Amplification, Cloning, Expression, and Purification of purQ gene from *Staphylococcus aureus;* along with an evolutionary study of the multi-subunit form of the FGAR amidotransferase enzyme using bioinformatics.

Presented to the Faculty of Lycoming College
In Partial Fulfillment of the Requirements
For Departmental Honors in Biology

By
Deanne M. Greene
Lycoming College
April 21, 2000

Approved by:

Jeffrey D. Newman

Dr. Jeffrey Newman, Department of Biology, Lycoming College

Edward D. Gabriel

Dr. Edward Gabriel, Department of Biology, Lycoming College

Jack D. Diehl, Jr.

Dr. Jack Diehl, Department of Biology, Lycoming College

Dr. Holly Bendorf, Department of Chemistry, Lycoming College

Brenda Terry Manchester MSN, BSN.

Ms. Brenda Terry Manchester, Department of Nursing, Lycoming College

Abstract

The purine biosynthetic pathway of most pathogenic bacteria differs from that of eukaryotes. A multi- subunit enzyme that converts FGAR, 5'- phosphoribosyl N-formyglycinamide into FGAM, 5'- phosphoribosyl N- formyglycinamidine has been found in many prokaryotes. The FGAR amidotransferase enzyme of many pathogenic bacteria is comprised of three subunits encoded by the *purQ*, *purY*, and *purL* genes; the human enzyme is a single subunit encoded by *purL*. This difference in purine synthesis can be used in the development of a new antibiotic. The goal of this study was the amplification, cloning, expression and purification of *purQ* for later biochemical studies. In order to get to the expression and purification stage, *purQ* was amplified and cloned into pTYB2, an expression vector used in the IMPACT™ T7 protein purification system developed by New England BioLabs. This construct was subsequently transformed into *E. coli* strain ER2566 and expressed. A supplementary study of protein sequences from various organisms was also completed to determine the evolutionary pattern of the type I and type II pathways.

Introduction

Pathogenic bacteria are developing resistance to many of the antibiotics prescribed to help the human body fight bacterial infections. Without antibiotics to fight and kill these organisms, medicine will not be able to cure the infections caused by these resistant strains. Pathogens are a fast growing menace; many strains have already developed this resistance and are prevalent in hospitals where patients are susceptible to these nosocomial infections. One such microbe known for achieving resistance is *Staphylococcus aureus*. These strains pose an even greater threat in their ability to pass

this resistance to other strains through a process known as conjugation. Resistant bacteria have encouraged a growth in research for new antibiotics and alternative ways to fight these resistant strains. A recent article in Scientific American demonstrates how prevalent resistance is.

> Last year an event doctors had been fearing finally occurred. In three geographically separate patients, an often-deadly bacterium, *Staphylococcus aureus*, responded poorly to a once reliable antidote the antibiotic vacomycin. *S. aureus*, a major cause of hospital-acquired infections, has thus moved one step closer to becoming an unstoppable killer. The looming threat of incurable *S. aureus* is just the latest twist in an international public health nightmare: increasing bacterial resistance to many antibiotics that once cured bacterial diseases readily. Strains of at least three bacterial species capable of causing life-threatening illnesses (*Enterococcus faecalis, Mycobacterium tuberculosis* and *Pseudomonas aeruginosa*) already evade every antibiotic in the clinician's armamentarium, a stockpile of more than 100 drugs. In part because of the rise in resistance to antibiotics, the death rates for some communicable diseases (such as tuberculosis) have started to rise again, after having declined in the industrial nations (Levy 1999).

This resistance calls for new drugs with new mechanisms of attack. The majority of drugs we have today have been developed from semi- rational optimization programs based on chemical compounds (Rosamond 2000). Most research today however has begun using the vast amount of information made available with the sequencing of organisms. This new method of attack is referred to as target based. Instead of random compounds found to inhibit growth, differences in metabolism and mechanisms of infection have become the targets for anti- microbial drugs. Ideal anti-microbial drugs are those that will inhibit mechanisms necessary for microbial growth, which are highly conserved and absent or different in humans (Rosamond 2000).

The purine nucleotide biosynthetic pathway has provided a new method of attack against resistant bacteria. This pathway is responsible for the *de novo* synthesis of adenine and guanine nucleotides; nucleoside bases with mono phosphate groups attached (fig 1). Adenine and guanine along with the pyrimidines are the building blocks of DNA and RNA; the genetic material of the cell. Inhibition of this pathway would block the cell's ability to replicate DNA, and transcribe mRNA. Without this ability the cell would not be able to copy its' genetic material and reproduce or make the necessary proteins for metabolism.

This pathway consists of fourteen steps catalyzed by a variety of enzymes (fig. 1). Even though the pathway is invariant, its gene organization and regulation differs among organisms (Zalkin 1992). The enzyme of interest in this study is FGAM synthetase also known as FGAR amidotransferase. This is a key enzyme that catalyses the fourth step of the biosynthetic pathway, which is the conversion of FGAR to FGAM (Schendel 1988) (fig 2). FGAR amidotransferase belongs to a family of enzymes known as glutamine amidotransferases. These enzymes catalyze glutaminase, NH_3 dependent, and glutamine dependent reactions (Zalkin 1992). In organisms such as *E.coli* and humans *purL* codes for the type I FGAR amidotransferase. On the basis of the amino acid sequence of PurL, the enzyme was dissected along its polypeptide chain into at least three discrete regions, designated as domains I, II, and III (Sampei 1989). Domain III (255 amino acids), which resides in the C-terminal region, is similar in amino acid sequence to several glutamine amidotransferases and catalyzes the transfer of the amide nitrogen of glutamine. Domain I (791 amino acids) resides in the N-terminal region and contains a potential ATP binding motif. Domain II (249 amino acids) is structural similar to family of triosephosphate

isomerases and is believed to play a role in the transfer of the carbonyl oxygen of FGAR. These results support a model that the *E. coli purL* gene is a fused gene of at least three different gene families (Sampei 1989). Organisms like *B. subtilis* and *Stapylococcus aureus* have a multi- subunit gene referred to as type II. Three separate genes *purL*, *purY*, and *purQ* comprise the FGAR amidotransferase of these organisms (fig 3). The PurL subunit was assigned to code for the aminator subunit. The PurQ subunit was assigned to the glutamine amide transfer subunit (Sampei 1989). *PurL* and *purQ* are homologous to the single subunit *purL* gene of humans and *E .coli*. The specific function of *purY* has not yet been solved and is not homologous to the *purL* gene found in organisms with the type I enzyme. *PurY* is also more commonly recognized as "purorf" or hypothetical protein. Earlier work in our lab has shown that *purY* mutants are deficient in FGAR amidotransferase (Cook, Ferguson, Newman, unpublished data).

The major goal of our lab is to isolate all three subunits of the multi- subunit enzyme that exists in the purine biosynthetic pathway of many pathogenic bacteria. Once these genes are cloned, expressed, and the proteins purified, biochemical studies can be performed to identify an inhibitor for possible antibiotic use. The antibiotic would target the multi- subunit form of FGAR amidotransferase resulting in the inability of bacteria with this multi- subunit enzyme to synthesize new genetic material. *Staphylococcus aureus* DNA was chosen as a model organism because it is a pathogen known for developing antibiotic resistance and contains the type II, multi subunit enzyme.

This project included the amplification and cloning of the *Staphylococcus aureus purQ* gene. Once amplification, cloning, and expression of the gene have been accomplished, purification of the protein coded for by the *purQ* gene can begin. A

supplementary in depth study, using bioinformatics has also been completed. Genomes of various organisms, including eukaryotes, and many different prokaryotic families were retrieved from databases and the FGAR amidotransferase genes were identified and compared. The resulting phylogenetic tree has provided insight into the evolution of the type I and type II enzymes.

Methods and Materials

Bioinformatics.

Template. The *Staphylococcus aureus* genome has not yet been sequenced completely so the genes had not yet been annotated. Entrez searches for FGAR amidotransferase sequences of other sequenced organisms were run. BLAST searches of unfinished genomes were then done using these sequences and the FGAR amidotransferase genes of *Staphylococcus aureus* were found. These sequences were then taken to the EditSeq component of DNAstar's Lasergene software and examined for an open reading frame corresponding to the expected size of *purQ*. Using MapDraw, anotherLasergene program, the *purQ* sequence was translated to its corresponding protein sequence. A map was also created for the pTyb2 vector (New England BioLabs) showing the restriction sites used in this study and the location of the *purQ* insert. A circular illustration was made in MapDraw and an enzyme filter with NdeI, EcoRI, and EcoRV was created (fig.6).

Gene annotation. Annotation of the FGAR amidotransferase genes was accomplished with MapDraw. The complete sequence was opened in EditSeq and open reading frames of greater than two hundred bases, the approximate size of the smallest subunit, *purY,* were found and a BLAST search was done to determine what protein each open reading

frame encoded. The complete sequence was then opened in MapDraw and features were made corresponding to *purY*, *purL*, and *purQ* and then labeled (fig. 7).

Sequence comparision and Analysis. Protein sequences of organisms from various families were retrieved through GeneQuest by doing an entrez search for *purL* and *purQ*. Since the function of the *purY* gene has not yet been published, it is listed as "purorf" or hypothetical protein. It was found by searching those sequences adjacent to the *purQ* gene for a ORF of about eighty amino acids, which corresponds to the size of PurY. Other sequences, such as those for eukaryotes, were retrieved through genomic research institutions by running a BLAST search using any *purL* sequence found in GeneQuest. The sequences were then saved and Lasergene software was used for further analysis. The nucleic acid sequences were first translated in Editseq; the protein sequences for *purL* and *purQ* of the same organism was then pasted together to facilitate alignment with the type I PurL protein in Megalign. The *purY* sequences were aligned separately because they do not have a *purL* homologue. The Megalign program also gives an alignment report, which shades the residues matching or mismatching the consensus sequence. Megalign also provided a phylogenetic tree based on the similarity of the sequences.

Amplification, Cloning, Expression, and Purification of *purQ* from *Stapylococcus aureus.*

Oligonucleotides. Primers are short oligonucleotides with free 3' OH groups needed for sequence extention. Primers with specific restriction enzyme sites were then designed and ordered through Sigma Genosys. The restriction sites for NdeI or EcoRI and EcoRV were added to the 5" end of the start and stop primer sequences respectively, and have been underlined on the sequences below. Each primer contained at least twenty bases complementary to the beginning and end of the *purQ* sequence.

5' GAGAGGCATATGAAGTTTGCGGTTCTTGTTTTTCCAGG 3' was used as the start primer with Nde1 incorporated into the sequence. A second start primer was also designed using the restriction site for EcoRI,

5' GAGAGGGAATTCATGAAGTTTGCGGTTCTTGTTTTTCCAGG 3'. Start primers attach to the 5' end of the gene.

5' GCACGTGATATCGACATGTTGTTCCCTCCAAC 3' is the stop primer with the EcoRV restriction site incorporated. Stop primers attach to the 3' end of the sequence. Both primers contain minor alterations in their sequences to prevent them from forming secondary structures, or folding up on themselves. Each alteration was made to conserve the resulting amino acid sequence.

Polymerase Chain Reaction (PCR). The primers were then used in a polymerase chain reaction (PCR). PCR amplifies the target sequence, identified by the primers, from bacterial genomic DNA. There are five components used in this reaction. Taq buffer, Taq polymerase enzyme, dNTPs, primers, and water to bring the volume up to fifty microliters. The Taq enzyme reads the sequence being copied and facilitates the addition of bases complementary to the template strand. One unit of Taq enzyme is added to the PCR mixture. dNTPs are added for a final concentration of 0.25 mM. Primers are added at a concentration of 0.75μM, and mineral oil is placed on top of the reaction mixture to prevent evaporation of the solution while in the thermocycler. PCR uses numerous cycles of different temperatures to denature the DNA, anneal the primers, and extend the sequence out from the 3' end using free nucleotides; a 60°C annealing temperature was used because of the primers high melting temperature and their moderate tendency to form secondary structures. A 1 % agarose gel was then run at one hundred twenty volts

to determine if product of the expected size is present; this is seen by examining the gel under ultraviolet light. Bands appear on the gel according to size; these bands are compared to markers, which are also run on the gel.

The PCR product was then purified using the Qiagen™ QIAquick system, which utilizes membrane that binds DNA and allows unwanted material to pass through. After washing, an elution buffer is used to release the DNA from the membrane allowing it to be collected in the filtrate. The pure PCR product is then cut with EcoRV and NdeI or EcoRI restriction enzymes, depending on which primers were used in the PCR reaction. The enzymes are added to 10 X H buffer and PCR product at 37°C for one hour. The host plasmid, pTyb2 was cut using NdeI/ EcoRI and Sma I restriction enzymes in combination with buffer # 4 at room temperature for one hour and then at 37°C for an additional hour. The buffers were chosen by examining enzyme- buffer compatibility tables.

The cut DNA fragments were then purified by phenol-chloroform or gel purification methods and used in the ligation. Phenol-chloroform purification uses an equal volume of phenol to chloroform to extract proteins. This is followed by a $CHCl_3$ extraction to remove residual phenol. The DNA in the resulting aqueous phase was then precipitated with 0.1 volume 3M Sodium acetate and 2.5 volumes 95% Ethanol. The pellet is then rinsed with 70% Ethanol dried in a speed vacuum apparatus and resuspended in water.

Gel purification of DNA is accomplished by running the sample on an agarose gel and then excising and weighing the gel slice containing the band of interest. Three volumes of purification kit binding buffer is then added. Once the gel slice has dissolved,

5µl Prep-A-Gene DNA binding resin (Bio-Rad) is added for each microgram of DNA. The suspension is then centrifuged to produce a pellet. The pellet was rinsed in binding buffer in the equivalent of 25 times the amount of matrix that was added. The pellet was then washed in 25 times matrix volume with wash buffer and resuspended in 1 pellet volume of elution buffer to elute the DNA from the membrane. The suspension is centrifuged and the DNA- containing supernatant is transferred to a clean tube for further use (Bio-Rad Instruction Manual).

The purified DNA, which has been cut with restriction enzymes, is then used in a ligation reaction. These enzymes cut the DNA producing either blunt or "sticky" ends, which are complementary to the cuts on the plasmid allowing insertion of the target sequence into the plasmid, which contains an ampicillin resistance gene. This resistance gene allows for selection of transformed bacteria, only colonies with the plasmid will grow on the AMP treated plates. The plasmid also contains an intein and chitin- binding domain immediately downstream from the inserted target gene, such that expression will produce a fusion protein used in purification. The fusion protein is downstream from a T7 promoter, which is indirectly controlled by a lactose promoter. These components are used in the expression and purification of the protein (fig. 4).

Cloning. In order for the ligated DNA fragments to be isolated, they are transformed into competent cells of bacterial strain TB1. Competent cells have been prepared to be capable of the uptake of foreign DNA. TB1 cultures are incubated and harvested while still in the log phase of growth. These cells are centrifuged and resuspended in .1M $CaCl_2$ and placed on ice. After a 30-minute incubation period on ice they are spun down again and resuspended in 500ul $CaCl_2$. These cells are then ready for transformation.

They are mixed with the ligated plasmid and placed on ice for another incubation after which they are heat shocked at 42°C for two minutes. LB broth is added and the cells are incubated for 1- 2 hours, spread on LB+ AMP plates and incubated overnight. Each colony that grew was "patched" on another LB+ AMP plate and analyzed to determine whether the plasmid contained the desired insert.

Plasmid mini preparations of each colony were done using the CTAB protocol (Del Sal et al 1989). 1.5ml of an overnight bacterial culture is centrifuged and resuspended in 175µl STET buffer. 12.5µl of a 10mg/ml lysozyme solution in TE buffer was added to the resuspended pellet. Boiling for 1 minute was immediately followed by a 10 minute centrifugation. Then 5µl RNase (10mg/ml) is added and heated at 68°C for 10 minutes. A 10% CTAB solution is then added. The mixture is centrifuged and the pellet is resuspended in 10µl of 1.2M NaCl. 95% Ethanol is then added to precipitate the DNA, this can be left overnight for better yield. The pelleted DNA is then rinsed in 70% Ethanol and dried in a centrifugal evaporator. The DNA is then cut with XbaI and HindIII restriction enzymes, which have sites in the host plasmid located on opposite sides of the insert. Each sample was then loaded on an agarose gel for analysis. Sequencing gels were also run to ensure the correct insert was present.

DNA Sequencing. Sequencing reactions require polymerase chain reactions to be run on the sample. The product is purified and distributed into eight tubes, four for each primer. The four tubes represent the four nucleotide bases, adenine, cytosine, guanine, and thymine. Termination mix is added to the appropriate tube, these are mixes containing ddNTPs. ddNTPs are nucleotides lacking the 3' OH group used in sequence extension. Added to these mixtures are dNTPs, Taq enzyme, DNA, buffer, and the appropriate

primer. Mineral oil is layered over the reaction mix before it is placed in a preheated 90°C thermocycler. An unwinding, annealing, extension cycle is then initiated, using 95° C, 42°C, and 70°C temperatures. After fifty- five cycles are completed a stop solution is added to the tubes.

The glass plates have to be thoroughly washed and treated. The long glass plate is treated with a mixture of bind silane, ethanol, and acetic acid. This is what causes the gel to stick to this plate. The integral plate chamber is treated with Rain X™ to coat the plate so the gel will not adhere to this side. Both plates are separated by 0.4mm spacers and clamped together. The gel is 0.4mm thick and is composed of acrylamide, urea, and tris-borate- EDTA (TBE) buffer. This solution was filtered; tetramethylenediamine (temed) and 10% ammonium persulfate (APS) are added to catalyze polymerization of the solution. The solution was quickly poured between two prepared glass plates. The gel is preheated to 55°C and the samples were warmed to 95°C and run on a sequencing gel. The gel was run for 15 minutes after the leading dye runs off the bottom. The plates are then separated and stained using Promega™ silver stain. After staining the gel is photographed and the sequence is ready to be read.

Expression and purification. Once the ligation is transformed into TB1 and the plasmids are isolated and characterized, desired clones are used in the transformation of ER2566 cells. ER2566 is a strain of *E.coli* used in the expression and purification of the target gene. The plasmid must be in ER2566 because it carries the T7 RNA polymerase gene, which is controlled by the lac promoter. Expression of the fusion protein is induced with the addition of IPTG, which activates the lac promoter inducing expression of T7 polymerase, which can then transcribe the fusion construct. Tubes of LB broth are

inoculated with the ER2566 colonies. Fifty milliliter flasks of broth were then inoculated with fresh culture until they reached an absorption at 600nm of 0.5 - 0.8. IPTG was then added for a final concentration of 5mM. Three flasks were set up and incubated at different temperatures to determine optimal expression temperature. One flask was incubated at 30°C for three hours, another at 20- 25°C for six hours, and the other at 12-15°C for sixteen hours. The samples were centrifuged and resuspended in B-Per bacterial protein extraction reagent from Pierce Incorporated. DNase was added at a concentration of 10 mg/ml along with 5mM $MgCl_2$ to breakdown DNA and reduce viscosity. Half of the culture was saved as crude extract; and the other half was centrifuged again and the supernatant saved as clarified extract. The crude extract contains the cell debris and is used to check for insoluble proteins. The clarified extract has the cell debris removed; if expression is found in the clarified extract the protein is in a soluble form.

Sample preparation involved mixing two parts undiluted or 5- fold diluted extract to one part SDS buffer and heated for five to ten minutes at 95°C, centrifuged, and loaded on two 4- 20% SDS protein Ready gels from Bio- Rad. Two SDS gels were run in the same manner and run at 150 volts for twenty minutes after the leading dye ran off the gel. One of the gels was used for coomassie blue staining and the other was used for a western blot. For the western blot, proteins were transferred to a PVDF membrane as follows, the gel was sandwiched in a cassette between filter pads, filter paper, and a PVDF membrane. The cassette is then placed in the chamber with a cooling unit and stir bar and run overnight at 35 volts at 4°C. The membrane is then incubated with 0.1ml per square centimeter of filter with a blocking solution of 5% nonfat dried milk in PBS (phosphate buffered saline) to prevent non- specific binding of antibodies. A primary

antibody (rabbit anti- chitin- binding domain) against the chitin-binding domain was then added and incubated overnight. After the incubation period the filter was washed three times with PBS. The filter was rinsed with a 150mM NaCl, 50mM Tris-HCL (pH 7.5) solution for ten minutes. The filter is then placed in a seal-a-meal bag with 0.1ml per cm^2 of phosphate free, azide free blocking solution composed of 5% nonfat dry milk in 150mM NaCl and 50mM Tris- HCL (pH 7.5). The secondary antibody against the first has a phosphate attached and was added for at a dilution of 1:1000. After incubation the filter is placed in a 150mM NaCl, 50mM Tris-HCL (pH 7.5) solution and washed three times. For detection of antibodies the filter is placed in an alkaline phosphatase buffer of 100mM NaCl, 5mM $MgCl_2$, and 100mM Tris- HCL (pH 9.5) with 66µl NBT and 33µl BCIP. The substrate NBT/BCIP is converted into a blue precipitate by immuno-localized alkaline phosphatse. The filter is then placed in a stop solution of 200µl 0.5M EDTA (pH 8.0) in 50ml PBS.

Results and Discussion

Bioinformatics.

