.table	COG3101	Uncharacterized BCR; Neutral zinc metallopeptidases, zinc-binding region signature.
ORFS_01651.table	COG0780	Enzyme related to GTP cyclohydrolase I; Chitinases family 18 active site.
ORFS_01652.table	COG0564	Pseudouridylate synthases, 23S RNA-specific (RluA)
ORFS_01653.table	N/A	Rlu family of pseudouridine synthase signature.
ORFS_01660.table	N/A	
ORFS_01662.table	COG3381	Uncharacterized component of anaerobic dehydrogenases (TorD)
ORFS_01673.table	COG1253	Uncharacterized CBS domain-containing proteins
ORFS_01676.table	N/A	
ORFS_01680.table	COG2915	Uncharacterized protein involved in purine metabolism
ORFS_01686.table	N/A	
ORFS_01687.table	N/A	
ORFS_01705.table	COG2850	Uncharacterized ACR
ORFS_01707.table	COG1660	Predicted P-loop-containing kinase
ORFS_01710.table	COG1934	Uncharacterized BCR. Possible

		membrane protein
ORFS_01711.table	COG3117	Uncharacterized BCR. Possible membrane protein
ORFS_01713.table	COG0767	ABC-type toluene export system, permease component. Possible membrane protein
ORFS_01714.table	COG1463	Permease component of an ABC-transporter. Possible membrane protein
ORFS_01715.table	COG2854	Uncharacterized periplasmic protein
ORFS_01716.table	COG3113	STAS domain protein
ORFS_01724.table	COG0861	Membrane protein TerC, possibly involved in tellurium resistance (terC). Possible membrane protein
ORFS_01728.table	COG0859	ADP-heptose:LPS heptosyltransferase (RfaF)
ORFS_01744.table	COG0477	Permeases of the major facilitator superfamily (proP). Possible membrane protein
VII.2 Escherichia coli		
ORFS_00051.table	COG1073	Hydrolases of the alpha/beta superfamily; Possible membrane protein. DLH, Dienelactone hydrolase family
ORFS_00042.table	COG1011	Predicted hydrolases of the HAD superfamily; Signal peptidases I signature 3. (YECHAGH)
ORFS_00122.table	Hypothetical	Probable membrane protein
ORFS_00249.table	COG3304	Uncharacterized membrane protein
ORFS_00281.table	Hypothetical cof family signature 1. / Hypothetical cof family signature 2.	
ORFS_00423.table	COG1011	Predicted hydrolases of the HAD superfamily
ORFS_00424.table	COG0494	NTP pyrophosphohydrolases including oxidative damage repair enzymes
ORFS_00433.table	N/A	
ORFS_00467.table	COG0561	Predicted hydrolases of the HAD

ORFS_00487.table	COG1553	superfamily
ORFS_00491.table	COG0621	Uncharacterized ACR involved in intracellular sulfur reduction 2-methylthioadenine synthetase (Translation, ribosomal structure and biogenesis); Uncharacterized protein family UPF0001 signature.
ORFS_00554.table	COG0391	
ORFS_00561.table	N/A	
ORFS_00585.table	COG2189	Adenine specific DNA methylase Mod (DNA replication, recombination and repair)
ORFS_00586.table	N/A	
ORFS_00643.table	COG0471	Di- and tricarboxylate transporters (Inorganic ion transport and metabolism)
ORFS_00666.table	COG0659	Sulfate permease and related transporters (MFS superfamily)-- Inorganic ion transport and metabolism. Probable membrane protien
ORFS_00714.table	N/A	
ORFS_00807.table	COG0433	Predicted ATPase; ATP/GTP-binding site motif A (P-loop).
ORFS_00874.table	COG0116	Predicted N6-adenine-specific DNA methylases; N-6 Adenine-specific DNA methylases signature. / Uncharacterized protein family UPF0020 signature.
ORFS_00983.table	N/A	
ORFS_01147.table	COG2220	Predicted Zn-dependent hydrolases of the beta-lactamase fold
ORFS_01148.table	COG1349	Transcriptional regulators of sugar metabolism (GlpR); Bacterial regulatory proteins, deoR family signature.