The *Staphylococcus aureus* sequence was retrieved from The Institute for Genomic Research (TIGR). The attached printout has the protein translation under the sequence with the primer sites labeled and underlined. The *purQ* sequence is 672bp long. The NdeI restriction site found to cut the insert in the middle is also labeled (fig.5). A map of the pTYB2 vector showing restriction sites where the insert was ligated (fig. 6).

The FGAR amidotransferase subunit genes are ordered with *purY* first then *purQ* and *purL*. The genes appear to be arranged in an operon. Shine- Delgarno sequences can be found in front of each gene sequence. Shine- Delgarno sequences are ribosomal

binding sites (labeled RBS in fig.5). *PurY* is 263 bases or 88 amino acids in length and corresponds to bases 21- 284 of the complete sequence. *PurQ* is 671 bases or 224 amino acids in length and corresponds to bases 286- 957. *PurL* is 2189 bases or 730 amino acids in length and corresponds to bases 950- 3139. *PurQ* and *purL* overlap each other by seven bases and are in separate reading frames. A printout of the annotation is attached (fig.7).

Sequence comparison and analysis. Sequences for both prokaryotes and eukaryotes were found. Single subunit and multi- subunit proteins were used in this analysis. The alignment of *purQ* and *purL* is attached (fig.8a) and does show areas of moderate to strong conservation in the protein coded for by *purL*. The resulting tree (fig. 8a) clearly shows the divergence of the multi- subunit and the single subunit types. All organisms on the branch with *E. coli* have the single subunit form. The organisms starting with *Thermatoga martima* have the multi- subunit form, including *Staphylococcus aureus*. A split of the proteobacteria can also be observed. *E. coli, Salmonella typhimurium*, *Neisseria meningitidis*, and *Haemophilus influenzae* belong to either the gamma or beta proteobacteria families and are consistently single subunit types whereas the delta and epsilon proteobacteria, such as *Camplobacter jejuni* have the multi- subunit type. The alignment report for *purY* (fig. 8b) show areas of strong conservancy. Residues LDP correspond to bases 17- 19 on the consensus sequence and are strongly conserved along with K43 and N71. The tree resulting (fig. 8b) from *purY* alignment predicts different evolutionary patterns then the *purL* and *purQ* trees. Eukaryotes such as *Drosophila melangaster,* and *Saccharomyces cerevisiae* and members of the gamma and beta proteobacteria have the single subunit enzyme and therefore do not have genes for *purQ*

and *purY*. All other families of bacteria do have the *purY* gene. This tells us that the single subunit form of FGAR amidotransferase evolved in the proteobacteria.

Amplification, Cloning, Expression, and Purification of *purQ* from *Staphylococcus aureus.*

Analysis of the PCR amplification of the Sa*purQ* gene shows a strong band slightly smaller than 700bp (fig. 9), which corresponds to the size of *purQ*. During this first attempt at cloning one of the minipreps from patch number 53 was found to have the correct insert with a band at about 1100 bp, corresponding to the size of the target sequence when cut with XbaI and HindIII. The gel picture for this prep was omitted because the band was too light and didn't show up well when scanned. A sequencing reaction was also run on PCR product derived from # 53 to ensure that the target gene was present. The gel picture has been attached (fig. 10). The portion corresponding to the *purQ* insert was read AAACGTCTTCAGC and is bolded on the attached sequence. Number 53 DNA was then isolated and used in a transformation into ER2566. Two ER2566 colonies grew and cultures of both were started in preparation for gene expression.

Expression and purification. The SDS gel did not show a band at 80KD, which would have corresponded to the size of the *purQ* protein (25KD) combined with the intein and chitin binding domain (55KD). Due to this result another miniprep of number 53 was done, and the gels showed no band at 1100 bp. The insert was present which the gel evidenced, however after multiple minipreps and different methods showing no bands number 53 was abandoned.

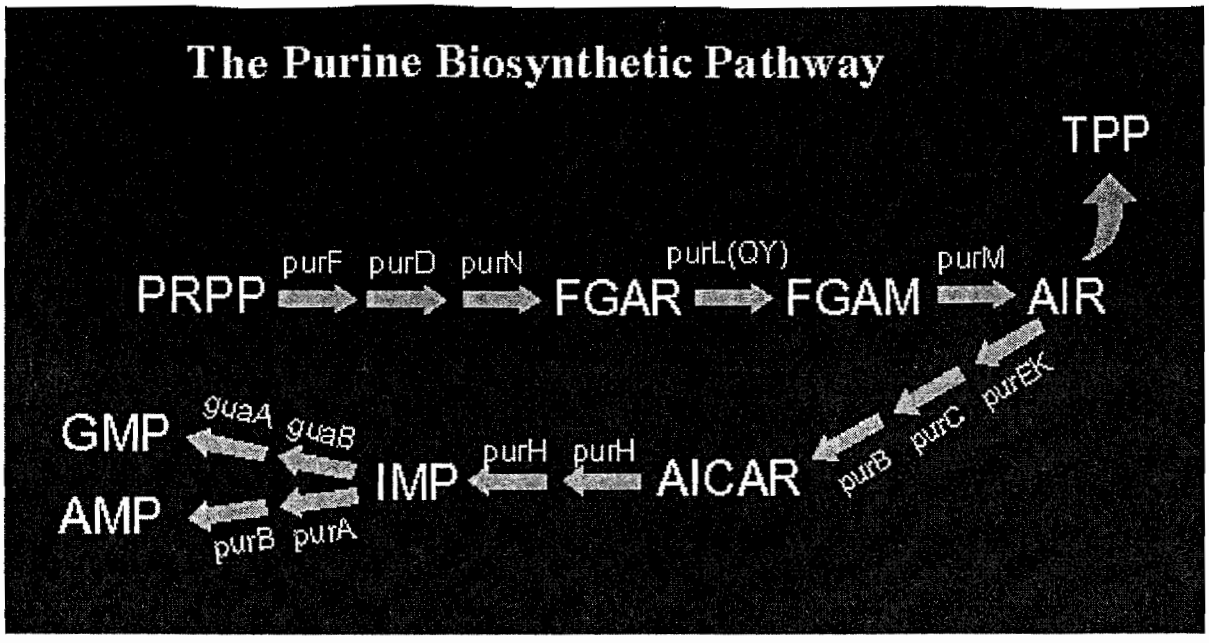
Cloning. More transformations were done and other miniprep methods were used in hopes of obtaining bands of greater intensity. After analysis of one gel, it was discovered that one of the enzymes was cutting the insert more then once (fig. 11). This was seen by the appearance of two bands close in size on the gel. Another sequencing reaction was done on the PCR of Sa*purQ* to determine what enzyme was cutting and where (fig. 12). It was determined the NdeI restriction enzyme cut the insert twice instead of once. This was supported by analysis of the sequence retrieved from the data banks. MapDraw showed that the NdeI enzyme being used cut the insert twice once at the restriction site created by the primer and once close to the end of our target. This was the reason for early cloning difficulties. A new primer using an EcoRI restriction site instead of NdeI was then designed. This primer was used for subsequent PCR reactions and the EcoRI enzyme was used in place of the NdeI enzyme. The first and second transformations did not work; there was no growth on either plate. More ligations were set up using varying amounts of PCR product in relation to the vector. Twenty- three colonies grew on plate one and sixty plus grew on plate two. Uncut plasmid preparations from these colonies were first run on gels; those that migrated slower were more likely to have the insert present and these were then tested further by cutting them with XbaI and HindIII. Colonies 1, 21, 23, and 27 migrated slower then other preps and were cut after a phenol-chloroform extraction, which showed no cutting probably due to the excess of RNA present (fig. 13). The preps were treated with RNase and recut to clean the samples and eliminate RNA contamination (fig. 14). The resulting gel shows two bands about 1100bp apart in size. The top band is about 8100bp this corresponds to the size of an uncut plasmid containing the insert. The bottom band is about 7000bp corresponding to the

size of a cut, insert containing plasmid. The band for the insert should appear at 1100bp, but is not visible this is because of its small size and therefore its low concentration. Additionally PCR reactions were run on all four colonies with the intien reverse and pTyb2, T7 primers, which correspond to vector sequences. Plasmids with no insert produced bands with a size of 450bp; those with the insert produced bands at 1100bp, both band sizes are present on the gel (fig. 15). Instead of using the CTAB miniprep method the Wizard™ miniprep sytem was used because it was found to produce higher concentrations of DNA. The samples were then cut and analyzed by gel electrophoresis (fig. 16). The bands closer to the top of the gel represent the plasmid with the insert removed. The lower, faster migrating bands represent the insert at about 1100bp. The gel showed a band at 1100bp for each sample, however the colony 21 and 23 bands are very light and difficult to see. Four transformations into ER2566 were set up using the purified plasmids extracted from the TB1 cultures. Each plate grew numerous colonies, which is indicative of a successful transformation of the plasmid.

Expression. Colony one was chosen for expression studies because it produced one of the strongest bands at 1100bp. The coomassie blue stained gel showed bands throughout the gel (fig. 17). Bovine serum albumin was run as a marker and has a band at 66kD. The fusion protein which contains the *purQ* protein plus the intein and chitin- binding domains is 80kD in size. Gel analysis shows expression of the protein at all three temperatures. The western blot produced bands for the crude samples, but not the clarified. This is indicative of non- soluble proteins, which are removed along with the cell debris for the clarified extract. Also the greatest amount of expression occurred in the 12- 15°C cultures.

Conclusion.

Amplification and cloning of the *purQ* insert into ER2566 has been accomplished. Expression of the protein has been determined by the western blot. Solubilization of the PurQ protein has to be accomplished for a succesful purification and better yeild of PurQ. The evolutionary analysis shows that the type I and type II enzyme evolved with the proteobacteria. The analysis shows a split placing the gamma and beta proteobacteria in the type I family along with eukaryotes and the epsilon, alpha, and delta proteobacteria with all other families of prokaryotes in the type II family.



ure 1. *The Purine Biosynthetic Pathway.* Genes encoding the enzyme that catalyzes each step is wn above or below each arrow.

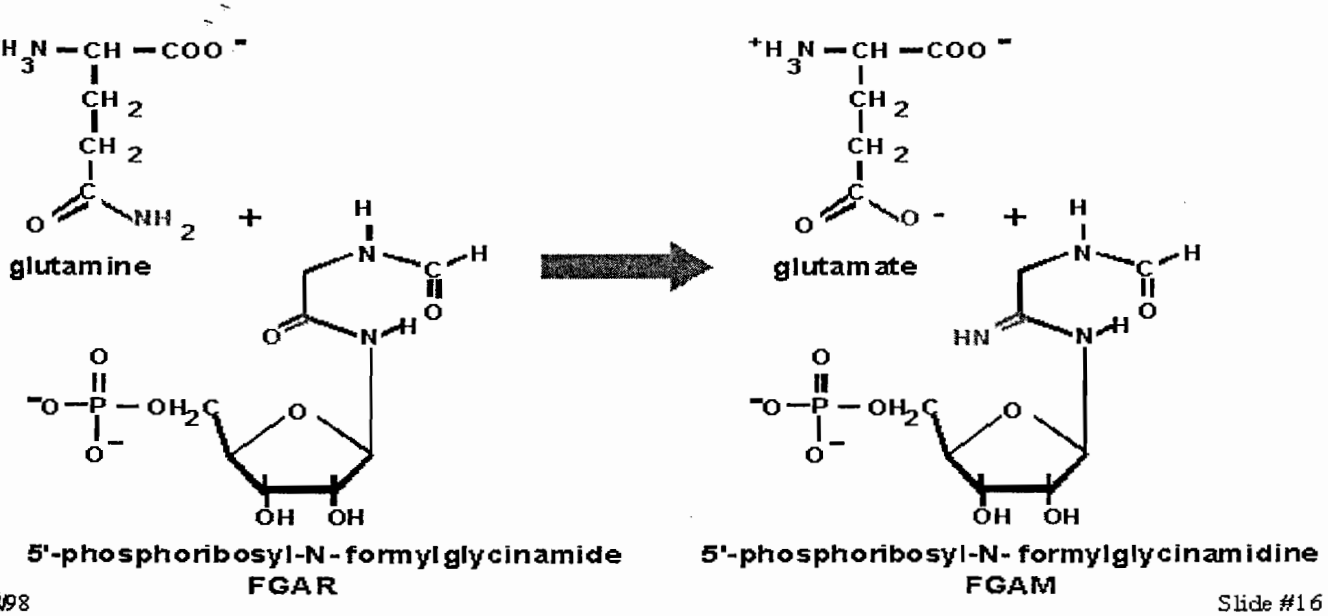


Figure 2. *FGAR Amidotransferase catalyzed reaction.* This figure shows the reaction catalyzed by FGAR amidotransferase. In this reaction an amino group is transferred from glutamine to FGAR producing FGAM and glutamate.

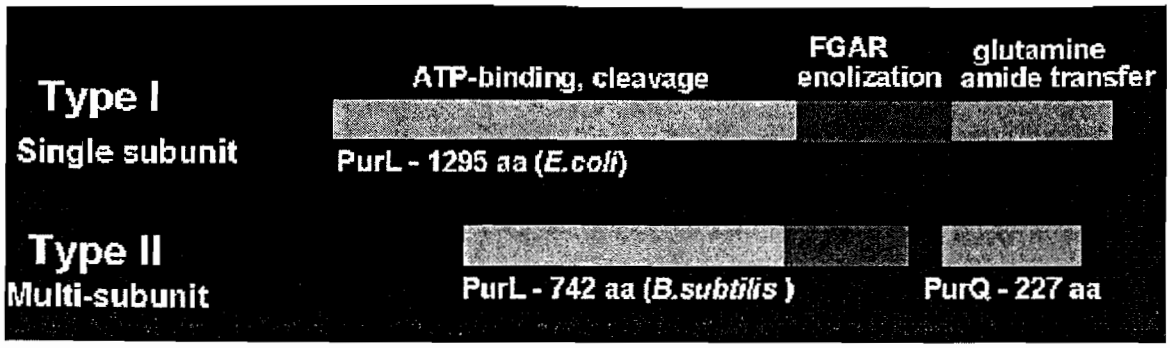


Figure 3. *Gene types.* Type I is the single subunit found in eukaryotes and some prokaryotes. Type II is the muti subunit enzyme found in many pathogenic bacteria and is composed of *purL*, *purY*, and *purQ*.

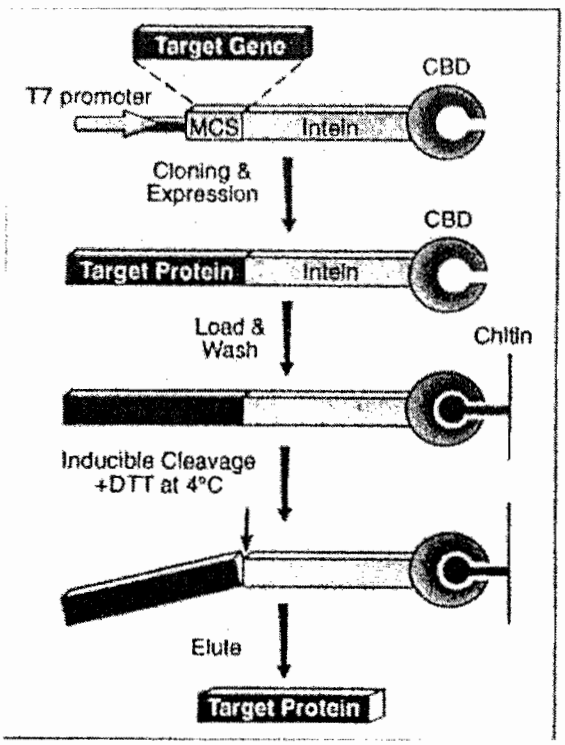
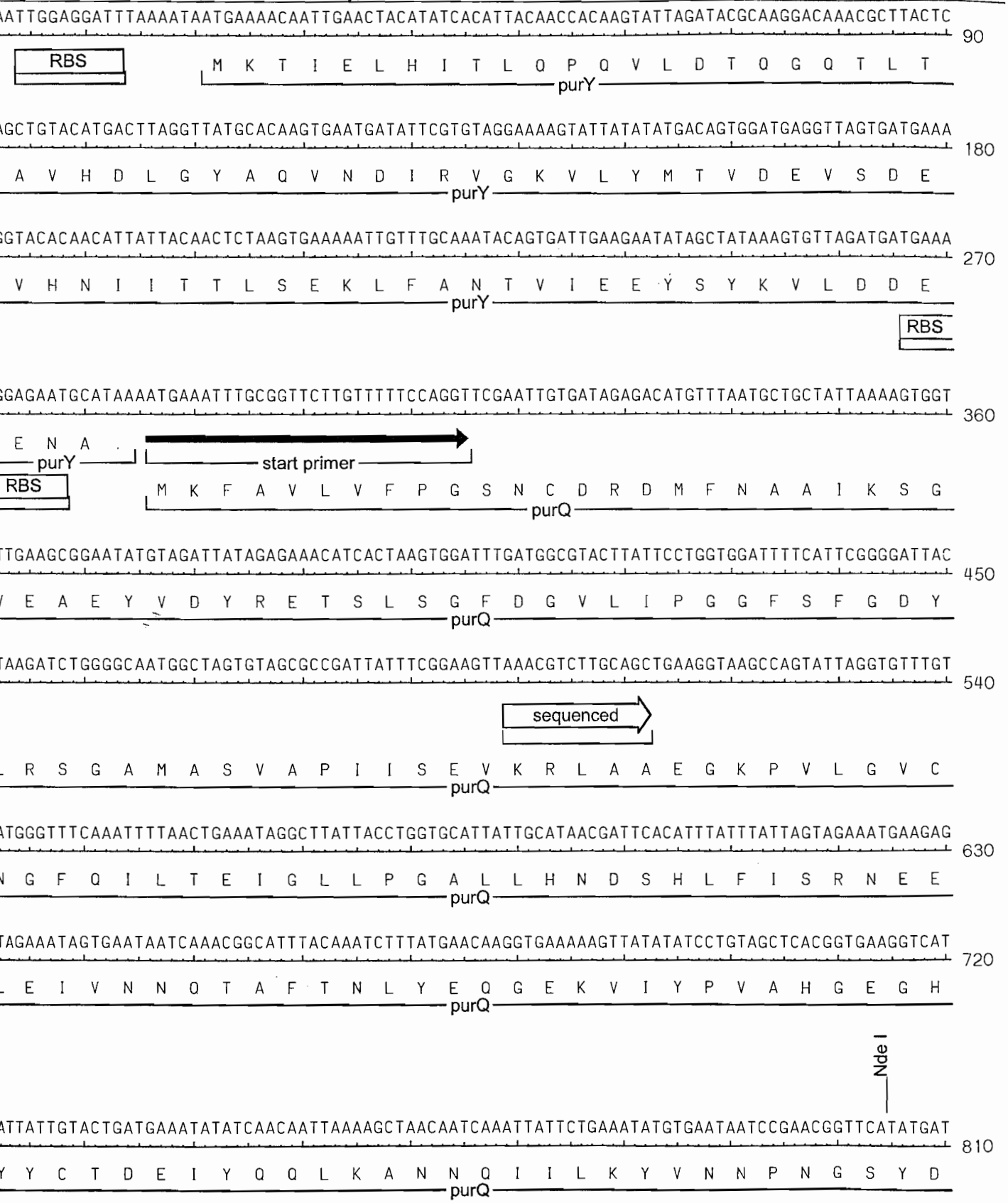
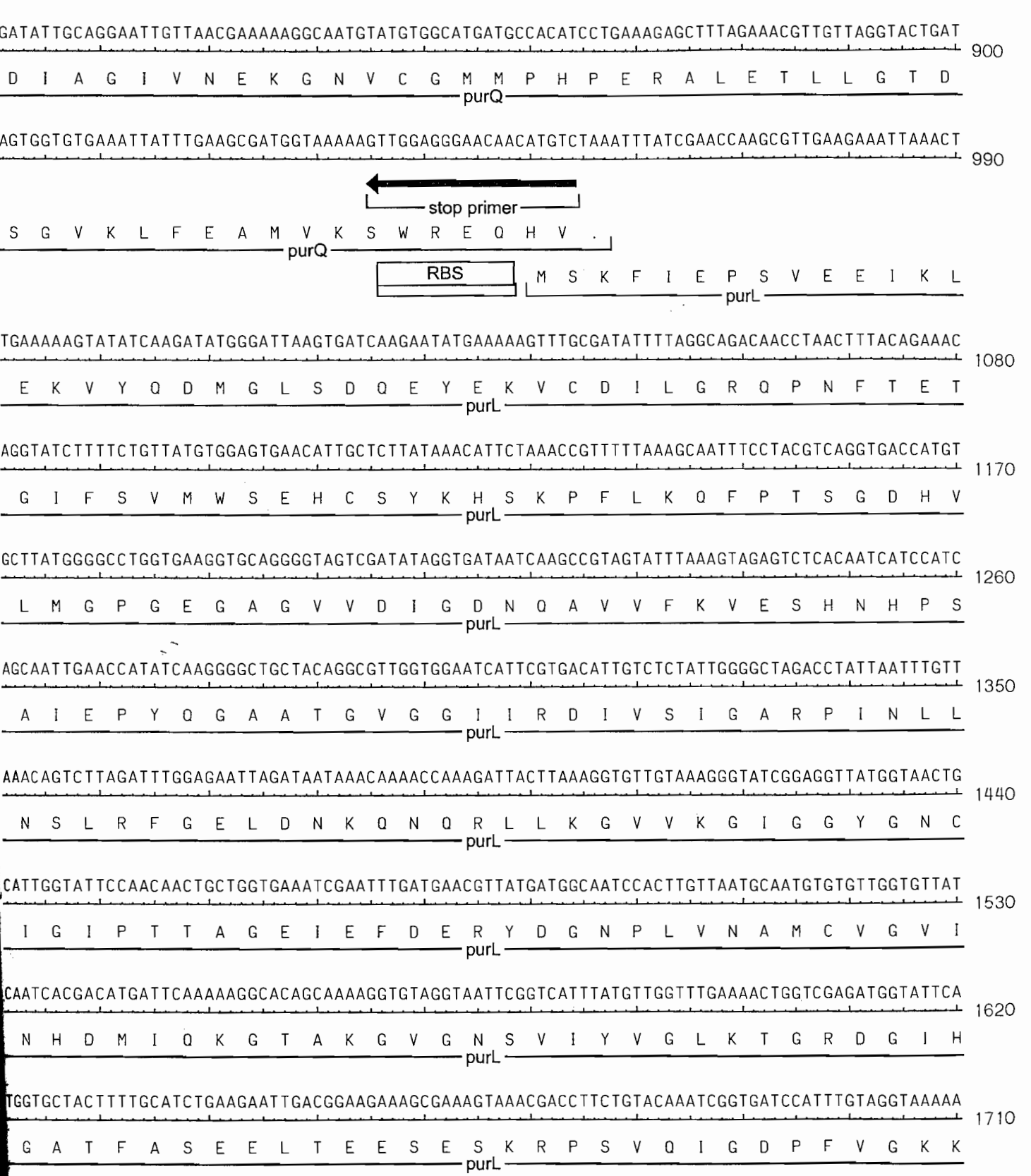
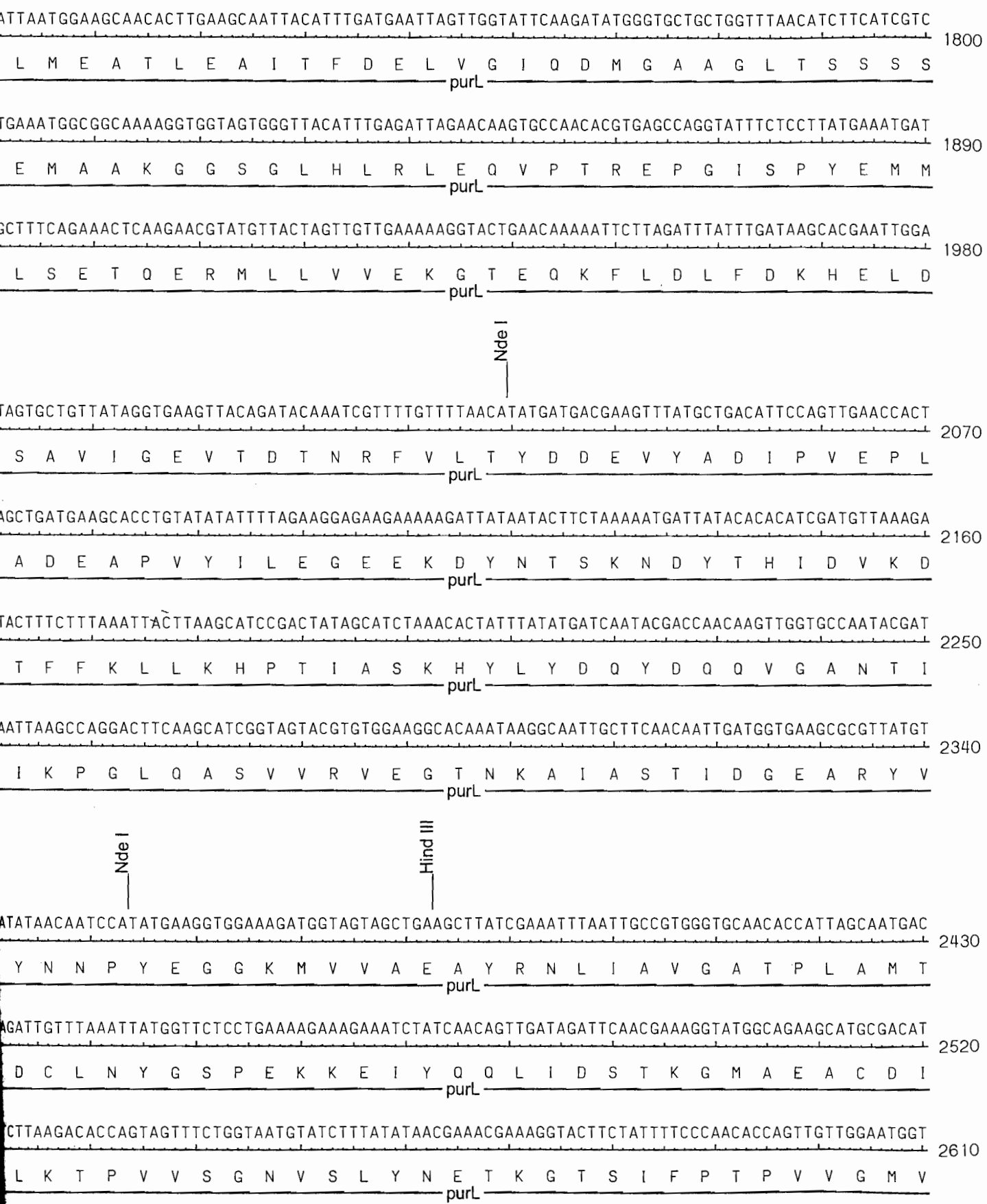


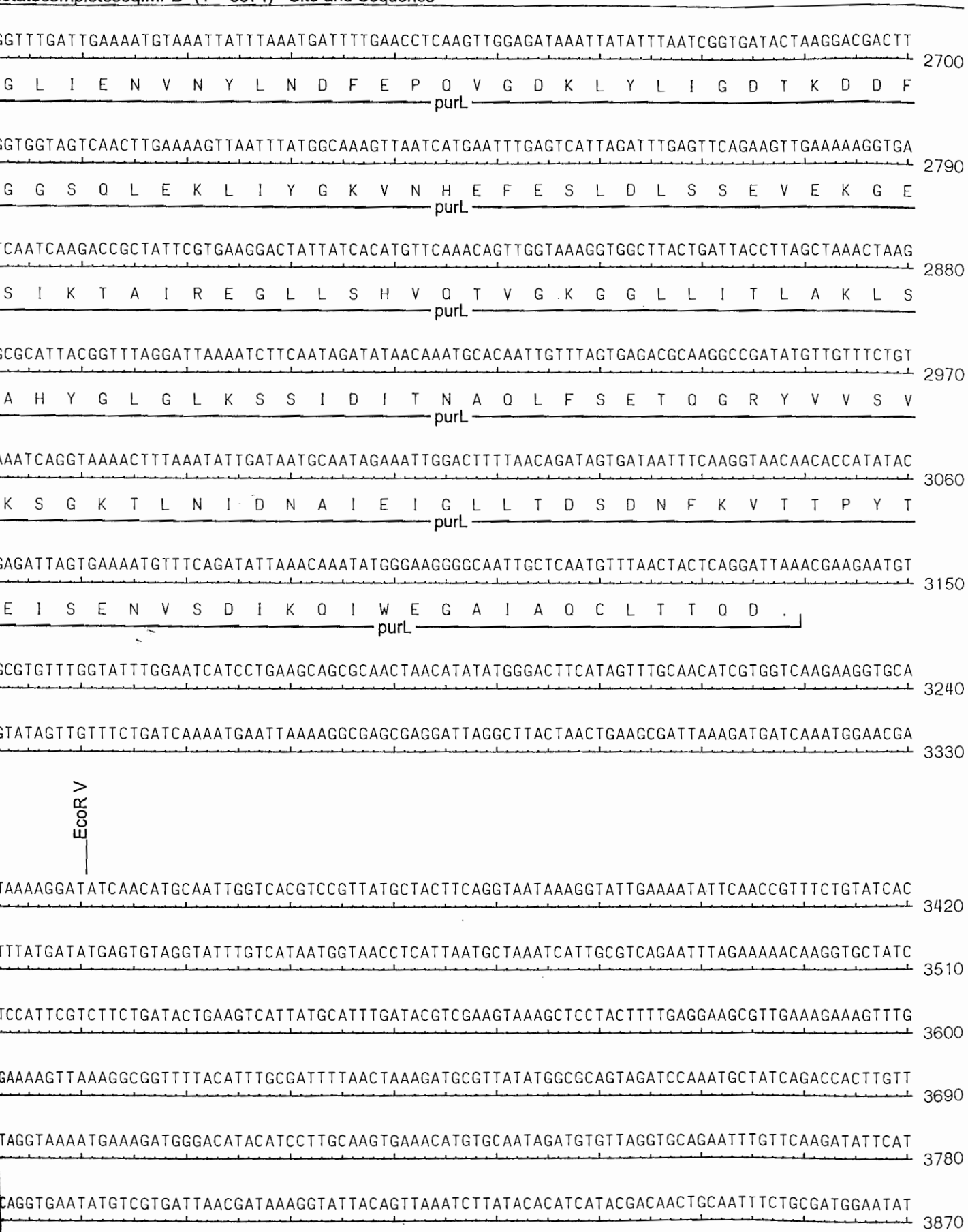
Figure 4. *IMPACT™ T7.* This figure shows the fusion protein created with the addition of the target insert.

Figure 5. *SapurQ sequence.* The genes for *purL, purQ*, and *purY* are shown. The sequence corresponding to *purQ* is underlined and labeled along with the primer sites and the bases corresponding to those read off of gel figure 8. The NdeI cut site is also shown.


AATTGGAGGATTTAAAATAATGAAAACAATTGAACTACATATCACATTACAACCACAAGTATTAGATACGCAAGGACAAACGCTTACTC
90
RBS
M K T I E L H I T L Q P Q V L D T Q G Q T L T
purY
AGCTGTACATGACTTAGGTTATGCACAAGTGAATGATATTCGTGTAGGAAAAGTATTATATATGACAGTGGATGAGGTTAGTGATGAAA
180
A V H D L G Y A Q V N D I R V G K V L Y M T V D E V S D E
purY
GGTACACAACATTATTACAACTCTAAGTGAAAAATTGTTTGCAAATACAGTGATTGAAGAATATAGCTATAAAGTGTTAGATGATGAAA
270
V H N I I T T L S E K L F A N T V I E E Y S Y K V L D D E
purY
RBS
GGAGAATGCATAAAATGAAATTTGCGGTTCTTGTTTTTCCAGGTTCGAATTGTGATAGAGACATGTTTAATGCTGCTATTAAAAGTGGT
360
E N A .
purY
start primer
RBS
M K F A V L V F P G S N C D R D M F N A A I K S G
purQ
TTGAAGCGGAATATGTAGATTATAGAGAAACATCACTAAGTGGATTTGATGGCGTACTTATTCCTGGTGGATTTTCATTCGGGGATTAC
450
E A E Y V D Y R E T S L S G F D G V L I P G G F S F G D Y
purQ
TAAGATCTGGGGCAATGGCTAGTGTAGCGCCGATTATTTCGGAAGTTAAACGTCTTGCAGCTGAAGGTAAGCCAGTATTAGGTGTTTGT
540
sequenced
R S G A M A S V A P I I S E V K R L A A E G K P V L G V C
purQ
ATGGGTTTCAAATTTTAACTGAAATAGGCTTATTACCTGGTGCATTATTGCATAACGATTCACATTTATTTATTAGTAGAAATGAAGAG
630
N G F Q I L T E I G L L P G A L L H N D S H L F I S R N E E
purQ
TAGAAATAGTGAATAATCAAACGGCATTTACAAATCTTTATGAACAAGGTGAAAAAGTTATATATCCTGTAGCTCACGGTGAAGGTCAT
720
E I V N N Q T A F T N L Y E Q G E K V I Y P V A H G E G H
purQ
Nde I
ATTATTGTACTGATGAAATATATCAACAATTAAAAGCTAACAATCAAATTATTCTGAAATATGTGAATAATCCGAACGGTTCATATGAT
810
Y Y C T D E I Y Q Q L K A N N Q I I L K Y V N N P N G S Y D
purQ


GATATTGCAGGAATTGTTAACGAAAAAGGCAATGTATGTGGCATGATGCCACATCCTGAAAGAGCTTTAGAAACGTTGTTAGGTACTGAT
900
D I A G I V N E K G N V C G M M P H P E R A L E T L L G T D
purQ
AGTGGTGTGAAATTATTTGAAGCGATGGTAAAAAGTTGGAGGGAACAACATGTCTAAATTTATCGAACCAAGCGTTGAAGAAATTAAACT
990
stop primer
S G V K L F E A M V K S W R E Q H V .
purQ
RBS
M S K F I E P S V E E I K L
purL
TGAAAAAGTATATCAAGATATGGGATTAAGTGATCAAGAATATGAAAAAGTTTGCGATATTTTAGGCAGACAACCTAACTTTACAGAAAC
1080
E K V Y Q D M G L S D Q E Y E K V C D I L G R Q P N F T E T
purL
AGGTATCTTTTCTGTTATGTGGAGTGAACATTGCTCTTATAAACATTCTAAACCGTTTTTAAAGCAATTTCCTACGTCAGGTGACCATGT
1170
G I F S V M W S E H C S Y K H S K P F L K Q F P T S G D H V
purL
GCTTATGGGGCCTGGTGAAGGTGCAGGGGTAGTCGATATAGGTGATAATCAAGCCGTAGTATTTAAAGTAGAGTCTCACAATCATCCATC
1260
L M G P G E G A G V V D I G D N Q A V V F K V E S H N H P S
purL
AGCAATTGAACCATATCAAGGGGCTGCTACAGGCGTTGGTGGAATCATTCGTGACATTGTCTCTATTGGGGCTAGACCTATTAATTTGTT
1350
A I E P Y Q G A A T G V G G I I R D I V S I G A R P I N L L
purL
AAACAGTCTTAGATTTGGAGAATTAGATAATAAACAAAACCAAAGATTACTTAAAGGTGTTGTAAAGGGTATCGGAGGTTATGGTAACTG
1440
N S L R F G E L D N K Q N Q R L L K G V V K G I G G Y G N C
purL
CATTGGTATTCCAACAACTGCTGGTGAAATCGAATTTGATGAACGTTATGATGGCAATCCACTTGTTAATGCAATGTGTGTTGGTGTTAT
1530
I G I P T T A G E I E F D E R Y D G N P L V N A M C V G V I
purL
CAATCACGACATGATTCAAAAAGGCACAGCAAAAGGTGTAGGTAATTCGGTCATTTATGTTGGTTTGAAAACTGGTCGAGATGGTATTCA
1620
N H D M I Q K G T A K G V G N S V I Y V G L K T G R D G I H
purL
TGGTGCTACTTTTGCATCTGAAGAATTGACGGAAGAAAGCGAAAGTAAACGACCTTCTGTACAAATCGGTGATCCATTTGTAGGTAAAAA
1710
G A T F A S E E L T E E S E S K R P S V Q I G D P F V G K K
purL


ATTAATGGAAGCAACACTTGAAGCAATTACATTTGATGAATTAGTTGGTATTCAAGATATGGGTGCTGCTGGTTTAACATCTTCATCGTC
1800
L M E A T L E A I T F D E L V G I Q D M G A A G L T S S S S
purL
TGAAATGGCGGCAAAAGGTGGTAGTGGGTTACATTTGAGATTAGAACAAGTGCCAACACGTGAGCCAGGTATTTCTCCTTATGAAATGAT
1890
E M A A K G G S G L H L R L E Q V P T R E P G I S P Y E M M
purL
GCTTTCAGAAACTCAAGAACGTATGTTACTAGTTGTTGAAAAAGGTACTGAACAAAAATTCTTAGATTTATTTGATAAGCACGAATTGGA
1980
L S E T Q E R M L L V V E K G T E Q K F L D L F D K H E L D
purL
Nde I
TAGTGCTGTTATAGGTGAAGTTACAGATACAAATCGTTTTGTTTTAACATATGATGACGAAGTTTATGCTGACATTCCAGTTGAACCACT
2070
S A V I G E V T D T N R F V L T Y D D E V Y A D I P V E P L
purL
AGCTGATGAAGCACCTGTATATATTTTAGAAGGAGAAGAAAAAGATTATAATACTTCTAAAAATGATTATACACACATCGATGTTAAAGA
2160
A D E A P V Y I L E G E E K D Y N T S K N D Y T H I D V K D
purL
TACTTTCTTTAAATTACTTAAGCATCCGACTATAGCATCTAAACACTATTTATATGATCAATACGACCAACAAGTTGGTGCCAATACGAT
2250
T F F K L L K H P T I A S K H Y L Y D Q Y D Q Q V G A N T I
purL
AATTAAGCCAGGACTTCAAGCATCGGTAGTACGTGTGGAAGGCACAAATAAGGCAATTGCTTCAACAATTGATGGTGAAGCGCGTTATGT
2340
I K P G L Q A S V V R V E G T N K A I A S T I D G E A R Y V
purL
Nde I
Hind III
ATATAACAATCCATATGAAGGTGGAAAGATGGTAGTAGCTGAAGCTTATCGAAATTTAATTGCCGTGGGTGCAACACCATTAGCAATGAC
2430
Y N N P Y E G G K M V V A E A Y R N L I A V G A T P L A M T
purL
AGATTGTTTAAATTATGGTTCTCCTGAAAAGAAAGAAATCTATCAACAGTTGATAGATTCAACGAAAGGTATGGCAGAAGCATGCGACAT
2520
D C L N Y G S P E K K E I Y Q Q L I D S T K G M A E A C D I
purL
CTTAAGACACCAGTAGTTTCTGGTAATGTATCTTTATATAACGAAACGAAAGGTACTTCTATTTTCCCAACACCAGTTGTTGGAATGGT
2610
L K T P V V S G N V S L Y N E T K G T S I F P T P V V G M V
purL