ORFS_01191.table	COG1830	DhnA-type fructose-1,6-bisphosphate aldolase and related enzymes (FbaB)
ORFS_01230.table	COG1444	Predicted P-loop ATPase fused to an acetyltransferase
ORFS_01242.table	COG0702	Predicted nucleoside-diphosphate-sugar epimerases
ORFS_01304.table	N/A	
ORFS_01327.table	COG3021	Uncharacterized BCR. Possible secretory protein
ORFS_01358.table	N/A	Possible membrane protein
ORFS_01360.table	COG0845	Membrane-fusion protein (AcrA). Possibile membrane protein
ORFS_01427.table	N/A	
ORFS_01509.table	COG0500	SAM-dependent methyltransferases (SmtA)
ORFS_01513.table	N/A	
ORFS_01529.table	COG3059	Uncharacterized membrane protein. Possible membrane protein
ORFS_01544.table	N/A	Probable secretory protein
ORFS_01576.table	COG2161	Uncharacterized ACR
ORFS_01581.table	N/A	Possible membrane protein
ORFS_01589.table	COG1896	Predicted hydrolases of HD superfamily
ORFS_01617.table	COG3111	Uncharacterized ACR. Probable secrewtory protien
ORFS_01674.table	COG0755	ABC-type transport system involved in cytochrome c biogenesis, permease component (CcmC)

VII.2 Neisseria meningitidis

ORFS_00029.table	COG0561	Predicted hydrolases of the HAD superfamily; Hypothetical cof family signature 2
ORFS_00141.table	N/A	
ORFS_00335.table	N/A	
ORFS_00162.table	COG0697	Permeases of the drug/metabolite transporter (DMT) superfamily

ORFS_00401.table	COG1914	Mn2+ and Fe2+ transporters of the NRAMP family (Inorganic ion transport and metabolism)
ORFS_00432.table	ATP/GTP-binding site motif A (P-loop).	
ORFS_00439.table	N/A	Probable secretory protein
ORFS_00439.table	N/A	
ORFS_00499.table	COG0042	Predicted TIM-barrel enzymes, possibly dehydrogenases, nifR3 family; Uncharacterized protein family UPF0034 signature.
ORFS_00503.table	COG1288	Uncharacterized membrane protein
ORFS_00555.table	N/A	
ORFS_00784.table	COG0204	1-acyl-sn-glycerol-3-phosphate acyltransferase (plsC). Probable secretory protein
ORFS_00786.table	N/A	Probable membrane protein
ORFS_00795.table	COG0764	3-hydroxymyristoyl/3-hydroxydecanoyl-(acyl carrier protein) dehydratases. FabA
ORFS_00878.table	N/A	
ORFS_00890.table	N/A	
ORFS_00891.table	COG0421	Spermidine synthase. Probable membrane protein
ORFS_01256.table	N/A	
ORFS_01274.table	N/A	
VII.3 Bacillus sp.		
ORFS_00035.table	N/A	
ORFS_00560.table	N/A	
ORFS_00697.table	N/A	
ORFS_00778.table	N/A	Probable membrane protein
ORFS_01035.table	COG1611	Predicted Rossmann fold nucleotide-binding protein
ORFS_01077.table	COG0577	ABC-type transport systems, involved in lipoprotein release, permease components