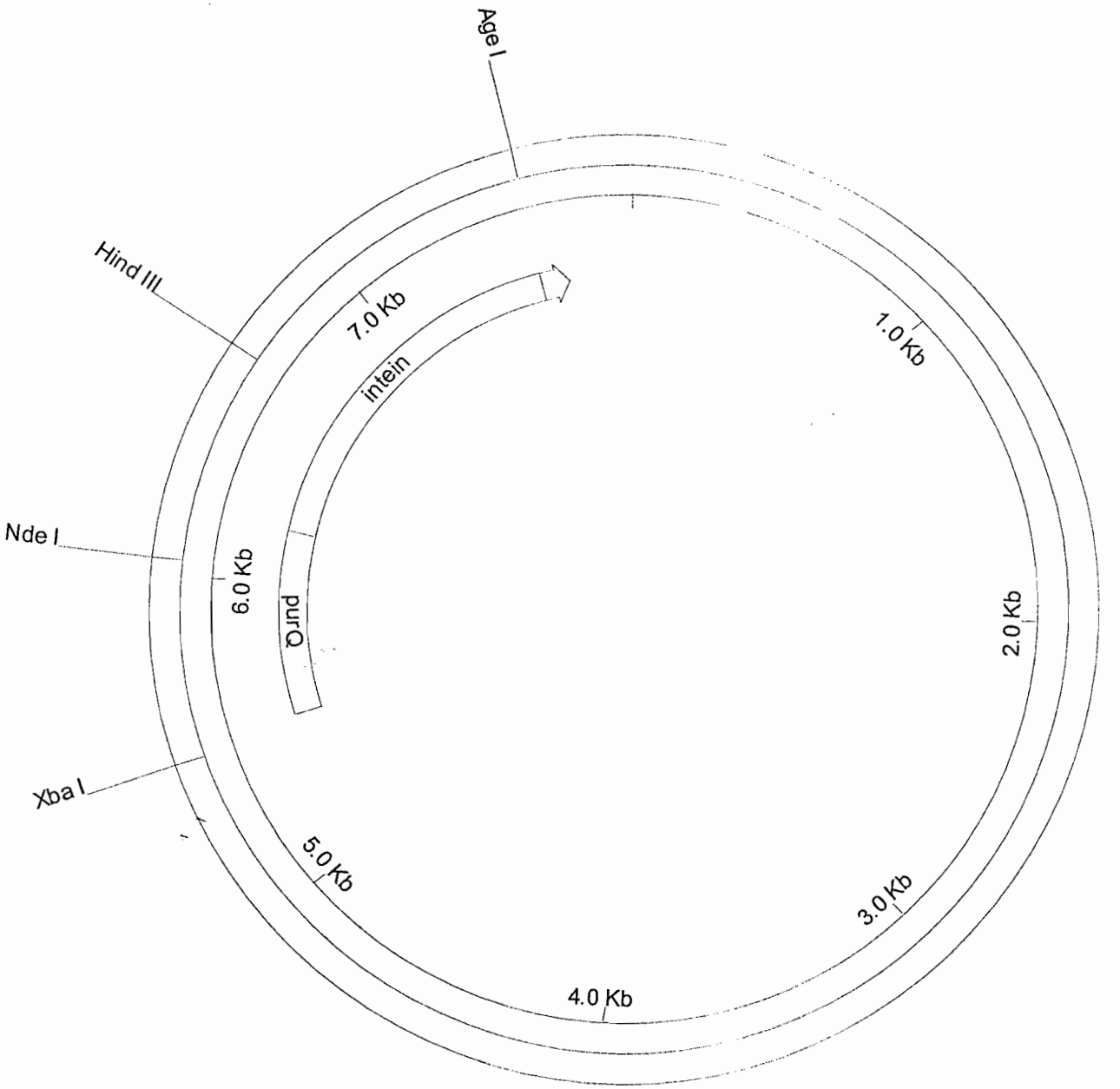

GGTTTGATTGAAAATGTAAATTATTTAAATGATTTTGAACCTCAAGTTGGAGATAAATTATATTTAATCGGTGATACTAAGGACGACTT
2700
G L I E N V N Y L N D F E P Q V G D K L Y L I G D T K D D F
purL
GGTGGTAGTCAACTTGAAAAGTTAATTTATGGCAAAGTTAATCATGAATTTGAGTCATTAGATTTGAGTTCAGAAGTTGAAAAAGGTGA
2790
G G S Q L E K L I Y G K V N H E F E S L D L S S E V E K G E
purL
TCAATCAAGACCGCTATTCGTGAAGGACTATTATCACATGTTCAAACAGTTGGTAAAGGTGGCTTACTGATTACCTTAGCTAAACTAAG
2880
S I K T A I R E G L L S H V Q T V G K G G L L I T L A K L S
purL
GCGCATTACGGTTTAGGATTAAAATCTTCAATAGATATAACAAATGCACAATTGTTTAGTGAGACGCAAGGCCGATATGTTGTTTCTGT
2970
A H Y G L G L K S S I D I T N A Q L F S E T Q G R Y V V S V
purL
AAATCAGGTAAAACTTTAAATATTGATAATGCAATAGAAATTGGACTTTTAACAGATAGTGATAATTTCAAGGTAACAACACCATATAC
3060
K S G K T L N I D N A I E I G L L T D S D N F K V T T P Y T
purL
GAGATTAGTGAAAATGTTTCAGATATTAAACAAATATGGGAAGGGGCAATTGCTCAATGTTTAACTACTCAGGATTAAACGAAGAATGT
3150
E I S E N V S D I K Q I W E G A I A Q C L T T Q D .
purL
GCGTGTTTGGTATTTGGAATCATCCTGAAGCAGCGCAACTAACATATATGGGACTTCATAGTTTGCAACATCGTGGTCAAGAAGGTGCA
3240
GTATAGTTGTTTCTGATCAAAATGAATTAAAAGGCGAGCGAGGATTAGGCTTACTAACTGAAGCGATTAAAGATGATCAAATGGAACGA
3330
EcoR V
TAAAAGGATATCAACATGCAATTGGTCACGTCCGTTATGCTACTTCAGGTAATAAAGGTATTGAAAATATTCAACCGTTTCTGTATCAC
3420
TTTATGATATGAGTGTAGGTATTTGTCATAATGGTAACCTCATTAATGCTAAATCATTGCGTCAGAATTTAGAAAAACAAGGTGCTATC
3510
TCCATTCGTCTTCTGATACTGAAGTCATTATGCATTTGATACGTCGAAGTAAAGCTCCTACTTTTGAGGAAGCGTTGAAAGAAAGTTTG
3600
GAAAAGTTAAAGGCGGTTTTACATTTGCGATTTTAACTAAAGATGCGTTATATGGCGCAGTAGATCCAAATGCTATCAGACCACTTGTT
3690
TAGGTAAAATGAAAGATGGGACATACATCCTTGCAAGTGAAACATGTGCAATAGATGTGTTAGGTGCAGAATTTGTTCAAGATATTCAT
3780
CAGGTGAATATGTCGTGATTAACGATAAAGGTATTACAGTTAAATCTTATACACATCATACGACAACTGCAATTTCTGCGATGGAATAT
3870

Hind III

TTTATTTTGCTAGACCAGACTCAACAATAGCTGGTAAAAATGTCCATGCAGTACGTAAAGCTTCTGGTAAAAAATTAGCCCAAGAAAGC
3960

CTGTAAATGC
3971

Figure 6. *pTYB2 vector.* The circular map of the pTYB2 vector shows the size of the vector and also the restriction sites used in this experiment. The *purQ* insert is inserted between these sites and the intein and chitin- binding domain are also shown.


Age I
Hind III
Nde I
Xba I
7.0 Kb
intein
purQ
1.0 Kb
2.0 Kb
3.0 Kb
4.0 Kb
5.0 Kb
6.0 Kb


ACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCAT
90
CAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTT
180
GCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTA
270
GAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTCTCCAATGATGAGCACTTTTAAAGTTCTGCT
360
GGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTC
450
GTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAA
540
CTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGA
630
GAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGG
720
CTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCC
810
GGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTC
900
ATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAA
990
TGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTACCCCGGTTGATAATCAGAAAAGCCCCAAAAACAGGAAG
1080
TATAAGCAAATATTTAAATTGTAAACGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAA
1170
CCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTA
1260
AGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAATCAAGTTTT
1350
GGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCG
1440
AGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCG
1530
ATGCGCCGCTACAGGGCGCGTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTC
1620
CACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAA
1710

CCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGAT 1800

AATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCT 1890

CCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGG 1980

ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGC 2070

CTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAA 2160

TGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATG 2250

AACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTC 2340

GATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC 2430

GTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGC 2520

CGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGG 2610

TGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGCAGCGATTC 2700

ATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAG 2790

GTTTTTTCCTGTTTGGTCACTTGATGCCTCCGTGTAAGGGGGAATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGG 2880

TCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGAC 2970

GAAAAATCACTCAGGGTCAATGCCAGCCGAACGCCAGCAAGACGTAGCCCAGCGCGTCGGCCGCCATGCCGGCGATAATGGCCTGC 3060

CGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATC 3150

TCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACA 3240

TAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCC 3330

GCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAA 3420

TCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATT 3510

TTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAA 3600

GGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGC 3690

ATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAA 3780

GACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAG 3870

GCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACC 3960

TCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCAC 4050

ATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGC 4140

ACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGG 4230

TGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATT 4320

TCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGA 4410

CCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCAT 4500

CGAAAGGTTTTGCGCCATTCGATGGTGTCCCGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTA 4590

GAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCCGCCCTTTCGTCTTCAAGAATTAATTCCCAATTCCAGGCATCAAAT 4680

CGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGG 4770

GATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTAATTCCAGGCATCAAA 4860

ACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGG 4950

GGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTAATTCCAGGCATCAA 5040

AACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCG 5130

CGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTAATTCCAGGCATCA 5220

AAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCC 5310

GCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGAATTGGGGATCGGAATT 5400

Xba I

CCCGGTTTAAACCGGGGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAG 5490

AATTTTGTTTAACTTTAAGAAGGAGATATAATGAAATTTGCGGTTCTTGTTTTTCCAGGTTCGAATTGTGATAGAGACATGTTTAA 5580

purQ

GCTATTAAAAGTGGTGTTGAAGCGGAATATGTAGATTATAGAGAAACATCACTAAGTGGATTTGATGGCGTACTTATTCCTGGTGG 5670

purQ

TCATTCGGGGATTACTTAAGATCTGGGGCAATGGCTAGTGTAGCGCCGATTATTTCGGAAGTTAAACGTCTTGCAGCTGAAGGTAA 5760

purQ

GTATTAGGTGTTTGTAATGGGTTTCAAATTTTAACTGAAATAGGCTTATTACCTGGTGCATTATTGCATAACGATTCACATTTATT 5850

purQ

AGTAGAAATGAAGAGTTAGAAATAGTGAATAATCAAACGGCATTTACAAATCTTTATGAACAAGGTGAAAAAGTTATATATCCTGT 5940

purQ

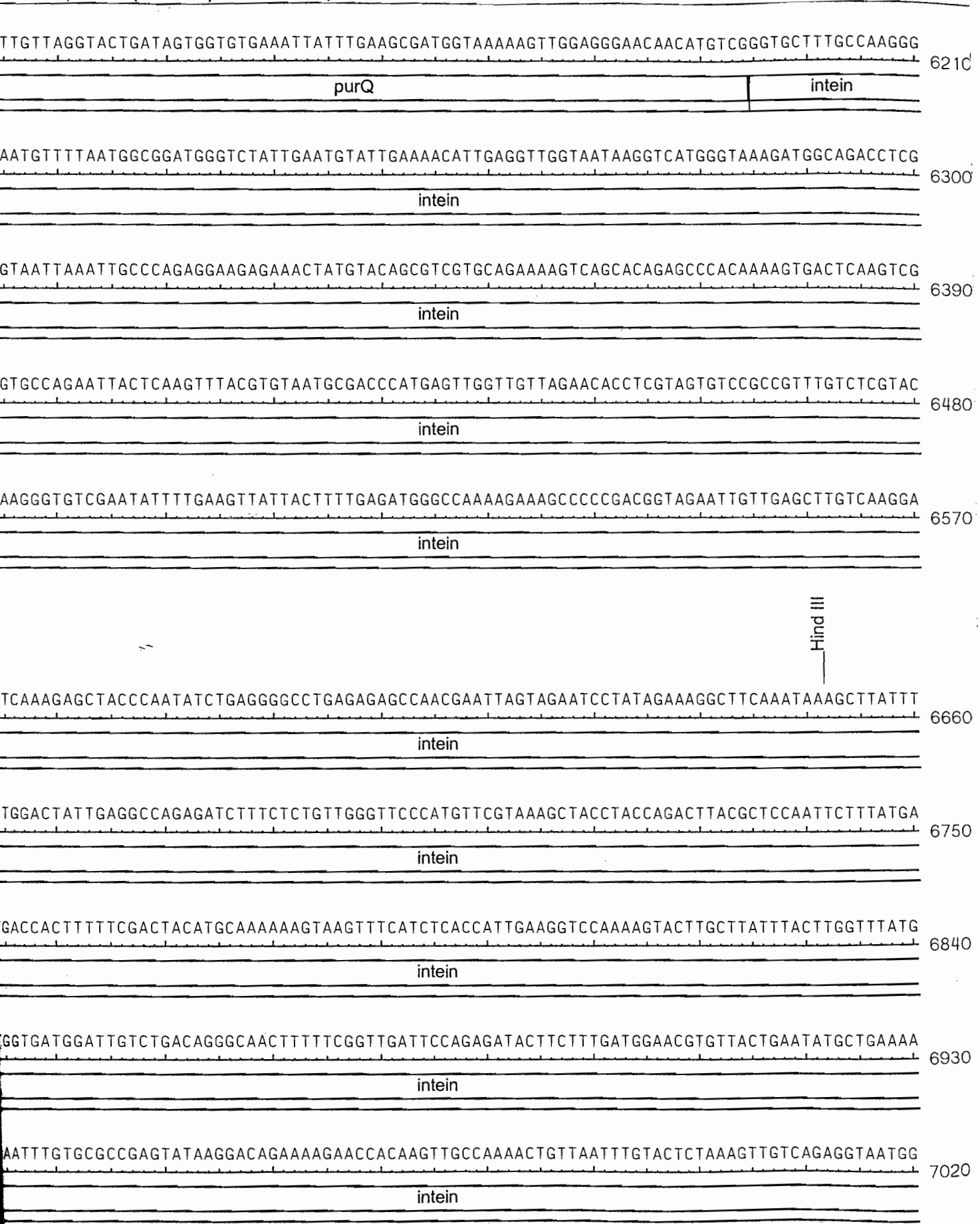
CACGGTGAAGGTCATTATTATTGTACTGATGAAATATATCAACAATTAAAAGCTAACAATCAAATTATTCTGAAATATGTGAATAA 6030

purQ

Nde I

AACGGTTCATATGATGATATTGCAGGAATTGTTAACGAAAAAGGCAATGTATGTGGCATGATGCCACATCCTGAAAGAGCTTTAGA 6120

purQ


TTGTTAGGTACTGATAGTGGTGTGAAATTATTTGAAGCGATGGTAAAAAGTTGGAGGGAACAACATGTCGGGTGCTTTGCCAAGGG
6210
purQ
intein
AATGTTTTAATGGCGGATGGGTCTATTGAATGTATTGAAAACATTGAGGTTGGTAATAAGGTCATGGGTAAAGATGGCAGACCTCG
6300
intein
GTAATTAAATTGCCCAGAGGAAGAGAAACTATGTACAGCGTCGTGCAGAAAAGTCAGCACAGAGCCCACAAAAGTGACTCAAGTCG
6390
intein
GTGCCAGAATTACTCAAGTTTACGTGTAATGCGACCCATGAGTTGGTTGTTAGAACACCTCGTAGTGTCCGCCGTTTGTCTCGTAC
6480
intein
AAGGGTGTCGAATATTTTGAAGTTATTACTTTTGAGATGGGCCAAAAGAAAGCCCCCGACGGTAGAATTGTTGAGCTTGTCAAGGA
6570
intein
Hind III
TCAAAGAGCTACCCAATATCTGAGGGGCCTGAGAGAGCCAACGAATTAGTAGAATCCTATAGAAAGGCTTCAAATAAAGCTTATTT
6660
intein
TGGACTATTGAGGCCAGAGATCTTTCTCTGTTGGGTTCCCATGTTCGTAAAGCTACCTACCAGACTTACGCTCCAATTCTTTATGA
6750
intein
GACCACTTTTTCGACTACATGCAAAAAAGTAAGTTTCATCTCACCATTGAAGGTCCAAAAGTACTTGCTTATTTACTTGGTTTATG
6840
intein
GGTGATGGATTGTCTGACAGGGCAACTTTTTCGGTTGATTCCAGAGATACTTCTTTGATGGAACGTGTTACTGAATATGCTGAAAA
6930
intein
AATTTGTGCGCCGAGTATAAGGACAGAAAAGAACCACAAGTTGCCAAAACTGTTAATTTGTACTCTAAAGTTGTCAGAGGTAATGG
7020
intein

TCGCAATAATCTTAATACTGAGAATCCATTATGGGACGCTATTGTTGGCTTAGGATTCTTGAAGGACGGTGTCAAAAATATTCCTTC 711

intein

CTTGTCTACGGACAATATCGGTACTCGTGAAACATTTCTTGCTGGTCTAATTGATTCTGATGGCTATGTTACTGATGAGCATGGTAT 720

intein

AGCAACAATAAAGACAATTCATACTTCTGTCAGAGATGGTTTGGTTTCCCTTGCTCGTTCTTTAGGCTTAGTAGTCTCGGTTAACGC 729

intein

ACCTGCTAAGGTTGACATGAATGTCACCAAACATAAAATTAGTTATGCTATTTATATGTCTGGTGGAGATGTTTTGCTTAACGTTCT 738

intein

GAAGTGTGCCGGCTCTAAAAAATTCAGGCCTGCTCCCGCCGCTGCTTTTGCACGTGAGTGCCGCGGATTTTATTTCGAGTTACAAGA 747

intein

GAAGGAAGACGATTATTATGGGATTACTTTATCTGATGATTCTGATCATCAGTTTTTGCTTGGATCCCAGGTTGTCGTCCATGCATG 756

intein C

Age I

TGGCCTGACCGGTCTGAACTCAGGCCTCACGACAAATCCTGGTGTATCCGCTTGGCAGGTCAACACAGCTTATACTGCGGGACAATT 7650

CBD

CACATATAACGGCAAGACGTATAAATGTTTGCAGCCCCACACCTCCTTGGCAGGATGGGAACCATCCAACGTTCCTGCCTTGTGGCA 7740

CBD

TCAATGACTGCAGGAAGGGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGC 7830

ACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT 7898

Figure 7. *Gene Annotation.* The following printout shows the annotation of the genes coding for FGAR amidotransferase, using the linear minimap illustration.

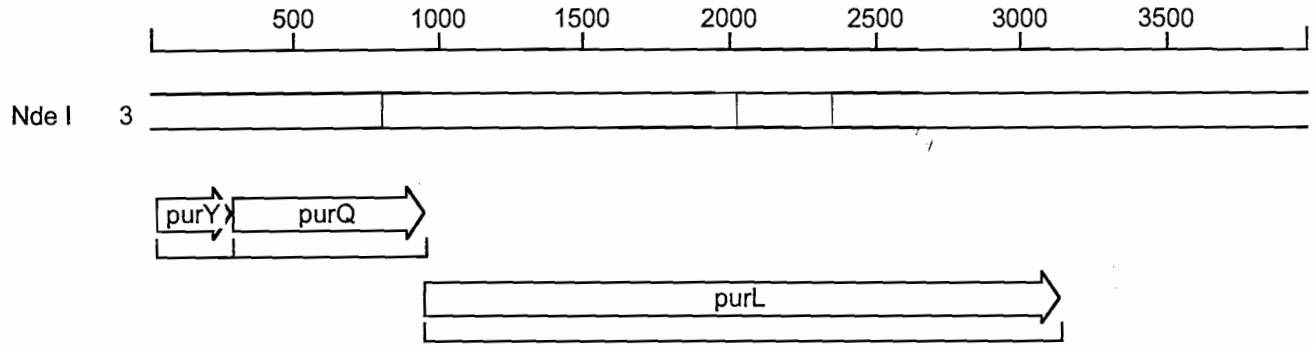
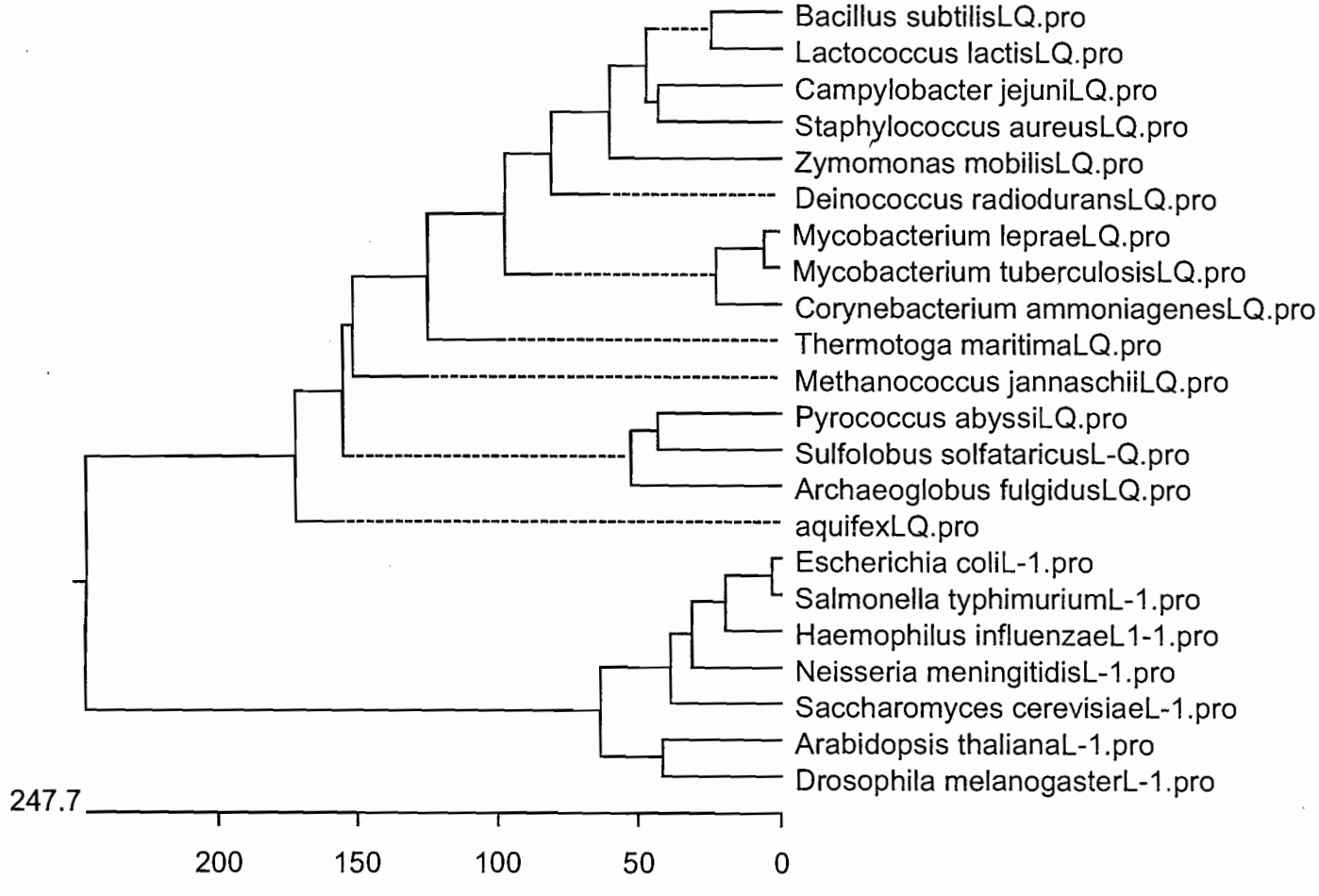

500
1000
1500
2000
2500
3000
3500
Nde I 3
purY
purQ
purL

Figure 8a. *Phylogenetic trees created in MegAlign.* The phylogenetic tree for the proteins coded for by *purL* and *purQ* genes are first, along with the alignment report.
Figure 8b. The next tree represents aligned protein sequences for the *purY* gene, along with the alignment report.