ORFS_01081.table	COG0672	High-affinity iron permease (FTR1); ATP/GTP-binding site motif A (P-loop). Possible membrane protien
ORFS_01204.table	N/A	
ORFS_01314.table	N/A	
ORFS_01499.table	N/A	
ORFS_01719.table	COG0463	Glycosyltransferases involved in cell wall biogenesis (WcaA); Eukaryotic and viral aspartyl proteases active site.
VII.4 Pseudomonas aeruginosa		
ORFS_00138.table	COG3471	Predicted periplasmic/secreted protein
ORFS_00703.table	N/A	
ORFS_00799.table	N/A	
ORFS_00806.table	N/A	
ORFS_01112.table	COG0687	Spermidine/putrescine-binding periplasmic protein (PotD); ATP/GTP-binding site motif A (P-loop).
ORFS_01263.table	COG3177	Uncharacterized BCR
ORFS_01703.table	COG2510	Predicted membrane protein
VII.5 Vibrio cholerae		
ORFS_00274.table	COG0053	Predicted Co/Zn/Cd cation transporters
ORFS_00275.table	COG0053	Predicted Co/Zn/Cd cation transporters
ORFS_00347.table	COG3001	Uncharacterized BCR
ORFS_00455.table	COG2966	Uncharacterized BCR; Phosphopantetheine attachment site. Probable membrane protein
ORFS_00480.table	N/A	
ORFS_00495.table	N/A	Probable membrane protein
ORFS_00665.table	COG0697	Permeases of the drug/metabolite transporter (DMT) superfamily (for amino acid and carbohydrate transport). Probable membrane protein

ORFS_00767.table	COG2020	Putative protein-S-isoprenylcysteine methyltransferase (Posttranslational modification, protein turnover, chaperones). Probable membrane protein
ORFS_00828.table	COG0316	Uncharacterized ACR; Hypothetical hesB/yadR/yfhF family signature.
ORFS_00843.table	COG1872	Uncharacterized ACR
ORFS_01389.table	COG2921	Uncharacterized BCR
ORFS_01704.table	COG2867	Oligoketide cyclase/lipid transport protein
ORFS_01727.table	COG1011	Predicted hydrolases of the HAD superfamily; Binding-protein-dependent transport systems inner membrane comp. sign.
ORFS_01087.table	COG3037	Uncharacterized ACR; ATP/GTP-binding site motif A (P-loop). Probable membrane protein
VII.6 Other species		
ORFS_01107.table (Caenorhabditis elegans)	COG3177	Uncharacterized BCR
ORFS_01638.table (Chlamydia muridarum)	N/A	
ORFS_00562.table (Deinococcus radiodurans)	COG1585	Membrane protein implicated in regulation of membrane protease activity (Cell motility and secretion; Posttranslational modification, protein turnover, chaperones)
ORFS_00406.table (Herpetosiphon aurantiacus)	COG0477	Permeases of the major facilitator superfamily (carbohydrate,amino acid , and inorganic ions transport and metabolism)
ORFS_00876.table (Mycobacterium tuberculosis)	COG1373	Uncharacterized ATPases of the AAA superfamily; ATP/GTP-binding site motif A (P-loop).

ORFS_01350.table (Methanosarcina barkeri)	COG0614	ABC-type Fe3+-siderophores transport systems, periplasmic components (FecB)
ORFS_01678.table (Paenibacillus popilliae)	COG1028 / COG0300	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases) (FabG) / Short-chain dehydrogenases of various substrate specificities (DltE); Short-chain dehydrogenases/reductases family signature.
ORFS_00166.table (Aquifex aeolicus)	Hypothetical	Probable membrane protein
ORFS_01681.table (Aquifex aeolicus)	COG2847	Uncharacterized BCR
ORFS_00800.table (Alcaligenes sp.)	N/A	
ORFS_01092.table (Rickettsia prowazekii)	COG2373	Large extracellular alpha-helical protein.possible membrane protein
ORFS_00081.table (Campylobacter jejuni)	Hypothetical	Probable membrane protein
ORFS_00157.table (Campylobacter jejuni)	COG2964	Uncharacterized BCR
ORFS_00853.table (Campylobacter jejuni)	N/A	
ORFS_01080.table (Campylobacter jejuni)	N/A	
ORFS_01521.table (Campylobacter jejuni)	COG2351	Transthyretin-like protein. Probable secretory protein
ORFS_00939.table (Cyanidium caldarium)	COG0583	Transcriptional regulator; Bacterial regulatory proteins, lysR family signature.
ORFS_01632.table (Chlorobium tepidum)	N/A	
ORFS_01633.table (Chlorobium tepidum)	N/A	

ORFS_00940.table (Ralstonia eutropha)	N/A	
ORFS_00334.table (Helicobacter pylori)	COG0846	NAD-dependent protein deacetylases, SIR2 family
ORFS_00879.table (Helicobacter pylori)	N/A	
ORFS_00185.table (Sinorhizobium meliloti)	N/A	
ORFS_00121.table (Methanococcus jannaschii)	COG2095	Integral membrane proteins of the MarC family
ORFS_01308.table (Methanococcus jannaschii)	COG0603	Predicted ATPase (PP-loop superfamily), confers aluminum resistance ()
ORFS_01429.table (Methanococcus jannaschii)	N/A	
ORFS_00126.table (Ureaplasma urealyticum)	COG3600	Uncharacterized phage-associated protein
ORFS_00127.table (Streptomyces coelicolor)	COG2357/COG0317	Uncharacterized BCR / Guanosine polyphosphate pyrophosphohydrolases/synthetases (Transcription)
ORFS_00255.table (Streptomyces coelicolor)	COG3037	Uncharacterized BCR; Protein kinases ATP-binding region signature.
ORFS_00131.table (Thermotoga maritima)	COG1434 / COG0457	Uncharacterized Ancient Conserved Region / repeat-containing proteins (mediates protein-protein interactions).
		Probabe secretory protein
ORFS_00803.table (Thermotoga maritima)	N/A	
ORFS_00867.table (Thermotoga maritima)	N/A	
ORFS_00220.table (Pasteurella multocida)	COG0621	2-methylthioadenine synthetase (Translation, ribosomal structure and biogenesis); Uncharacterized protein family UPF0004 signature.

ORFS_00468.table (Pasteurella multocida)	N/A	Probable membrane protein
ORFS_01637.table (Pyrococcus horikoshii)	N/A	
ORFS_01622 (Pasteurella multocida)	COG0642	Sensory transduction histidine kinases (BaeS). Possible membrane protein
ORFS_00484.table (Synechocystis sp.)	COG0500	SAM-dependent methyltransferases
ORFS_01462.table (Synechocystis sp.)	ATP/GTP-binding site motif A (P-loop) Two motifs	
ORFS_01300.table (Streptococcus pneumoniae)	N/A	
ORFS_00219.table (Actinobacillus pleuropneumoniae)	N/A	
ORFS_01375.table (Actinobacillus pleuropneumoniae)	N/A	
ORFS_01596.table (Moraxella bovis)	N/A	Possible membrane protein
ORFS_00687.table (Archaeoglobus fulgidus)	N/A	
ORFS_00688.table (Archaeoglobus fulgidus)	N/A	
ORFS_01079.table (Acanthamoeba castellanii)	N/A	Probable membrane protein
ORFS_01554.table (Salmonella sp.)	N/A	
ORFS_01550.table (Salmonella sp.)	Epimerase domain; 3Beta_HSD, 3-beta hydroxysteroid dehydrogenase/isomerase family; Polysacc_synt_2, Polysaccharide biosynthesis protein.	Possible secretory protein
ORFS_00789.table (Xylella fastidiosa)	COG1560	Lauroyl/myristoyl acyltransferase involved in lipid A biosynthesis
ORFS_01074.table (Yersinia pestis)	COG2863	Cytochrome c553; Cytochrome c family heme-binding site signatur