Bacillus subtilisLQ.pro
Lactococcus lactisLQ.pro
Campylobacter jejuniLQ.pro
Staphylococcus aureusLQ.pro
Zymomonas mobilisLQ.pro
Deinococcus radioduransLQ.pro
Mycobacterium lepraeLQ.pro
Mycobacterium tuberculosisLQ.pro
Corynebacterium ammoniagenesLQ.pro
Thermotoga maritimaLQ.pro
Methanococcus jannaschiiLQ.pro
Pyrococcus abyssiLQ.pro
Sulfolobus solfataricusL-Q.pro
Archaeoglobus fulgidusLQ.pro
aquifexLQ.pro
Escherichia coliL-1.pro
Salmonella typhimuriumL-1.pro
Haemophilus influenzaeL1-1.pro
Neisseria meningitidisL-1.pro
Saccharomyces cerevisiaeL-1.pro
Arabidopsis thalianaL-1.pro
Drosophila melanogasterL-1.pro
247.7
200
150
100
50
0

```
1    MSL---L---------LEPSKEQI KEEKLYQQMGVSDDEFALI-ESI LG--RLPNYTEI G  Bacillus subtilisLQ.pro
1    MTL---E---------MSP--EQI QESKI YREWGLTDEEYLKI KDEI LG-GRLPNFTETG  Lactococcus lactisLQ.pro
1    M-----------------DKETI KAHK------I SDEEYAQI L-EI LG--REPNLLELG  Campylobacter jejuniLQ.pro
1    MSK---F---------IEPSVEEI KLEKVYQDMGLSDQEYEKV-CDI LG--RQPNFTETG  Staphylococcus aureusLQ.pro
1    MTEENSS---------ITAETVASH--------GLSPEEYDTI KQAL---GRTPNLVKLG  Zymomonas mobilisLQ.pro
1    MT-----------------SQSLRDQA--ATFGLTTEEYDLFVSQL---GREPNALEAA  Deinococcus radiodiransLQ.pro
1    MI---DT---------VEYAATTPDQPQPFAELGLREDEYQRV-REI LG-RR-PTDTELA  Mycobacterium lepraeLQ.pro
1    ML---DT---------VEHAATTPDQPQPYGELGLKDDEYRRI-RQI LG-RR-PTDTELA  Mycobacterium tuberculosisLQ.pro
1    MTVSNDT---------VDNAKATPELDQPWEELGLKQDEYDKI-VGI LG-RR-PTDAELT  Corynebacterium ammoniagenesLQ.pro
1    MKLR-------------------------------------YLNI LKEKLG--REPTFVELQ  Thermotoga maritimaLQ.pro
1    M-----------------------------------DENDLKYI EKVLG--RKPNHI ELA  Methanococcus jannaschiiLQ.pro
1    M-------------------------------LPREE--KI I RERLG--REPNEVEKA  Pyrococcus abyssiLQ.pro
1    M---------------------------GI NLLPI EMDEI RKRL---GREPNETEWR  Sulfolobus solfataricusL-Q.pro
1    MYRKVDVPFELYEVEI LDASEEELAKI SEEMGLALSVDEMKRI QDYFRQKGRNPYDI ELQ  Archaeoglobus fulgidusLQ.pro
1    M-----------------------ERVWEAYGLTEEEYRKI LKTL---KREPNHVELG  aquifexLQ.pro
1    M-----MEI-LRGSPALSAFRI NKLLAR-FQAARLP-------------------------  Escherichia coliL-1.pro
1    M-----MEI-LRGSPALSAFRI NKLLAR-FQAANLQ-------------------------  Salmonella typhimuriumL-1.pro
1    M----TVKT-FRGSPALSEFRLTQLQQK-CQQYQLP-------------------------  Haemophilus influenzaeL1-1.pro
1    M----SVVLPLRGVTALSDFRVEKLLQK-AAALGLP-------------------------  Neisseria meningitidisL-1.pro
1    M----TDYI-LPGPKALSQFRVDNLI KD-INSYTNS----------------------T  Saccharomyces cerevisiaeL-1.pro
1    MLLQRSSMSQLWGSVRMRTSRLS--LNR-TKAVSLRCSAQPNKPKAAVSTGSFVTADELP  Arabidopsis thalianaL-1.pro
1    MV-------------------------------------------------------------  Drosophila melanogasterL-1.pro
```

```
46  IFSVMWSEHCSYKNSKPILRKFPTSGER-----VLQGPGE--GAGIVDIGDNQAVVFKIE  Bacillus subtilisLQ.pro
46  MYAVMWSEHCCYKNSKPVLKKFPTTGPQ-----VLMGPGE--GAGVVDIGDDLAVVFKAE  Lactococcus lactisLQ.pro
34  VISAMWSEHCSYKSSKKYLNGFPTKAPW-----VIQGPGE--NAGVIDIGQGMAAVFKVE  Campylobacter jejuniLQ.pro
46  IFSVMWSEHCSYKHSKPFLKQFPTSGDH-----VLMGPGE--GAGVVDIGDNQAVVFKVE  Staphylococcus aureusLQ.pro
41  IFSAMWSEHCSYKSSRKHLRELPTTGSQ-----VICGPGE--NAGVVDIGDGQAAIFKME  Zymomonas mobilisLQ.pro
38  IVGAMWSEHCGYKNSRPLFRAFPTTGPQ-----VLQGPGE--NAGVVDIGDGWGVAFKME  Deinococcus radioduransLQ.pro
46  MYSVMWSEHCSYKSSKVHLRYFGETTTEEMRTGMLAGIGE--NAGVVDIGDGWAVTFKVE  Mycobacterium lepraeLQ.pro
46  MYSVMWSEHCSYKSSKVHLRYFGETTSDEMRAAMLAGIGE--NAGVVDIGDGWAVTFKVE  Mycobacterium tuberculosisLQ.pro
49  VYSVMWSEHCSYKSSKTHLRYFGETTTEEMASKILAGIGE--NAGVVDIGDGDAVTFRVE  Corynebacterium ammoniagenesLQ.pro
24  AFSVMWSEHCGYSHTKKYIRRLPKT----------GFEG--NAGVVNLDDYYSVAFKIE  Thermotoga maritimaLQ.pro
24  MFENLWSEHCAYRTSKKLLRMFAKTVNEKTSKNIVVGIGD--DAAVIRLKNDICLAIAME  Methanococcus jannaschiiLQ.pro
24  MLEVMWSEHVSYKSSRKWLKLLPTKNEH-----VILGPGE--DAGVVKFDESTWIVIGIE  Pyrococcus abyssiLQ.pro
28  VIDAVWSEHCSYKSSKIFLKSFSIDSPN-----VIMGIKDWQDAGAVDIGDGWAVVIKVE  Sulfolobus solfataricusL-Q.pro
61  SLAQAWSEHCCYKSSKYYLRQYLLEASK--ADYVISAIEE--DAGVVEFDDEYAYVTAFE  Archaeoglobus fulgidusLQ.pro
33  VLGALWSEHCSYKSSKKHLKKFPTKAEW-----VVQGPGE--NAGVVKIDEKVWAFKVE  aquifexLQ.pro
30  --VH--------NIYAEYVHF----------AD-L-------NAPLNDDEHAQLERL---  Escherichia coliL-1.pro
30  --VH--------NIYAEYVHF----------AD-L-------NAPLNDSEQAQLTRL---  Salmonella typhimuriumL-1.pro
31  --IT--------SVYAEYLHF----------VE-Q-------KTSLVEDEIVKLQAL---  Haemophilus influenzaeL1-1.pro
32  --EV--------KLSSEFWYF----------VG-S-------EKALDAATVEKLQAL---  Neisseria meningitidisL-1.pro
32  SVIN--------ELRSCYIHY----------VNGI-------AQNLSEQDTKLLEVL---  Saccharomyces cerevisiaeL-1.pro
58  SLVE--------KPAAEVIHFYRVPLIQESANAEL-------LKAVQTKISNQIVSLTTE  Arabidopsis thalianaL-1.pro
3   ----------------ILRYYDVQAHSAAEEESV-------LRRLREE-DGAVVSVRME  Drosophila melanogasterL-1.pro
```

```
99   SHNHPSALEPYQGAATGVGGIIRDVFSMGARPIAVLNSLRFGEL--------T-SPRVKY  Bacillus subtilisLQ.pro
99   SHNHPSYVEPYEGAATGSGGIIRDIFSMGARPIAILDSLRFGPI--------D-NGKTRH  Lactococcus lactisLQ.pro
87   SHNHPSFIEPFAGAATGVGGILRDVFTMGARVVAGLNSLKFGDIHDE-----KCGKHQKY  Campylobacter jejuniLQ.pro
99   SHNHPSAIEPYQGAATGVGGIIRDIVSIGARPINLLNSLRFGEL--------D-NKQNQR  Staphylococcus aureusLQ.pro
94   SHNHPSYIEPYQGAATGVGGILRDVFTMGARPVANLKHLRFGSP--------K-HPKTPH  Zymomonas mobilisLQ.pro
91   SHNHPSAVEPVQGAATGVGGILRDIFAMGARPFAVLDSLRFGNP--------D-SPRTRF  Deinococcus radioduransLQ.pro
104  SHNHPSYVEPYQGAATGVGGIVRDIMAMGARPVAVMDQLRFGAA--------D-ALDTRR  Mycobacterium lepraeLQ.pro
104  SHNHPSYVEPYQGAATGVGGIVRDIMAMGARPVAVMDQLRFGAA--------D-APDTRR  Mycobacterium tuberculosisLQ.pro
107  SHNHPSFVEPYQGAATGVGGIVRDIMAMGARPIAVMDQLRFGPA--------D-APDTAR  Corynebacterium ammoniagenesLQ.pro
71   SHNHPSAIEPYNGAATGVGGIIRDVLAMGARPTAIFDSLHMS----------------R  Thermotoga maritimaLQ.pro
82   SHNHPSYIDPYNGAATGVGGIVRDVLSMGAKPIALLDPLRFGDIFGK-----E-GDKVRW  Methanococcus jannaschiiLQ.pro
77   SHNHPSAVEPYGGAATGIGGIVRDILCMGARPIALLDPIRFGPL--------E-KEKNRY  Pyrococcus abyssiLQ.pro
83   SHNHPSAIDPFNGAATGVGGIIRDIISKGAKPIALMDMIRVGNLKIK---------KNVW  Sulfolobus solfataricusL-Q.pro
117  SHNHPSAIEPYGGAATGIGGILRDVLCMGAQPIALIDPLFFGNLDTPYEKLPRGTKHPLY  Archaeoglobus fulgidusLQ.pro
86   SHNHPSYIEPFHGAATGVGGIIRDVLSMGARPIALADSLRFGEF--------N-YHETKR  aquifexLQ.pro
59   ----LKYGPALASHA----------PQG-------KL---------  Escherichia coliL-1.pro
59   ----LQYGPALSSHT----------PAG-------KL---------  Salmonella typhimuriumL-1.pro
60   ----LHYGSMFSELK----------PAG-------YC---------  Haemophilus influenzaeL1-1.pro
61   ----LAAQSVEQTPK----------ARE-GL----HL---------  Neisseria meningitidisL-1.pro
64   ----LTYDSALDIANDPLARQLNDLVA---------NNLPSS-ALGEDTYL---------  Saccharomyces cerevisiaeL-1.pro
103  QSFNIGLESKLKDEKLSV---LKWMLQ---------ETYEP-ENLGTDSFLERKKQEGLH  Arabidopsis thalianaL-1.pro
38   RCYHLEYSAQ-AEHSLALDELLVWLVK---------QPLSKGQSLSRQPALQS---TGSS  Drosophila melanogasterL-1.pro
```

```
150  LFEEVVAGI AGYGNCI GI PTVGGEVQFDSSYEGNPLVNAMCVGLI NHEDI KKGQAKGVGN  Bacillus subtilisLQ.pro
150  I VDQVTAGI AGYGNCI GI PTVGGEVAFDESYAGNPLVNVMCVGLI EHKHI QKGQAKGVGN  Lactococcus lactisLQ.pro
142  LVKGVVNGI SHYGNCMGVPTI GGECAFDECFNGNI LVNAFALGVCKSEDI FYAKAEGVGN  Campylobacter jejuniLQ.pro
150  LLKGVVKGI GGYGNCI GI PTTAGEI EFDERYDGNPLVNAMCVGVI NHDMI QKGTAKGVGN  Staphylococcus aureusLQ.pro
145  LVSGVVAGI GGYGNCVGVPTVGGEVNFHPAYDGNNLVNAMTVAVAETNKI FYSAASGAGN  Zymomonas mobilisLQ.pro
142  LVNGVVDGI AHYGNAI GVPTVGGEVTFHPSYQENPLVNVMALGLLRHEDLATGTMGEVGN  Deinococcus radioduransLQ.pro
155  VLDGVVRGI GGYGNSLGLPNI GGETVFDSCYDGNPLVNALCVGVLRQEDLHLAFASGAGN  Mycobacterium lepraeLQ.pro
155  VLDGVVRGI GGYGNSLGLPNI GGETVFDPCYAGNPLVNALCVGVLRQEDLHLAFASGAGN  Mycobacterium tuberculosisLQ.pro
158  VLPGVVSGI GGYGNSLGLPNI GGETVFDESYAGNPLVNALCVGTLRVEDLKLAFASGTGN  Corynebacterium ammoniagenesLQ.pro
114  I I DGI I EGI ADYGNSI GVPTVGGELRI SSLYAHNPLVNVLAAGVVRNDMLVDSKASRPGQ  Thermotoga maritimaLQ.pro
136  LI EGVVKGI GDYGNRI GVPTVGGECEFDSSFDYNNLVNVVCVGLVKENEI I TGKAKEPGL  Methanococcus jannaschiiLQ.pro
128  LFEYVVKGI ADYGNRI GVPTVGGETEFDESLDNYTLVNVVCVGI MKPEHLVHSYVTKPGL  Pyrococcus abyssiLQ.pro
134  LLKNI I AGI AAYGNSI GVPVVGGELSFDDTYNDNPLVDVAAI GI VRKDKI KPSI VDKAGL  Sulfolobus solfataricusL-Q.pro
177  LLKGVVAGI RDYGNRVGI PTVAGMVFFDNSYLTNCLVNVGCVGI VRKDRI I HSRAGGAGD  Archaeoglobus fulgidusLQ.pro
137  LVKGVVSGI SFYGNCI GVPTVAGETVFEPSYKTNPLVNAFCLGVI PAGRMYRARATREGQ  aquifexLQ.pro
75   ----------LLVTPRPGTI SPWSSKATDI AHNCGL-------QQ-VNRLER  Escherichia coliL-1.pro
75   ----------LLVTPRPGTI SPWSSKATDI AHNCGL-------QQ-VDRLER  Salmonella typhimuriumL-1.pro
76   ----------LI VTPRVGTI SSWSSKATDI AHNCGL-------SK-VNRI ER  Haemophilus influenzaeL1-1.pro
79   ----------FLVTPRLGTI SPWASKATNI AENCGL-------AG-I ERI ER  Neisseria meningitidisL-1.pro
101  ----------I RVVPRSGTI SPWSSKATNI AHVCGL-------QDKVQRI ER  Saccharomyces cerevisiaeL-1.pro
150  AV----------I VEVGPRLSFTTAWSTNAVSI CRACGL-------DE-VTRLER  Arabidopsis thalianaL-1.pro
85   QL----------LLEI GPRFNFSTPYSTNCVNI FQNLGY-------SE-VRRMET  Drosophila melanogasterL-1.pro
```

```
210  TVMYVGAKTGRDGIHGA-TFASEEMSDSSEE-KRSAVQVGDPFMEKLLLEACLEVIQC--  Bacillus subtilisLQ.pro
210  SIFYVGAKTGRDGIHGA-SFASKEFGSGSET-QRSAVQVGDPFMEKLLLEACIEVIQNHG  Lactococcus lactisLQ.pro
202  PVIYVGSKTGRDGLGGA-VMASDSFNEES-KSLRPTVQIGDPFSEKLLMEACLEL--FKT  Campylobacter jejuniLQ.pro
210  SVIYVGLKTGRDGIHGA-TFASEELTEESES-KRPSVQIGDPFVGKKLMEATLEAITF--  Staphylococcus aureusLQ.pro
205  PIVYVGSKTGRDGIHGA-TMASADFGKDAEEK-RPTVQVGDPFSEKLLIEACLELMA--S  Zymomonas mobilisLQ.pro
202  QIVYVGSKTGRDGLGGA-VFSSADLSAASQA-DRPAVQVGDPFMEKLLLEATLEAIQ--A  Deinococcus radioduransLQ.pro
215  KIILFGACTGLDGIGGVSVLASDTFDAEGARKKLPSVQVGDPFMEKVLIECCLEL--YAG  Mycobacterium lepraeLQ.pro
215  KIILFGARTGLDGIGGVSVLASDTFDAEGSRKKLPSVQVGDPFMEKVLIECCLEL--YAG  Mycobacterium tuberculosisLQ.pro
218  KVMLFGSRTGLDGIGGVSVLGSASFE-EGEERKLPAVQVGDPFAEKVLIECCLEL--YAA  Corynebacterium ammoniagenesLQ.pro
174  VIVIFGGATGRDGIHGAS-FASEDLTGDKATKL--SIQVGDPFAEKMLIEAFLEMVE--E  Thermotoga maritimaLQ.pro
196  SLILIGS-TGRDGIGGAS-FASKDLTEESEEE-RPSVQVGDAFSEKCLIDAVLEAVK--T  Methanococcus jannaschiiLQ.pro
188  KLVLVGNRTGRDGIHGV-TFASEELSENAEEEDRSAVQIPDPFTEKLLIEATLEAV--YT  Pyrococcus abyssiLQ.pro
194  KLVLAGL-TGVDGLGGAS-FASRKLSGEDEI---GAVQIADPFAGKIILDVTLEI----A  Sulfolobus solfataricusL-Q.pro
237  LFVLAGGKTGRDGIHGV-TFASAELDEKSEEEARGAVQLGNPIMKEPLIHACLEVVE--K  Archaeoglobus fulgidusLQ.pro
197  LLFLIGSSTGRDGIFGA-VMASGEFSEDVEE-KRPNVQIGDPYFGKKLVEAIMEIVE--K  aquifexLQ.pro
109  GVAYY--------IEAG-TLTNEQWQQVTAE--LHDRMMETVFFALDDAEQ-LFAHHQPT  Escherichia coliL-1.pro
109  GVAYY--------IEAS-TLTAEQWRQVAAE--LHDRMMETVFSSLTDAEK-LFIHHQPA  Salmonella typhimuriumL-1.pro
110  GIAYY--------FNIERDLTEAELATLKDL--LHDRMLETVLNHETEAAL-LFTQQEPK  Haemophilus influenzaeL1-1.pro
113  GMAVW--------LEGA--LTDEQQQQWAAL--LHDRMTESVLPDFQTASK-LFHHLESE  Neisseria meningitidisL-1.pro
136  GLALL--------IKTVPGFPLLENLNDISLKCVYDRMTQQLYLTEPPNTMSIFTHEEPK  Saccharomyces cerevisiaeL-1.pro
187  SRRYL--------LFSKEPLLENQIKEFAAM--VHDRMTECVYTQKLV---SFETNVVPE  Arabidopsis thalianaL-1.pro
122  STRYL--------VTFGEGSKAPEAARFVPL--LGDRMTQCLYTEENTPKASFDEQLPER  Drosophila melanogasterL-1.pro
```

```
266  DALVGIQDMGAAGLTSSSAEMASKAGSGIEMNLDLIPQRETGMTAYEMMLSESQERMLLV  Bacillus subtilisLQ.pro
268  DILVGIQDMGAAGLVSSTSEMASKAGSGLRLNLDNVPQRETEMIPYEMMLSESQERMVLC  Lactococcus lactisLQ.pro
258  DYIVGIQDMGAAGLTSSSFEMAGRSGSGMKLYLDKTPMRESGMTPYELMLSESQERMLIC  Campylobacter jejuniLQ.pro
266  DELVGIQDMGAAGLTSSSSEMAAKGGSGLHLRLEQVPTREPGISPYEMMLSETQERMLLV  Staphylococcus aureusLQ.pro
261  DAIVAIQDMGAAGLTSSAVEMASKGEVGIELDMDMVPCREEGMTPYEMMLSESQERMLMV  Zymomonas mobilisLQ.pro
258  GLVAGVQDMGAAGLVSSTCEMAYRASLGITMDLDKVPTREEGMVPMELCLSESQERMILV  Deinococcus radiodoransLQ.pro
273  GLVIGIQDLGGAGLSCATSELASAGDVGMAIQLDTVPRRAKDMTPAEVFCSESQERMCAV  Mycobacterium lepraeLQ.pro
273  GLVIGIQDLGGAGLSCATSELASAGDGGMTIQLDSVPLRAKEMTPAEVLCSESQERMCAV  Mycobacterium tuberculosisLQ.pro
275  GVVVGIQDLGGGGLACATSELAAAGDGGMVVNLDNVPLRAENMSAAEILASESQERMCAV  Corynebacterium ammoniagenesLQ.pro
229  GLVEGAQDLGAGGVLSATSELVAKGNLGAIVHLDRVPLREPDMEPWEILISESQERMAVV  Thermotoga maritimaLQ.pro
251  GKVKAMKDLGAAGLSGASSEMCYGGGVGCELYLENVVLREP-LTPYEIMVSESQERMLLA  Methanococcus jannaschiiLQ.pro
245  GKVKALKDLGGGGLTCAASEMAGKKGLGAVIYADRVPLREPGMTPLEVMISESQERMLFA  Pyrococcus abyssiLQ.pro
245  DKVEAIKDLGGGGLAVAVTEITN--GLGATVDIEKIPLRVKNMNPSDVIISETQERMLYA  Sulfolobus solfataricusL-Q.pro
294  GLLTGMKDLGGGGLSCVIGEMALAAGFGAEVYLDKVPLKEEGMAPWEIWISESQERMMLT  Archaeoglobus fulgidusLQ.pro
253  DLIVGMQDLGAAGLAGSASEIAAKSEKGVELYLENVPLREKDMNPYEILLSESQERMLLV  aquifexLQ.pro
157  PVT--SVDLLGQG----RQALIDANLRLGLALAEDEIDYLQDAF-TKLGRNPNDIELYM  Escherichia coliL-1.pro
157  PVS--SVDLLGEG----RQALIDANLRLGLALAEDEIDYLQEAF-TKLGRNPNDIELYM  Salmonella typhimuriumL-1.pro
159  ALT--TIDILNGG----RQALEQANIALGLALADDEMDYLVESF-TALKRNPQDVELYM  Haemophilus influenzaeL1-1.pro
160  TFS--TVDVLGGG----KEALVKANTEMGLALSADEIDYLVENY-QALQRNPSDVELMM  Neisseria meningitidisL-1.pro
188  PLV--HVPLTPKDTKQSPKDILSKANTELGLALECGEMEYVIHAFVETMKRDPTDVELFM  Saccharomyces cerevisiaeL-1.pro
234  EVK--YVPVMEKG----RKALEEINQEMGLAFDEQDLQYYTRLFREDIKRDPTNVELFD  Arabidopsis thalianaL-1.pro
172  QANWHFVPVLEEG----RAALERINQELGLAFNDYDLDYYHDLFAKELGRNPTTVELFD  Drosophila melanogasterL-1.pro
```

```
326  IERGREQEIIDIFDKYDLEAVSVGHVTDDKMLRLTH-KGEVVCELPVDALAEEAPVYHKP  Bacillus subtilisLQ.pro
328  VKKGHEQEIIDLFKKYDLDAVNIGEVTDDGFYTLYH-KGQMVAHVPVDSLAEDAPTYYRE  Lactococcus lactisLQ.pro
318  AKKGYEDKVIEIFKKWDLDAVVMGEVTNTGKMELFWH-DELVGLIPIEPLSEKAPILSRP  Campylobacter jejuniLQ.pro
326  V-----------------------------------------------------------  Staphylococcus aureusLQ.pro
321  LKPGREAEAEAIFKKWELDFAIIGRVTDSKHMVLTW-KGDIVCDIPLAPLADNAPCYDRP  Zymomonas mobilisLQ.pro
318  PVPGKEQALHDLLAKWELDVVTIGEVEAHDRYRLTW-KGEVVCDLPV-ALLNEAPKYTRE  Deinococcus radioduransLQ.pro
333  VAPENVDAFLAVCRKWEVLATVIGEVTDGDRLRITWH-GETVVDVPPRTVAHEGPVYQRP  Mycobacterium lepraeLQ.pro
333  VSPKNVDAFLAVCRKWEVLATVIGEVTDGDRLQITWH-GETVVDVPPRTVAHEGPVYQRP  Mycobacterium tuberculosisLQ.pro
335  VSPDNVEKFREICEKWDVTCAEIGEVTDKKDTYLVYHNGELVVDAPPSTI-DEGPVYERP  Corynebacterium ammoniagenesLQ.pro
289  TSPQKASRILEIARKHLLFGDVVAEVIEEPVYRVM-YRNDLVMEVPVQ-LLANAPEEDIV  Thermotoga maritimaLQ.pro
310  VEPGSEEEIIEIFKKYELPASVIGKTIPEKRI-IAKYKGEVVVDLPLD-LLCEAPLYDRE  Methanococcus jannaschiiLQ.pro
305  VEPEDVEELAKIFEKYELEWAVVGEIIEEPRFVVYW-KGDKVADLPIE-LLTNVPTIEWP  Pyrococcus abyssiLQ.pro
303  VEEKNVKEVCEAFEEYEYPCSVIGEITNEPVIKF-RYIGKDLVSLPTNVLL-NPPRFLWP  Sulfolobus solfataricusL-Q.pro
354  VRPEHIDEVLYIFQKWDVPATVVGKVIPEKITRIY-YKGYKIYEMDTE-FVTSGPEYCRP  Archaeoglobus fulgidusLQ.pro
313  VEEENVEKVKEIANKWHLEGAVVGKITDDDTFRAYY-KGELVAELPVSLIVDEAPVYDRP  aquifexLQ.pro
209  FAQANSEHCRHKI--FNADWVIDGEQQPKSLFKMIKNTFETTPDHVLSAYKDNAAVM---  Escherichia coliL-1.pro
209  FAQANSEHCRHKI--FNADWMIDGKPQPKSLFKMIKNTFETTPDYVLSAYKDNAAVM---  Salmonella typhimuriumL-1.pro
211  FAQANSEHCRHKI--FNADWMIDGKKQDKSLFKMIKNTFEQTPDFVLSAYKDNAAVM---  Haemophilus influenzaeL1-1.pro
212  FAQANSEHCRHKI--FNADFILNGEKQPKSLFGMIRDTHNAHPEGTVVAYKDNSSVI---  Neisseria meningitidisL-1.pro
246  FAQVNSEHCRHKI--FNADWTIDGIKQQFTLFQMIRNTHKLNPEYTISAYSDNAAVL---  Saccharomyces cerevisiaeL-1.pro
287  IAQSNSEHSRHWF--FAGNMVIDGKPMDKSLMQIVKSTWEANRNNSVIGFKDNSSAI---  Arabidopsis thalianaL-1.pro
227  CAQSNSEHSRHWF--FRGRMVIDGVEQPKSLIRMIMDTQAHTNPNNTIKFSDNSSAM---  Drosophila melanogasterL-1.pro
```

```
385  SQEPAYYREFLETDVP--APQIEDA--NEMLKALLQQPTIASKEWVYDQYDYMVRTNTVV  Bacillus subtilisLQ.pro
387  AKVPERIQKFTDSEKY--LPEITDSAVSEIFKKLLAQPTIASKKSIYETYDSRVMTNTVV  Lactococcus lactisLQ.pro
377  TSEPKYLSEIKNYKFE------LKSSVQELFIQMLQNENINNKAFIYDQFDSSVQTNTIK  Campylobacter jejuniLQ.pro
327  ----------------------------------------------------------  Staphylococcus aureusLQ.pro
380  W-------VATPKAKALGAVPASGSITDNLVTLVGSPDLASRRWMWEQYDNMVGADTVQ  Zymomonas mobilisLQ.pro
376  GVESADIRAARERDLS--GVPLPGD-LGAVLLELLSHPTIASKRPIFERYDHQVMTNTVV  Deinococcus radioduransLQ.pro
392  VSRPESQEALNADSSKGLPRPVSGDELRATLLALLGSPHLCSRAFITEQYDRYVRGNTVL  Mycobacterium lepraeLQ.pro
392  VARPDTQDALNADRSAKLSRPVTGDELRATLLALLGSPHLCSRAFITEQYDRYVRGNTVL  Mycobacterium tuberculosisLQ.pro
394  YARPQWQDEI--QQAPEIARP---ESLVQAFKDMVSSPALSSRAFITEQYDRYVRGNTVK  Corynebacterium ammoniagenesLQ.pro
347  EYTPGKIPEFKRVEFEEVNARE----------------------VFEQYDHMVGTDTVV  Thermotoga maritimaLQ.pro
368  GKEDLKEKEDDKEKIK------MPEDLNAVLLKLLESPNICSKEWMYQQYDHEVQIRTVV  Methanococcus jannaschiiLQ.pro
363  MKEYK-----LEEDVE--TPDIA---LSKAFDLVWSSPNIVAKRWVWEQYDHEVQGRTVV  Pyrococcus abyssiLQ.pro
361  IKNTK----------KNVEEKIVDLPLESAIYTVLTHPDLVSKGWAYSQFDYEVNTSTVV  Sulfolobus solfataricusL-Q.pro
412  YVARKPEKELHEEV-------EPPADYVKTFMKMLSHPNAAFKEWMVRQYDHEVRASTVL  Archaeoglobus fulgidusLQ.pro
372  YKEPEYMKEV--RNFN--QEELPQTDVKEALKKLLSSPNISCKEWVYTQYDYQVGTNTLL  aquifexLQ.pro
264  ------------EGSEVGRYFADHETGR-Y----DFHQEPAHILMKVETHNH---PTAI  Escherichia coliL-1.pro
264  ------------EGSAVGRYFADHNTGR-Y----DFHQEPAHILMKVETHNH---PTAI  Salmonella typhimuriumL-1.pro