ORFS_01309.table (Borrelia burgdorferi)	COG3613	Nucleoside 2-deoxyribosyltransferase ()
VII.7 Unique		
ORFS_00093.table	N/A	
ORFS_00184.table	COG1373	uncharacterized ATPases of the AAA superfamily; ATP/GTP-binding site motif A (P-loop).
ORFS_00186.table	N/A	
ORFS_00187.table	N/A	
ORFS_00194.table	N/A	
ORFS_00212.table	N/A	
ORFS_00213.table	N/A	
ORFS_00215.table	N/A	
ORFS_00216.table	N/A	
ORFS_00217.table	N/A	
ORFS_00225.table	COG1359	Uncharacterized Ancient Conserved Region
ORFS_00226.table	N/A	
ORFS_00301.table	N/A	
ORFS_00329.table	N/A	
ORFS_00369.table	N/A	
ORFS_00405.table	N/A	Probable membrane protein
ORFS_00465.table	ATP/GTP-binding site motif A (P-loop).	
ORFS_00511.table	N/A	
ORFS_00513.table	N/A	
ORFS_00574.table	ATP/GTP-binding site motif A (P-loop).	
ORFS_00615.table	COG1752	Predicted esterase of the alpha-beta hydrolase superfamily. Probable secretory protein
ORFS_00664.table	N/A	
ORFS_00692.table	N/A	
ORFS_00743.table	ATP/GTP-binding site motif A (P-loop).	
ORFS_00754.table	N/A	
ORFS_00783.table	N/A	

ORFS_00790.table	COG0824	Predicted thioesterase
ORFS_00791.table	N/A	Probable secretory protein
ORFS_00792.table	COG1033	Predicted exporters of the RND superfamily.
ORFS_00793.table	N/A	
ORFS_00834.table	N/A	
ORFS_00861.table	N/A	
ORFS_00894.table	N/A	Probable membrane protein
ORFS_00908.table	COG2926	
ORFS_00942.table	N/A	
ORFS_00945.table	N/A	
ORFS_00953.table	N/A	Probable membrane protein
ORFS_01021.table	N/A	
ORFS_01043.table	N/A	
ORFS_01045.table	N/A	
ORFS_01070.table	N/A	Probable membrane protein
ORFS_01071.table	N/A	
ORFS_01121.table	COG1917	Uncharacterized ACR, double-stranded beta-helix domain
ORFS_01137.table	N/A	
ORFS_01144.table	N/A	
ORFS_01160.table	N/A	
ORFS_01161.table	N/A	
ORFS_01213.table	COG1579	Zn-ribbon protein, possibly nucleic acid-binding ; Lipases, serine active site.
ORFS_01223.table	COG0455	ATPases involved in chromosome partitioning ()
ORFS_01227.table	COG0477	Permeases of the major facilitator superfamily (ProP). Possible membrane protien
ORFS_01229.table	N/A	Probable secretory protein
ORFS_01238.table	N/A	
ORFS_01243.table	N/A	
ORFS_01255.table	N/A	

ORFS_01294.table	N/A	
ORFS_01297.table	N/A	
ORFS_01298.table	ATP-dependent DNA ligase AMP-binding site.	
ORFS_01299.table	N/A	
ORFS_01303.table	N/A	
ORFS_01307.table	COG0778	Nitroreductase (NfnB)
ORFS_01329.table	N/A	Possible membrane protein
ORFS_01330.table	N/A	Possible membrane protein
ORFS_01372.table	N/A	
ORFS_01408.table	N/A	
ORFS_01417.table	N/A	
ORFS_01420.table	N/A	
ORFS_01421.table	N/A	
ORFS_01423.table	COG1204	Predicted helicases
ORFS_01424.table	N/A	
ORFS_01425.table	N/A	
ORFS_01426.table	N/A	
ORFS_01430.table	N/A	
ORFS_01431.table	N/A	
ORFS_01432.table	N/A	
ORFS_01433.table	N/A	
ORFS_01436.table	N/A	
ORFS_01459.table	N/A	
ORFS_01460.table	N/A	
ORFS_01463.table	N/A	
ORFS_01478.table	N/A	Possible membrane protein
ORFS_01522.table	N/A	
ORFS_01523.table	N/A	
ORFS_01524.table	N/A	
ORFS_01527.table	N/A	
ORFS_01543.table	N/A	
ORFS_01548.table	N/A	Possible membrane protein
ORFS_01551.table	N/A	

ORFS_01556.table	N/A	Possible membrane protein
ORFS_01558.table	COG2246	Uncharacterized membrane protein
ORFS_01559.table	N/A	Possible membrane protein
ORFS_01560.table	COG0451	Nucleoside-diphosphate-sugar epimerases (WcaG)
ORFS_01561.table	N/A	
ORFS_01564.table	COG1682	Membrane permeases involved in cell wall biosynthesis;ABC-2 type transport system integral membrane proteins signature.
ORFS_01572.table	N/A	
ORFS_01573.table	COG3306	Glycosyltransferase involved in LPS biosynthesis
ORFS_01574.table	N/A	
ORFS_01620.table	N/A	
ORFS_01634.table	N/A	
ORFS_01635.table	N/A	
ORFS_01636.table	N/A	
ORFS_01639.table	N/A	
ORFS_01659.table	N/A	
ORFS_01669.table	N/A	
ORFS_01693.table	N/A	Probable secretory protein
ORFS_01698.table	N/A	
ORFS_00041.table	Immunoglobulins and major histocompatibility complex proteins signature	
ORFS_01699.table	N/A	
ORFS_01725.table	N/A	

Appendix-B. Regions of significantly different GC content in *A. actinomycetemcomitans.*

Homology	***Organism***	**E value**	**GC%**
aspartate-semialdehyde dehydrogenase	*H. inf*	0e+00	49.3
OmpH	*P. multocida*	5e-31	35.45
ORFs_00157; Hypothetical; Uncharacterized BCR (COG2964)	*Campylobacter jejuni*	7e-59	35.54
ORFs_ 00185; Hypothetical	*Sinorhizobium meliloti*	3e-14	32
ORFs_ 00186; Unique	*N/A*	N/A	32
ORFs_ 00187; Unique	*N/A*	N/A	29.61
ORFs_00403; Hypotetical; Uncharacterized BCR. Probable membrane protein (COG1755)	*H. inf.*	2e-52	26
ORFS_00432; Hypothetical; ATP/GTP-binding site motif A (P-loop).	*N.meningitidis*	1e-18	33.3
ORFs_ 00434 ; VapD	*Xylella fastidiosa*	4e-79	36.6
ORFs_00513; Unique	*N/A*	N/A	35.8
ORFs_ 00468; Hypothetical; Probable membrane protein	*P. multocida*	8e-75	36.8
ORFS_00574; Unique ; ATP/GTP-binding site motif A (P-loop).	*N/A*	N/A	26.7
Uronate isomerase	*E. coli*	0e00	36.6
D-arabinitol 2-dehydrogenase	*H. inf.*	0e00	
2-dehydro-3-deoxygluconokinase (kdgK)	*H. inf*	0.00	
C4-dicarboxylate periplasmic binding protein precursor (dctP)	*K. pneumoniae*	7e-58	
integral C4-dicarboxylate membrane transport protein (dctQ)	*K. pneumoniae*	3e-26	