266  ------------EGSKVGRWFPDPD-GQ-Y----RVHQEDVHILMKVETHNH---PTAI  Haemophilus influenzaeL1-1.pro
267  ------------EGAKVERFYPNAAENQGY----RFHEEDTHIIMKVETHNH---PTAI  Neisseria meningitidisL-1.pro
301  ------------DSENDAFFFAPNSTTKEW----TSTKERIPLLIKVETHNH---PTAV  Saccharomyces cerevisiaeL-1.pro
342  ------------RGFLVNQLRP---LLPGSVCLLDVSARDLDILFTAETHNF---PCAV  Arabidopsis thalianaL-1.pro
282  ------------VGFDHQTIVPSSVVAPGAVRLQSVQS---DLIFTAETHNM---PTAV  Drosophila melanogasterL-1.pro
```

```
441  AP--------GS-DAGVLRIR-------GTKKALAMTTDCNARYLYLDPEVGGKIAVAEA  Bacillus subtilisLQ.pro
445  AP--------GS-DAAVLRVR-------GTNKALAMTTDCNARYLYLDPEKGGAIAVAEA  Lactococcus lactisLQ.pro
431  AD--------GRLGASVIRIK-------ENGASVAMAIECNSRLNYVNSKIGAALAVASA  Campylobacter jejuniLQ.pro
327  ------------------------------------------------------------  Staphylococcus aureusLQ.pro
432  CP--------G-GDAAVVRVH-------GTEKALAMSVDVTPRYCRADPEEGGKQAVAEC  Zymomonas mobilisLQ.pro
433  VP--------GAADAAVLRVK-------GSPMGVAATSDCNPRFVQLDPYAGAAAAVAEA  Deinococcus radioduransLQ.pro
452  A---------EHADAGVLRI-------DESTGRGIALSTDASGRYTRLDPYAGAQLALAEA  Mycobacterium lepraeLQ.pro
452  A---------EHADGGMLRI-------DESTGRGIAVSTDASGRYTLLDPYAGAQLALAEA  Mycobacterium tuberculosisLQ.pro
449  A---------KQSDSGVLRI-------NEETSRGVAISADGSGRYTKLDPNMGARLALAEA  Corynebacterium ammoniagenesLQ.pro
384  PP--------GFG-AAVMRIKR-------DGGYSLVTHSRADLALQDTYWGTLIAVLES  Thermotoga maritimaLQ.pro
422  KP--------GK-DAAVLRI-------NEVYPMGIALTTDCNSRYCKLNPYVGAVNAVAEA  Methanococcus jannaschiiLQ.pro
413  KP--------GF-DAAVLKIN-------GEYGLAITSDGNPSYCYLNPYHGAMGTVAEV  Pyrococcus abyssiLQ.pro
411  KP--------GDADSAVVSLP-------NGKLLAIKADANPDMCAEDGYECGKGIVAEA  Sulfolobus solfataricusL-Q.pro
465  KPLQGKMNFETHGDAAVIKPTRSFR-------GLAITADVNPWMCKVDPYWGAASSFDEM  Archaeoglobus fulgidusLQ.pro
428  IP--------GH-DAAVLRLKWWLRPELTTEKGIAISSEGNGRMVYLNPYEGGKFVVAEV  aquifexLQ.pro
303  SP------WPGAATGSGGEI----RDEGATGRGAKPKAG-------LVGFSVSNLRIPGF  Escherichia coliL-1.pro
303  SP------WPGAATGSGGEI----RDEGATGRGAKPKAG-------LVGFSVSNLRIPGF  Salmonella typhimuriumL-1.pro
304  SP------FPGAATGSGGEI----RDEGATGRGAKPKAG-------LTGFSVSNLVIPNF  Haemophilus influenzaeL1-1.pro
307  AP------FAGAATGAGGEI----RDEGATGKGSRPKAG-------LTGFTVSNLNIPGL  Neisseria meningitidisL-1.pro
341  SP------FPGAATGSGGEI----RDEGATGRGSKTKCG-------LSGFSVSDLLIPGN  Saccharomyces cerevisiaeL-1.pro
383  AP------YPGAETGAGGRI----RDTHATGRGSFVVAS-------TSGYCVGNLNMEGS  Arabidopsis thalianaL-1.pro
323  AP------FSGATTGTGGRL----RDVQGVGRGGVPIAG-------TAGYCVGALHIPGY  Drosophila melanogasterL-1.pro
```

```
485  ARNI I CSGAEPLAVTDNLNFGNPEK- PEI LWQI EKAADGI SEACNVLSTPVI GGNVSLYN  Bacillus subtilisLQ.pro
489  ARNI VASGGKPLAI TDCLNFGNPEK- PEQFWELTTAADGI SRSCLALDTPVI SGNVSLYN  Lactococcus lactisLQ.pro
476  GRKVACTGAKPLAI SDCLNYGNPQN- PEVMWQFAQGCEGI KEACKELNTPVVSGNVSLYN  Campylobacter jejuniLQ.pro
327  ------------------------------------------------------------  Staphylococcus aureusLQ.pro
476  YRNI TAVGALPLASTDCLNFGNPER- PEI MGQI VGAI KGI GEACRALDMPI VSGNVSLYN  Zymomonas mobilisLQ.pro
478  ARNLACVGATPLAI TDNLNFGNPHR- PEVYYQLQQAVQGI ADACRALNTPVTGGNVSLYN  Deinococcus radioduransLQ.pro
497  YRNVAVTGATPVAVTNCLNFGSPED- PGVMWQFAQAVRGLADGCAALKI PVTGGNVSFYN  Mycobacterium lepraeLQ.pro
497  YRNVAVTGATPVAVTNCLNFGSPED- PGVMWQFTQAVRGLADGCADLGI PVTGGNVSFYN  Mycobacterium tuberculosisLQ.pro
494  YRNVAVTGARPYAVTNCLNFGSPEN- TDVMWQFREAVHGLADGSKELNI PVSGGNVSFYN  Corynebacterium ammoniagenesLQ.pro
427  VRKTLSVGAEPLAI TNCVNYGDPDV- DPV- - GLSAMMTALKNACEFSGVPVASGNASLYN  Thermotoga maritimaLQ.pro
467  VRNLATVGAEPI AMLDNLNFGNPER- PERFWQLAECI KGLADAAEFFEI PVVGGNVSLYN  Methanococcus jannaschiiLQ.pro
456  VRNLVSVGAKPLALVDNLNFASPER- PEVYWSFVETVKGLADAAKAFDLAYVSGNVSFYN  Pyrococcus abyssiLQ.pro
455  YRNLATVGARGMVAVDHLQFGDPKK- AEVYYTFVEAI RGI GEATRFFNI PI VGGKVSFYN  Sulfolobus solfataricusL-Q.pro
518  VRNLVAVNAVPHSFNDCLNFGNPEK- PERMGEFVEAVKALGWMAKGVGVPCVSGNVSFYN  Archaeoglobus fulgidusLQ.pro
479  CRNLACVGAKPLAI TDCLNFGNPER- PEI MWQFVKAVEGMAEACEELGI PVVSGNVSLYN  aquifexLQ.pro
346  EQPW- - - - - - - - - - EE- DFGKPERI VTALDI MTEGPLGGAAFNNEFGRPALNGYFRTYE  Escherichia coliL-1.pro
346  EQPW- - - - - - - - - - EE- DFGKPERI VTALDI MTEGPLGGAAFNNEFGRPALTGYFRTYE  Salmonella typhimuriumL-1.pro
347  EQPW- - - - - - - - - - EN- PLSKPNRI ASALDI MI DAPLGSAAFNNEFGRPALLGYFRTYE  Haemophilus influenzaeL1-1.pro
350  KQPW- - - - - - - - - - EQ- DYGKPEHI SSPLDI MI EGPI GGAAFNNEFGRPNLLGYFRTFE  Neisseria meningitidisL-1.pro
384  EQPW- - - - - - - - - - EL- NI GKPYHI ASALDI MI EAPLGSAAFNNEFGRPCI NGYFRTLT  Saccharomyces cerevisiaeL-1.pro
426  YAPW- - - - - - - - - - EDSSFQYPSNLASPLQI LI DASNGASDYGNKFGEPMI QGYTRTFG  Arabidopsis thalianaL-1.pro
366  KQPY- - - - - - - - - - EPLDFKYPATFAPPLQVLI EASNGASDYGNKFGEPVI SGFALSYG  Drosophila melanogasterL-1.pro
```

```
544  E----SNGT-AIYP-TPVIGMVGLIEDTAHITTQHFK--QAGDLVYVIGET-----KPEF  Bacillus subtilisLQ.pro
548  E----TNGS-AILP-TPMIGMVGLIEDVKNITTQEFK--KAGDLIVLVGQI-----FDDF  Lactococcus lactisLQ.pro
535  ETEG-----VSIYP-SPTIVSVGVLEDANKTLKASFE--KENLSVYLLGES-----LGEF  Campylobacter jejuniLQ.pro
327  ----------------------------------------------------------  Staphylococcus aureusLQ.pro
535  ETRQDDGSSLAILP-TPTIGGVGLLQDWRDSTTIAFK--NTGEEIYLVGNSGQ----GHL  Zymomonas mobilisLQ.pro
537  QYTEGDHKV-AIHP-TPTIGMVGVLPDVTVRASLNLK--AAGQTLLLLGHRAEKGWSDSI  Deinococcus radioduransLQ.pro
556  QTGAV-----AILP-TPVVGVLGVLDNVARRIHTSLGT-EPGEILMLLGDT-----YDEF  Mycobacterium lepraeLQ.pro
556  QTGSA-----AILP-TPVVGVLGVIDDVRRRIPTGLGA-EPGETLMLLGDT-----RDEF  Mycobacterium tuberculosisLQ.pro
553  QTGDE-----PILP-TPVVGVLGVIDDVHKALAHDLGGIDEPETLILLGET-----KEEF  Corynebacterium ammoniagenesLQ.pro
484  TYQGK-----PIPP-TLVVGMLGKVN------PQKVAKPKPSKVFAV-------GWND--  Thermotoga maritimaLQ.pro
526  ETVIEGKEH-PINP-TPAIFVLGKVEDVEKVPGVLDNKIKEGDILIITNET-----KDEM  Methanococcus jannaschiiLQ.pro
515  EVVD----R-PVKP-TPVVAGIG---KVKLKDIPRGP--RDGDVIALIGST-----RREL  Pyrococcus abyssiLQ.pro
514  E----NNQGRPIKP-TPLIVMAGLVQD-------KLLKNRVEDNLYVVSVGYTR---KEL  Sulfolobus solfataricusL-Q.pro
577  ETPYG-----AVAP-TPSLLGVGIVEDVRKAITSEF---KGRGAVILVGET-----HNEF  Archaeoglobus fulgidusLQ.pro
538  ETVEKNEIR-NVFP-TPIVVGVGVLEKAEKYTPSKVE--KESEL-YLVGNLEENLR---L  aquifexLQ.pro
394  EKVNSHNG-EELRGYHKPIMLAGGIGNIRADHVQKGE-INVGAKLVVLGGPAMNIGLGGG  Escherichia coliL-1.pro
394  EKVNSHNG-EELRGYHKPIMLAGGIGNIRADHVQKGE-IVVGAKLIVLGGPAMNIGLGGG  Salmonella typhimuriumL-1.pro
395  EKVNSFAG-KEVRGYHKPIMLAGGIGNIRGEQVQKGE-IPIGAKLIVLGGAAMNIGLGGG  Haemophilus influenzaeL1-1.pro
398  EKFDG-----QVRGYHKPIMIAGGLGSIQAQQTHKDE-IPEGALLIQLGGPGMLIGLGGG  Neisseria meningitidisL-1.pro
432  TKVLNHQGKEEIRGFHKPIMIAGGFGTVRPQFALKNTPITPGSCLIVLGGQSMLIGLGGG  Saccharomyces cerevisiaeL-1.pro
475  MRLPS--G--DRREWLKPIMFSAGIGQIDHTHITKGEP-EVGMLVVKIGGPAYRIGMGGG  Arabidopsis thalianaL-1.pro
415  LNSAA--DASQRDEYVKPIMFSGGLGTMPATMREKLPP-ARGQLLAKIGGPVYRIGVGGG  Drosophila melanogasterL-1.pro
```

```
591  AGSELQKMTEGRIY---GKAPQIDLDVELSRQKAL---LDAIKKGFVQSAHDVSEGGLGV  Bacillus subtilisLQ.pro
595  SGSELQKMLTGEIS---GKI-DFDLETEKVNQDFV---LKAITDGLINSAHDLSEGGLAI  Lactococcus lactisLQ.pro
582  SGSMVMKIQDKKVS---GSLKELDYKAELALWDLL---YKANQNSLLECANSVGIGGIAM  Campylobacter jejuniLQ.pro
327  ----------------------------------------------------------  Staphylococcus aureusLQ.pro
588  GQSIWLRKIAGREE---GTAPSVDLAQEKATGDFI---RAMIQDGMLCAVHDISDGGLAV  Zymomonas mobilisLQ.pro
593  GASQYLETVHGLEA---GQVPPVDLDLAQKVVDGT---LALIRAGLTDTAHDCAEGGLAV  Deinococcus radioduransLQ.pro
604  DGSVWAQVMAGHLG---GLPPMVDLAREKLLAEVL---SSASRDELVSAAHDLSEGGLAQ  Mycobacterium lepraeLQ.pro
604  DGSVWAQVTADHLG---GLPPVVDLAREKLLAAVL---SSASRDGLVSAAHDLSEGGLAQ  Mycobacterium tuberculosisLQ.pro
602  GGSIWQQVSGGGLQ---GLPPQVDLANEAKLADFF---VGNTS---VAASHDLSEGGLAI  Corynebacterium ammoniagenesLQ.pro
523  ----------------------FELEREKELWRAI---RKLSEEG---------------  Thermotoga maritimaLQ.pro
579  GGSEYYKVIHNTEE---GRVPRVDLEKEKKIYEEV---REVVKEGLVSEAVDCSRGGLAV  Methanococcus jannaschiiLQ.pro
559  GGSELYRVL-GIKG---GIAPRVNLEEEKGNALAI---LNLIENDLVTFVHDVSRGGVAV  Pyrococcus abyssiLQ.pro
559  GGSLLSKIF--KIP---SQAPKVRLQEDLLSSEVV---IDSINEGKITFAKDVSRGGLAA  Sulfolobus solfataricusL-Q.pro
623  GGSLYSAVM-GQRC---HKVPRTSPERLKTYSDAM---LESFRKFEVKACHDVSMGGLAV  Archaeoglobus fulgidusLQ.pro
590  DGSEYLKVIHGLIK---GDVPPVDLEKEKILINLL---ISFNNKELITCAHDVSVGGLLI  aquifexLQ.pro
452  AASSMA-SGQSDADLDFASVQRDNPEMERRCQEVIDRCWQLGDANPILFIHDVGAGGLSN  Escherichia coliL-1.pro
452  AASSMA-SGQSDADLDFASVQRDNPEMERRCQEVIDRCWQLGDANPILFIHDVGAGGLSN  Salmonella typhimuriumL-1.pro
453  AASSMD-SGKSKEDLDFASVQRENPEMERRCQEVIDRCWQLGEENPILFIHDVGAGGLSN  Haemophilus influenzaeL1-1.pro
452  AASSMD-TGTNDASLDFNSVQRGNPEIERRAQEVIDRCWQLGDKNPIISIHDVGAGGLSN  Neisseria meningitidisL-1.pro
492  PASSVA-SREGSADLDFASVQRGNPEMERRCQQVIDACVALGNNNPIQSIHDVGAGGLSN  Saccharomyces cerevisiaeL-1.pro
530  AASSM-VSGQNDAELDFNAVQRGDAEMSQKLYRVVRACIEMGEKNPIISIHDQGAGGNCN  Arabidopsis thalianaL-1.pro
472  AASSVEIQGSGDAELDFNAVQRGDAEMENKLNRVVRACLDLGEQNPILAIHDQGAGGNGN  Drosophila melanogasterL-1.pro
```

```
645  AIAESV-MTTENLGANVTVEG-----EAALLF--SESQSRFVVSVKKEHQAAFEATVKDA  Bacillus subtilisLQ.pro
648  ALAESA-FA-NGLGIDVEVDL-----SNAQLF--SETQGRFVLSISPENQAAFEKLLTES  Lactococcus lactisLQ.pro
636  TLAKMFAISS------VGANLTSDFDDEKMIF--DESASRAIIGLSKENEEAFLNLAKEF  Campylobacter jejuniLQ.pro
327  ------------------------------------------------------------  Staphylococcus aureusLQ.pro
642  ALAEMA-LAGN-IGATVEAHD-KAIAEHAYYF--GEDQGRYLVS--STNAVALVSAAEKA  Zymomonas mobilisLQ.pro
647  ALAEMA-IAG-GLGLNVSLDAPASVRADALLF--GEAHSRVIVAV--EDAAAAGAKLDEL  Deinococcus radioduransLQ.pro
658  AIVESA-LAGE-TGCRIALPE--DADPFVMLF--SESAGRVLVAVPRPEESRFRSMCEAR  Mycobacterium lepraeLQ.pro
658  AIVESA-LAGE-TGCRIVLPE--GADPFVLLF--SESAGRVLVAVPRTEESRFRGMCEAR  Mycobacterium tuberculosisLQ.pro
653  AAFEMA-QKNN-VGVDLDLSV-VHEDALTALF--SESASRVLISTASDHLDGILQRASEL  Corynebacterium ammoniagenesLQ.pro
543  ----------------------------AF--ILSSSQLLTRTH--------------  Thermotoga maritimaLQ.pro
633  ALAKMAVLNNIGLEVDLTEYNKNNLRDDILLF--SETSGRIILAVRDENKDKV-------  Methanococcus jannaschiiLQ.pro
612  ALAEL----SAWFNVGVKAKFTSSFKSIDFAF--SESHGRYIITLPEDKVEEAKEIAKIS  Pyrococcus abyssiLQ.pro
611  SLFSILV-HGYGVEISTKSILSDTDNVIENLF--SESSGRFIVLTN--EPEWIVEKSKSK  Sulfolobus solfataricusL-Q.pro
676  CIAEMSFGRGLGFEAARELSFVELF--SESNTRWVVEVPESVAEGYAEFFRAK  Archaeoglobus fulgidusLQ.pro
644  ALLEMV-FRTP---YGLEVEVYTDERPDVFFF--SENPTRVIIGVESDKAEEVKNAVEKA  aquifexLQ.pro
511  AMPELVSDGGRGGKFELREILSDEPGMSPLEIWCNESQERYVLAVAADQLPLFDELCKRE  Escherichia coliL-1.pro
511  AMPELVSDGGRGGKFELRDILSDEPGMSPLEIWCNESQERYVLAVAADQLPLFDELCKRE  Salmonella typhimuriumL-1.pro
512  AMPELVHDGKRGGKFDLRSILCDEKGMSPLEIWCNESQERYVLAVAPENLELFTALCERE  Haemophilus influenzaeL1-1.pro
511  AFPELVNDAGRGAVFKLREVPLEEHGLNPLQIWCNESQERYVLSILEKDLDIFRSICERE  Neisseria meningitidisL-1.pro
551  TLPELVHDNDLGAKFDIRKVLSLEPGMSPMEIWCNESQERYVLGVSPQDLSIFKEICKRE  Saccharomyces cerevisiaeL-1.pro
589  VVKEIIYP--QGAEIDIRAVVVGDHTMSVLEIWGAEYQEQDAILVKAESREILQSICKRE  Arabidopsis thalianaL-1.pro
532  VLKELVEPGFAGAVIFSKEFQLGDPTITALELWGAEYQENNAILCNADQRELLEKICRRE  Drosophila melanogasterL-1.pro
```

```
697  VH----IGEVTADGILAIQ-------NQD---------GQQMIHAQTKELERVWKGAIPC  Bacillus subtilisLQ.pro
699  SASSEVIGKVTDNGILKIN-------E---------------LSISTDEAVSIYEGALPC  Lactococcus lactisLQ.pro
688  GVKAYKLGVSTSQKHFKLD-SIELSK--------------------AELDKLYFESFKE  Campylobacter jejuniLQ.pro
327  ------------------------------------------------------------  Staphylococcus aureusLQ.pro
695  GIPVFRLGVTGGDAVVLNSQSVSLEKLRK---------SHEAFLPELMQ-------....MTE  Zymomonas mobilisLQ.pro
701  GLPYAVLGETVEAPKVTIA-------APA---------QHVHLSVNLESLKTAWEEPLKG  Deinococcus radioduransLQ.pro
712  GLPAMRIG-VVDQGSDSIE-------VRG-----------QFTVSLAELRMTFEAVLPR  Mycobacterium lepraeLQ.pro
712  GLPAVRIG-VVDQGSDAVE-------VQG-----------LFAVSLAELRATSEAVLPR  Mycobacterium tuberculosisLQ.pro
708  GIPAVVVG-TTNDSGN-IT-------FAG-----------E-EVATAELREAWSATLPN  Corynebacterium ammoniagenesLQ.pro
557  ----------------------VETFRE---------YGLKIEVKLPEVRPAHQMVL--  Thermotoga maritimaLQ.pro
684  ------------------------------------------------------------  Methanococcus jannaschiiLQ.pro
666  -----IVGRVGGD---NFA-------LEV---------NGEKVEKDIEELSRIY------  Pyrococcus abyssiLQ.pro
666  GIVASIIGRVNKKTNILTIDNIDYN-------------------------------------  Sulfolobus solfataricusL-Q.pro
727  ELKAEVIGYSGGER-------LDFGA---------------FSVELGEADKSWREGLTK  Archaeoglobus fulgidusLQ.pro
698  GLEWMYIGKTTEEKKIKVT-------F-N---------GDTLLEDELEEYEKLWRTSLEK  aquifexLQ.pro
571  RAPYAVIGEATEELHLSLHDR---------------HFDNQPIDLPLDVLLGKTPKMTRD  Escherichia coliL-1.pro
571  RAPYAVIGDATEEQHLSLHDN---------------HFDNQPIDLPLDVLLGKTPKMTRD  Salmonella typhimuriumL-1.pro
572  RAPFAVIGEATQAEHLILHDS---------------HFDNNPIDLPMNVLLGKTPKMTRE  Haemophilus influenzaeL1-1.pro
571  RCPFAVVGTATDDGHLKVRDD---------------LFSNNPVDLPLNVLLGKPPKTTRT  Neisseria meningitidisL-1.pro
611  RAPFAVVGHATAEQKLIVEDP---------------LLKTTPIDLEMPILFGKPPKMSRE  Saccharomyces cerevisiaeL-1.pro
647  RLSMAVIGTINGGGRCTLIDSTAAAKCSKEGLPPPP----PAVDLELEKVLGDMPKKTFK  Arabidopsis thalianaL-1.pro
592  RCPISFVGVVTGDGRVTLLEK-PAPKDLEQALNASNRSEVSPFDLELKYVLGDMPKRTYD  Drosophila melanogasterL-1.pro
```

```
737  LLKSKA----------------------------------------------------------  Bacillus subtilisLQ.pro
737  LMK-------------------------------------------------------------  Lactococcus lactisLQ.pro
726  QIQ-------...MKILLTLIFS-IVVL----FGRG--EISV------FSGQDENMQKELQK  Campylobacter jejuniLQ.pro
327  -------EK-------------------------------------------------------  Staphylococcus aureusLQ.pro
742  ENSSITAETVASHGLSPEEYDTIKQALGRTPNLV--KLGIFSAMWSEHCSYKSSRKHLRE  Zymomonas mobilisLQ.pro
745  ILG---------...MRLR-----------------------PTPF-PLWRNCT---------  Deinococcus radioduransLQ.pro
752  FFG-------------------------------------------------------------  Mycobacterium lepraeLQ.pro
752  YFG-------------------------------------------------------------  Mycobacterium tuberculosisLQ.pro
746  LFGHAVGAN-------------------------------------------------------  Corynebacterium ammoniagenesLQ.pro
583  ------------------------------V--FS-----------------------------  Thermotoga maritimaLQ.pro
684  ----------------------------LSKLS-------------------------------  Methanococcus jannaschiiLQ.pro
696  ------------------------------------WN-Y----------------------  Pyrococcus abyssiLQ.pro
691  -------LKNIVDNYFNFLEEVMGNG-----...-----------------------------  Sulfolobus solfataricusL-Q.pro
764  FL--------------------------------------------------------------  Archaeoglobus fulgidusLQ.pro
741  LLGST-----------------------------------------------------------  aquifexLQ.pro
616  VQTLKAKGDALAREGI-TIADAVKRVLHLPTVAEKTFLVTIGDRSVTGMVARDQMVGPWQ  Escherichia coliL-1.pro
616  VQTLKAKGDALNRADI-TIADAVKRVLHLPTVAEKTFLVTIGDRTVTGMVARDQMVGPWQ  Salmonella typhimuriumL-1.pro
617  VLSKTVENQSLKIESI-QLKEAFHRVLRLPVVAEKTFLITIGDRSVTGMVARDQMVGPWQ  Haemophilus influenzaeL1-1.pro
616  DKTVAPSKKPFHAGDI-DITEAAYRVLRLPAVAAKNFLITIGDRSVGGMTHRDQMVGKYQ  Neisseria meningitidisL-1.pro
656  TITEALNLPEANLSEIPSLQDAIQRVLNLPSVGSKSFLITIGDRSVTGLIDRDQFVGPWQ  Saccharomyces cerevisiaeL-1.pro
703  FNRIAYAREPLDIAPGITLMDALKRVLRLPSVSSKRFLTTKVDRCVTGLVAQQQTVGPLQ  Arabidopsis thalianaL-1.pro
651  LKREQTPLKELSLPKGLLLDEALERVLSLVAVGSKRFLTNKVDRCVGGLIAQQQCVGPLQ  Drosophila melanogasterL-1.pro
```

```
743  ------------------------------------------------------------  Bacillus subtilisLQ.pro
740  ------------------------------------------------------------  Lactococcus lactisLQ.pro
769  LPQEE----QKIYQNIPPSD------ENNDFESNVD------DPFVAKGSLVLT------  Campylobacter jejuniLQ.pro
329  ------------------------------------------------------------  Staphylococcus aureusLQ.pro
800  LPTTGSQVICGPGENAGVVDIGDGQAAIFKMESHNHPSYIEPYQGAATGVGGILRDVFTM  Zymomonas mobilisLQ.pro
767  ------------------------------------------------------PKAS  Deinococcus radioduransLQ.pro
755  ------------------------------------------------------------  Mycobacterium lepraeLQ.pro
755  ------------------------------------------------------------  Mycobacterium tuberculosisLQ.pro
755  ------------------------------------------------------------  Corynebacterium ammoniagenesLQ.pro
586  ---------ERTPVVDV------------------------PVKEIGTLSR---------M  Thermotoga maritimaLQ.pro
689  ----------------------------SAYI------------------------  Methanococcus jannaschiiLQ.pro
699  ------------------------------------------------------------  Pyrococcus abyssiLQ.pro
714  ------------------------------------------------------------  Sulfolobus solfataricusL-Q.pro
766  ----------....MDREEIKVAV------------------------------  Archaeoglobus fulgidusLQ.pro
746  ------------------------------------------------------------  aquifexLQ.pro
675  VPVANCAVTT-ASLDSYY---GEAMAIGERAPVALLDFAASARLAVGEALTNIAATQIGD  Escherichia coliL-1.pro