Hypothetical	*S. typhimurium*	0e00	
glucosidase II	*Sus scrofa*	1e-75	
uxu operon regulator (uxuR)	*H. inf.*	4e-80	
mannonate dehydratase (uxuA)	*H. inf.*	8e-61	
mannonate dehydratase (uxuA)	*H. inf.*	0e00	
Unique	N/A		
Hypothetical	*Pyrococcus abyssi*	8e-23	
modification methylase, type III R/M system	Archaeoglobus fulgidus	3e-78	35.2
ORFs_00692; Unique	N/A	N/A	30.6
Hemoglobin & haptoglobin binding protein	*H. inf.*	0.00	36.2
Hypothetical (806915..807673)	*H. inf.*	1e-32	36.9
ORFs_00852; protein envolved in LPS biosynthesis	*E. coli*	3e-65	26.4
ORFs_00853; Hypotheical	*Campylobacter jejuni*	2e-35	30.14
eemB. Hypothetical protein envolved in LPS biosynthesis			36
ORFs_00889; Hyothetical; Predicted membrane-associated, metal-dependent hydrolase. putatively phase variable gene (dca) (COG2194)	*H. inf.*	2e-83	31.4
ORFs_00890; Hypothetical; Permeases of the major facilitator superfamily (ProP); ATP/GTP-binding site motif A (P-loop) (COG0477)	***N. meningitidis***	2e-53	
ORFs_00891; Hypothetical; Spermidine synthase. Probable membrane protein (COG0421)	*N. meningitidis*	0e00	
CDP-abequose synthase (RfbJ)	*Synechocystis sp.*	2e-69	
ORFs_00893; Hypothetical; Probable membrane protein	Bacteriophage SfII	5e-21	
ORFs_00894; Unique	N/A	N/A	
ORFs_ 00939; Hypothetical; Transcriptional regulator; Bacterial regulatory proteins, lysR family signature (COG0583)	red alga (Cyanidium caldarium	6e-35	34
o-succinylbenzoate synthase (menC)	*H. inf.*	0e00	49.4

VirB4	***Brucella abortus***	2e-42	36.4
ORFs_01070; Unique	N/A	N/A	24
Tad operon (1385228..1396676)			35
ORFs_01303; Unique			26
ParB protein	*Campylobacter jejuni*	8e-07	33.3
ORFs_01299; Unique			32.3
adenine-N6--methyltransferase	***Enterobacter aerogenes***	3e-42	34
ORFs_01306; Endonuclease III (DNA-(apurinic or apyrimidinic site) lyase)	*M. jannaschii*	6e-11	
ORFS_01307; Unique; Nitroreductase (NfnB) (COG0778)			
ORFs_01308; Hypothetical; Predicted ATPase (PP-loop superfamily), confers aluminum resistance (COG0603)	*M. jannaschii*	2e-13	
ORFs_01309; Hypothetical; Nucleoside 2-deoxyribosyltransferase (COG3613)	*B. burgdorferi*	2e-07	
dolichol phosphate mannose	*B. subtilis*	8e-68	30
ORFs_01329; Unique; Possible membrane protein			
ORFs_01330; Unique; Possible membrane protein			
type III DNA modification enzyme(methyltransferase)	*H. pylori*	7e-61	32
ORFs_01423; Unique; Predicted helicases (COG1204)			
ORFs_01424; Unique			
ORFs_01425; Unique			
ORFs_01426; Unique			
ORFs_01427; Hypothetical	*E. coli*	4e-12	
UvrD/REP helicase	***Chlamydia muridarum***	3e-56	
purine NTPase	***M. jannaschii***	1e-06	
ORFs_01430; Unique			

ORFs_01431; Unique			
ORFs_01433; Unique			
Hypothetical	*Synechocystis sp.*	4e-15	24
ORFs_01463; Unique			27
Unique (1719798..1720454)			
Transposon			35.6
ORFs_01522 ; Unique			35.6
endo alpha-1,4polygalactosaminidase	*Pseudomonas sp.*	5e-05	32.4
F200	*E. coli*	2e-39	29.8
Glycosyl transferase (rfaG) (1815691..1816560)	*Deinococcus radiodurans*	2e-25	28.4
ORFS_01556 Unique	Possible membrane protein		
Glycosyl transferase; lipopolysaccharide biosynthesis protein (lgtF)	*Enterococcus faecalis*	1e-132	
Unique; Uncharacterized membrane protein (COG2246)			
ORFS_01559 ; Unique; Possible membrane protein			
ORFS_01560 ; Unique ; Nucleoside-diphosphate-sugar epimerases (COG0451)			
ORFS_01561 ; Unique; (1820008..1821264)			
Rhamnosyltransferase	*Salmonella typhimurium*	4e-23	
ABC-type polysaccharide/polyol phosphate transport system, ATPase component (COG1134)			
(1822972..1823763)			