675  VPVADCAVTT-ASLDSYY---GEAMSIGERAPVALLDFAASARLAVGEALTNIAATQIGD  Salmonella typhimuriumL-1.pro
676  IPVSDVAVTT-ASLDSYH---GEAMAIGERSPVALLDFSASARLAVAEAITNIAGTLIGE  Haemophilus influenzaeL1-1.pro
675  TPVADCAVTM-MGFNTYR---GEAMSMGEKPTVALFDAPASGRMCVGEAITNIAAANIGD  Neisseria meningitidisL-1.pro
716  VPVADVGVTG-TSLGETII STGEAMAMGEKPVNALISASASAKLSVAESLLNIFAADVKS  Saccharomyces cerevisiaeL-1.pro
763  ITLADVAVIA-QTFTDL---TGGACAIGEQPIKGLLDPKAMARLAVGEALTNLVWAKVTA  Arabidopsis thalianaL-1.pro
711  APLADYALTT-VSHFSH---SGIATSIGTQPLKGLLDPAAMARMCVAEALSNLVFVKISE  Drosophila melanogasterL-1.pro
```

```
743  ----------------------------------------------------------..--------  Bacillus subtilisLQ.pro
740  ----------------------------------------------------------..--------  Lactococcus lactisLQ.pro
807  ----------NDEYPSRVYVGEVFPI----------TIYARTTEN-------  Campylobacter jejuniLQ.pro
329  ----------------------------------------GTMKF--------  Staphylococcus aureusLQ.pro
860  GARPVANL--KHLRFGSPKHPKTPHLVSGVVAGIGG----YGNCVGVPTVGGEVNFHPAY  Zymomonas mobilisLQ.pro
771  TARPKPLL--LLAPLRFRPLA----------RTQRGG--------  Deinococcus radioduransLQ.pro
755  ----------...----------M--------  Mycobacterium lepraeLQ.pro
755  ----------...----------M--------  Mycobacterium tuberculosisLQ.pro
755  ----SVVE------...----------.M--------  Corynebacterium ammoniagenesLQ.pro
609  ----------K----P----------  Thermotoga maritimaLQ.pro
693  ----------IGKVGGS-RLKIKINEKD  Methanococcus jannaschiiLQ.pro
699  --MYDLLE--L-------...----------.--------  Pyrococcus abyssiLQ.pro
714  ----------  Sulfolobus solfataricusL-Q.pro
780  ----------  Archaeoglobus fulgidusLQ.pro
746  ----------..--------  aquifexLQ.pro
731  IKRIKLSANWMAAAGHPGEDAGLYEAVKAVGE-ELCPALGLTIPVGKDSMSMKTRWQEGN  Escherichia coliL-1.pro
731  IKRIKLSANWMAAAGHPGEDAGLYDAVKAVGE-ELCPQLGLTIPVGKDSMSMKTRWQEGN  Salmonella typhimuriumL-1.pro
732  MKRIKLSANWMSAAGHTGEDAGLYEAVKAVGE-ELCPALGLTIPVGKDSMSMKTTWIDNG  Haemophilus influenzaeL1-1.pro
731  IGNIKLSANWMAACGNEGEDEKLYRTVEAVSK-A-CQALDLSIPVGKDSLSMKTVWQDGE  Neisseria meningitidisL-1.pro
775  LNHIKLSANWMSPASHQGEGSKLYEAVQALGL-DLCPALGVAIPVGKDSMSMKMKWDD--  Saccharomyces cerevisiaeL-1.pro
819  LSDVKASGNWMYAAKLEGEGSAMYDAAIALSE-AM-IELGIAIDGGKDSLSMAAH----A  Arabidopsis thalianaL-1.pro
767  LADVKCSGNWMWAAKLPGEGARMFDACKELCQ-IL-EELHIAIDGGKDSLSMAAK----V  Drosophila melanogasterL-1.pro
```

```
746  ----------------------------------.MKFAVIVLPGSNCDI----DMYH  Bacillus subtilisLQ.pro
743  ----------------------------------.MKFAVIQFPGSNCDF----DLLW  Lactococcus lactisLQ.pro
832  ----IKFDFNISVEKTNLS----------------FLNPDAKWEFINNEYKT----TLWF  Campylobacter jejuniLQ.pro
334  ----------------------------------AVLVFPGSNCDRDMFN----AAIK  Staphylococcus aureusLQ.pro
914  DGNNLVNAMTVAVAETNKIFY------------SAASGAGNPIVYVGSKTGR----DGIH  Zymomonas mobilisLQ.pro
796  ----------------------------------VMKTAVIQFPGSNCDA----DALH  Deinococcus radioduransLQ.pro
760  ----------------------------------NARIGVITFPGTLDDV----DAAR  Mycobacterium lepraeLQ.pro
760  ----------------------------------TARIGVVTFPGTLDDV----DAAR  Mycobacterium tuberculosisLQ.pro
764  ----------------------------------TAKIGVITFPGTLDDV----DAQR  Corynebacterium ammoniagenesLQ.pro
611  ----------------------------------RACVVVYPGSNCDR----D---  Thermotoga maritimaLQ.pro
710  VVN---LDVEEMKKRYYEA----------------FP----------------  Methanococcus jannaschiiLQ.pro
710  ----------------------------------MPRFAVIVFPGTNCDF----ETVE  Pyrococcus abyssiLQ.pro
714  ----------------------------------MIAIIKFPGTTCE---------  Sulfolobus solfataricusL-Q.pro
780  ------------------------------------LRMEGTNCED----ETV-  Archaeoglobus fulgidusLQ.pro
749  ----------------------------------.MKFAVCVFPGSNCDY----DTYY  aquifexLQ.pro
790  EEREMTSPLSLVISAFARVEDVRHTITPQLSTED--NALLLIDLG--KGNNALGATALAQ  Escherichia coliL-1.pro
790  EQREMTSPLSLVISAFARVEDVRHTLTPQLSTED--NALLLIDLG--KGHNALGATALAQ  Salmonella typhimuriumL-1.pro
791  EQKSVTAPLSLVISAFARVEDVRKTLTPQLRTDKGFSSLLLIDLG--EGHNRLGATALAQ  Haemophilus influenzaeL1-1.pro
789  EKKSVVSPLSLIISAFAPVQDVRKTVTPELKNVED-SVLLFVDLG--FGKARMGGSAFGQ  Neisseria meningitidisL-1.pro
832  --KEVTAPLSLNITAFAPVFNTSKTWTPLLNRNTDDSVLVLVDLSAKQETKSLGASALLQ  Saccharomyces cerevisiaeL-1.pro
873  DGEVVKAPGNLVISAYVTCPDITKTVTPDLKLGGDDG--ILLHVDLAKGKRRLGGSALAQ  Arabidopsis thalianaL-1.pro
821  GGETIKSPGTLVISTYAPCPDVRLKVTPDLKGPGAGSKTSLLWINLENSAR-LGGSALAQ  Drosophila melanogasterL-1.pro
```

```
766  A-----------------------------------------------------------V  Bacillus subtilisLQ.pro
763  A-----------------------------------------------------------I  Lactococcus lactisLQ.pro
868  EAKNSNA-------------------------SLSKISIKLLRNNEAFQ------  Campylobacter jejuniLQ.pro
354  S-------------------------------------------------------  Staphylococcus aureusLQ.pro
958  GATMASADFGKDAEEKRPTVQVG--------DPFSEKLLIEACLELMASDAIVAIQDMGA  Zymomonas mobilisLQ.pro
816  A-----------------------------------------------------------A  Deinococcus radioduransLQ.pro
780  A-------------------------------------------------------  Mycobacterium lepraeLQ.pro
780  A-------------------------------------------------------  Mycobacterium tuberculosisLQ.pro
784  A-------------------------------------------------------  Corynebacterium ammoniagenesLQ.pro
626  ---------------------------------AYH-------------------  Thermotoga maritimaLQ.pro
728  ----------------------------KMMGEL-----.-----------------  Methanococcus jannaschiiLQ.pro
730  A-------------------------------------------------------  Pyrococcus abyssiLQ.pro
727  ----------------------------TDVYK-----------------------  Sulfolobus solfataricusL-Q.pro
793  ---KAFRSLG----VEAEAVHIK--------QF-----------------------  Archaeoglobus fulgidusLQ.pro
769  V-----------------------------------------------------I  aquifexLQ.pro
846  VFRQLGDKPADVRDVAQLKGFYDAIQALVAQRK--LLAYHDRSDGGLLVTLAEMAFAGHC  Escherichia coliL-1.pro
846  VYRQLGDKPADVRDVAQLKGFYDAMQALVAARK--LLAWHDRSDGGLLVTLAEMAFAGHC  Salmonella typhimuriumL-1.pro
849  VYKQLGDKPADVVKVQRLKDFYNAMQTLVAEDK--LLAYHDRSDGGLITTLAEMAFAGHC  Haemophilus influenzaeL1-1.pro
846  VYNNMSGDAPDLDDTSRLKAFYNVIQQLVAEDK--LLAYHDRSDGGLFAVLVEMAFAGRC  Neisseria meningitidisL-1.pro
890  VYNQVGNKSPTVYDNAILKGFLESLIQLHQQKEDIVLAYHDRSDGGLLITLLEMAFASRC  Saccharomyces cerevisiaeL-1.pro
931  VFGQIGNDCPDLDDVPYLKNVFDGVQALIAE--NLVSAGHDISDGGLVVTALEMAFAGNK  Arabidopsis thalianaL-1.pro
880  AYAQQGKDTPNLTRSDVLGKAFAVTQSLLGD--GLIQAGHDVSDGGLLVCVLEMAIGGLS  Drosophila melanogasterL-1.pro
```

```
768   KDELGHEVEYVWHEE-TSLDG----------------------------FDGV-----  Bacillus subtilisLQ.pro
765   RDVMGAEAEFVWHDE-KSLAG----------------------------FDGV-----  Lactococcus lactisLQ.pro
892   --EADINLNPIKFENTLFNKDFSHLVASSLEIKKVKASYFDDANI------IMMIELNAT  Campylobacter jejuniLQ.pro
355   GVEAEYVDYRETSLS-GFDGV----------------------------LIPG-----  Staphylococcus aureusLQ.pro
1010  AGLTSSAVEMASKGEVGIELDMDMVPCREEGMTPYEMMLSESQER-----MLMVLKPGR  Zymomonas mobilisLQ.pro
818   RLLLDDGAQFVWHTETALPEG----------------------------TELV-----  Deinococcus radioduransLQ.pro
781   ARHVGAEAVSLWHA-DADLKG----------------------------VDAV-----  Mycobacterium lepraeLQ.pro
781   ARQVGAEVVSLWHA-DADLKG----------------------------VDAV-----  Mycobacterium tuberculosisLQ.pro
785   VRLAGAEAVSLWHA-DTDLKG----------------------------VDAV-----  Corynebacterium ammoniagenesLQ.pro
629   ----------ALEINGFEPS---YVGLDDKLD----------D------YELI-----  Thermotoga maritimaLQ.pro
736   ----------------------------------------------------------  Methanococcus jannaschiiLQ.pro
731   IRKAGGEAERVW--YKESIRE----------------------------YDGV-----  Pyrococcus abyssiLQ.pro
732   ----------------ALIEAGVPTVIVKYKDFDPD------R------YNGV-----  Sulfolobus solfataricusL-Q.pro
811   ------YSDMIRFEEQRSVFD----------------------------YQCL-----  Archaeoglobus fulgidusLQ.pro
771   RDILEKDVEFVYWEEK-NLSK----------------------------YDVV-----  aquifexLQ.pro
904   GIDADIATLGDDRLAA----------------------------LFNEELGAVIQVRAADR  Escherichia coliL-1.pro
904   GVQVDIAALGDDHLAA----------------------------LFNEELGGVIQVRAEDR  Salmonella typhimuriumL-1.pro
907   GVEVDISALGDNDLAV----------------------------LFNEELGAVIQVADSQL  Haemophilus influenzaeL1-1.pro
904   GLDIDLNLLLAQTFITNHTALSQSLRTEEVKALAEWQETIARTLFNEELGAVIQVRKQDV  Neisseria meningitidisL-1.pro
950   GLEINID---GGDLESQLTNL----------------------FNEELGAVFQISAKNL  Saccharomyces cerevisiaeL-1.pro
989   GINLDLASNG-----ISL--FET--------------------LFSEELGLVLEISKTNL  Arabidopsis thalianaL-1.pro
938   GLRVDLSEPL-----AKLKNFDKSVEKLNRPELA--------VLFAEECGWVEVLDTDL  Drosophila melanogasterL-1.pro
```

```
792   ----------LI PGGF------------------------------SYGD-------  Bacillus subtilisLQ.pro
789   ----------LI PGGF------------------------------SYGD-------  Lactococcus lactisLQ.pro
944   NTN---LKS-FFI EGI QKQGI ENLKGDFNASSAFY---YAI LPLSKTNFEFS-------Y  Campylobacter jejuniLQ.pro
379   ----------GFSFGD------------------------------YLRS-------  Staphylococcus aureusLQ.pro
1064  EAE---AEAI FKKWELDFAI I GRVTDSKHMVLTWKGDI VCDI PLAPLADNAPCYDRPW-V  Zymomonas mobilisLQ.pro
843   ----------FLPGGF------------------------------SYGD-------  Deinococcus radioduransLQ.pro
805   ----------VVPGGF------------------------------SYGD-------  Mycobacterium lepraeLQ.pro
805   ----------VVPGGF------------------------------SYGD-------  Mycobacterium tuberculosisLQ.pro
809   ----------VVPGGF------------------------------SYGD-------  Corynebacterium ammoniagenesLQ.pro
653   ----------I LPGGF------------------------------SYGD-------  Thermotoga maritimaLQ.pro
736   --------. . MTK---------------VRVYVAYK------------------------  Methanococcus jannaschiiLQ.pro
754   ----------VLPGGF------------------------------SYAD-------  Pyrococcus abyssiLQ.pro
757   ----------I LPGGF------------------------------SFGD-------  Sulfolobus solfataricusL-Q.pro
830   ----------VFPGGF------------------------------SAGD-------  Archaeoglobus fulgidusLQ.pro
795   ----------VLPGGF------------------------------SFGD-------  aquifexLQ.pro
937   EAVESVLAQHGLA-DCVHYVGQAV---SGDRF-VI TANGQTVFSESRTTLRVWWAETTWQ  Escherichia coliL-1.pro
937   DAVEALLAQYGLA-DCVHYLGQAL---AGDRF-VI TANDQTVFSESRTTLRVWWAETTWQ  Salmonella typhimuriumL-1.pro
940   ESVREVLKAHNLL-GI I HQLGTVT---ADDRF-EI SRGSHKLFSEKRSELRSI WAELTYQ  Haemophilus influenzaeL1-1.pro
964   ADI I NLFYQQQLH-HNVFEI GTLT---DENTL-I I RDGQTHLI SDNLI KLQQTWQETSHQ  Neisseria meningitidisL-1.pro
984   SKFEKI LNENGVAKEYI SI VGKPSFQSQEI KI-I NSTTNDVI YANSRSELEQTWSKTSYE  Saccharomyces cerevisiaeL-1.pro
1022  DAVMEKLRAFDVTAEI I GNVTDSPL------I-EVKVDGI THLSEKTSFLRDMWEDTSFQ  Arabidopsis thalianaL-1.pro
985   ERVRSTYEKAGVPNYYLGVTEGFGLDS---RV-VLKNGKSELLDQPLRVLYKKWERTSYE  Drosophila melanogasterL-1.pro
```

```
802   ----YLRCGAIARFANIMPAV----------------------------------------  Bacillus subtilisLQ.pro
799   ----YLRCGAIASFANIMPEI----------------------------------------  Lactococcus lactisLQ.pro
990   FNKDSKKLENINLKLK--ISDDEIST-----------------------------------  Campylobacter jejuniLQ.pro
389   ----GAMASVAPIISEVKRLA----------------------------------------  Staphylococcus aureusLQ.pro
1120  ATPKAKALGAVPASGSITDNLVTLVGSPDLASRRWMWEQYDNMVGADTVQCPGGDAAVVR  Zymomonas mobilisLQ.pro
853   ----HLRSGAIAARSPIMNAV----------------------------------------  Deinococcus radioduransLQ.pro
815   ----YLRAGAIARLSPIMTEV----------------------------------------  Mycobacterium lepraeLQ.pro
815   ----YLRAGAIARFAPVMDEV----------------------------------------  Mycobacterium tuberculosisLQ.pro
819   ----YLRSGAISALAPVMQSV----------------------------------------  Corynebacterium ammoniagenesLQ.pro
663   ----YLRPGAVAAREKI--------------------------------------------  Thermotoga maritimaLQ.pro
749   ----------------------------------------------------------  Methanococcus jannaschiiLQ.pro
764   ----YLRAGAIAARQRIMEEV----------------------------------------  Pyrococcus abyssiLQ.pro
767   ----YLRAGSIAASTETMKKV----------------------------------------  Sulfolobus solfataricusL-Q.pro
840   ----------YIRAGAIFSARVKSVL-------RKDIEE---------------------  Archaeoglobus fulgidusLQ.pro
805   ----YLRPGALAARTPLAQAI----------------------------------------  aquifexLQ.pro
992   MQRLRDNPECADQEHQAKSNDADPGLNVKL----SFDINEDVAAPYIATGARP-KVAVLR  Escherichia coliL-1.pro
992   MQRLRDNPQCADQEHEAKANDTDPGLNVKL----SFDINEDIAAPYIATGARP-KVAVLR  Salmonella typhimuriumL-1.pro
995   MQRLRDNPECAEQEFEAKKNPDDKGLSAFL----TYDVNEDITAPFINKGVKP-TIAILR  Haemophilus influenzaeL1-1.pro
1019  IQRLRDNPACADSEFALIGDNGRSALFANL----KFDVNEDIAAPFINSGAKP-KIAILR  Neisseria meningitidisL-1.pro
1043  MQKLRDNPKTAEEEFASITDDRDPGLQYAL----TYNPADDMKIGLELSSQRP-KVAILR  Saccharomyces cerevisiaeL-1.pro
1075  LEKLQRLASCVEMEKEGLKFRHEPNW--KL----SFIPSSTNNNY-MSQGMIP-KVAVIR  Arabidopsis thalianaL-1.pro
1041  LEKLQANPECAEAEYNSLEYRQAPQY--R-------GPQNVQAELTLKRSSAPVRVAVLR  Drosophila melanogasterL-1.pro
```

```
819   ----------KQAAAEGKPVLGVCNGFQILQELGLLPGAMRRNK------  Bacillus subtilisLQ.pro
816   ----------KRLAKEGKPVFGTCNGFQILVESGLLPGVLIRNE------  Lactococcus lactisLQ.pro
1014  -QSDLNPVNKDLNI----------YKQY------  Campylobacter jejuniLQ.pro
406   ----------AEGKPVLGVCNGFQILTEIGLLPGALLHNDSHLF------  Staphylococcus aureusLQ.pro
1180  VHGTEKALAMSVDVTPRYCRADPEEGGKQA--VAECYRNITAV--GALPLASTDCLNFGN  Zymomonas mobilisLQ.pro
870   ----------KAHAEAGGYVLGVCNGFQVLTEAGLLPGALSRNK------  Deinococcus radioduransLQ.pro
832   ----------VDAVQRGMPVLGICNGFQVLCEAGLLPGALIRNV------  Mycobacterium lepraeLQ.pro
832   ----------VAAADRGMPVLGICNGFQVLCEAGLLPGALTRNV------  Mycobacterium tuberculosisLQ.pro
836   ----------VDRARQGMPTLGICNGFQILTEAGLLEGALTRNK------  Corynebacterium ammoniagenesLQ.pro
676   ----------AFEIAKAAERGKLIMGICNGFQILIEMGLLKGALLQNS------  Thermotoga maritimaLQ.pro
749   ----------ASILDPQAQAIKAATHKMGYQEVS----------  Methanococcus jannaschiiLQ.pro
781   ----------REFAEEGRPVLGICNGFQVLTEAGLLPGALRPNK------  Pyrococcus abyssiLQ.pro
784   ----------KQMAEDGKIVIGICNGFQILVESGLLKGALLPNL------  Sulfolobus solfataricusL-Q.pro
862   ----------FIKMGYPILGICNGFQVLVELG-ALPGFDE----DKPL  Archaeoglobus fulgidusLQ.pro
822   ----------YDFAQKGKYVIGICNGFQILTELGLLPGALLPNL------  aquifexLQ.pro
1047  EQGVNSHVEMAAAFHRAGFDAIDVHMSDLLTGRTGLEDFHALVACGGFSYGDVLGAGEGW  Escherichia coliL-1.pro
1047  EQGVNSHVEMAAAFHRAGFDAIDVHMSDLLGGRIGLGNFHALVACGGFSYGDVLGAGEGW  Salmonella typhimuriumL-1.pro
1050  EQGVNSHYEMAAAFDRAGFNAIDVHMSDLMIGRRNLAEFNAMVACGGFSYGDVLGAGGGW  Haemophilus influenzaeL1-1.pro
1074  EQGVNGQIEMAAAFTRAGFDAYDVHMSDLMAGRVHLADFKMLAACGGFSYGDVLGAGKGW  Neisseria meningitidisL-1.pro
1098  EQGVNGQMEMAWCFQQAGFNSVDVTMTDLLEGRFHLDDFIGLAACGGFSYGDVLGAGAGW  Saccharomyces cerevisiaeL-1.pro
1127  EEGSNGDREMSAAFYAAGFEPWDVTVSDLLAGDITLDQFRGIVFVGGFSYADVLDSAKGW  Arabidopsis thalianaL-1.pro
1092  EEGVNSEREMMACLLRANFEVHDVTMSDLLQGTASVSQYRGLIFPGGFSYADTLGSAKGW  Drosophila melanogasterL-1.pro
```

```
853   ---------------DLKFICRPVELIVQNDETLFTASYEKGESI--------------  Bacillus subtilisLQ.pro
850   ---------------GLKFVSKWQVLKVENNQSNFTTEYAKDALI--------------  Lactococcus lactisLQ.pro
1031  ------------------------------------ALWF-------LAVLFGV  Campylobacter jejuniLQ.pro
440   ---------------ISRNEELEIVNNQTAFTNLYEQGEKVIYPV-------------  Staphylococcus aureusLQ.pro
1236  PERPEIMGQIVGAIKGIGEACRALDMPIVSGNVSLYNETRQDDGSSLAILPTPTIGGVGL  Zymomonas mobilisLQ.pro
904   ---------------ELHFMCKPVHLRVENNATDFSRAYGPGQII-------------  Deinococcus radiodoransLQ.pro
866   ---------------GLHFICRDVWLRVISTSTAWTSRFEPETDL-------------  Mycobacterium lepraeLQ.pro
866   ---------------GLHFICRDVWLRVASTSTAWTSRFEPDADL-------------  Mycobacterium tuberculosisLQ.pro
870   ---------------GLHFHCVDTHLEVVNNNTAWTTEFEAGDKI-------------  Corynebacterium ammoniagenesLQ.pro
714   ---------------SGKFICKWVDLIVENNDTPFTNAFEKGEKI-------------  Thermotoga maritimaLQ.pro
773   ------------------------------------------------------  Methanococcus jannaschiiLQ.pro
815   ---------------IPRFICKWMYLKVNDTNTAFTQLYEEGEII-------------  Pyrococcus abyssiLQ.pro
818   ---------------KLRFISKWVYLKVIRADTILTK-GLDKKII-------------  Sulfolobus solfataricusL-Q.pro
895   AEKPEMALAMND---SSRFECRPTLLKKESEKCIFVK-NLKKDVVM-------------  Archaeoglobus fulgidusLQ.pro
856   ---------------NMRFVCKWVNLRVENERSAFTRKLEKGDVL-------------  aquifexLQ.pro
1107  AKSILFNDRVRD--EFAT--FFH-RPQTLALGVCNGCQMMSNLRELIPGSEL-------W  Escherichia coliL-1.pro
1107  AKSILFNHRVRD--EFET--FFH-RPQTLALGVCNGCQMMSNLRELIPGSEL-------W  Salmonella typhimuriumL-1.pro
1110  AKSILFNPKLHE--QFSQ--FFI-NPNTLTLGVCNGCQMISNLAEIIPGTEN-------W  Haemophilus influenzaeL1-1.pro
1134  AKSILFHPALRD--QFAA--FFT-DPNTLTLGVCNGCQMVSNLAEIIPGTAG-------W  Neisseria meningitidisL-1.pro
1158  AKSVLYHEGVRTSSKFSK--FFNERQDTFAFGACNGCQFLSRLKDIIPGCEN-------W  Saccharomyces cerevisiaeL-1.pro
1187  AASIRFNEPVL--SQFQE--FYK-RPDTFSLGICNGCQLMA-LLGWVPGPQVGGSLDTSQ  Arabidopsis thalianaL-1.pro
1152  AANILHNPRLL--PQFEA--FKR-RQDVFSLGICNGCQLMT-LIGFVGSAKSEVGAD-PD  Drosophila melanogasterL-1.pro
```

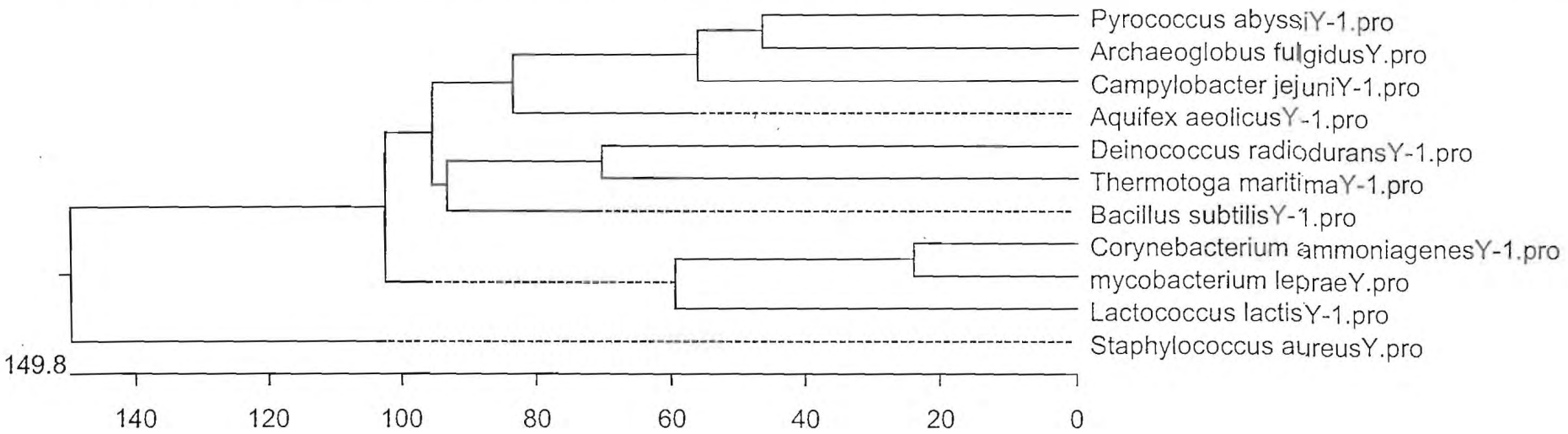
```
883   ---------TIPVAHG---EGNFYCDDETLATLKENNQIAF--------------T----  Bacillus subtilisLQ.pro
880   ---------NLPIAHG---EGQYVADEATLAELKENGQIVFT-------------Y----  Lactococcus lactisLQ.pro
1042  LFV-------------W---RKNYIIL------ALAVVCFALSFLVDTNT----------  Campylobacter jejuniLQ.pro
470   ---------AHGEGHY---YCTDEIYQQLKANNQIILKYVN--------------N----  Staphylococcus aureusLQ.pro
1296  LQDWRDSTTIAFKNTG---EEIYLVGNSGQGHLGQSIWLRKIAGREEGTAPSVDLA----  Zymomonas mobilisLQ.pro
934   ---------EIPIAHG---EGNYYADAATIAELEEGGRVV-F-------------R----  Deinococcus radioduransLQ.pro
896   ---------LVSLKSG---EGRYVASENVLDELDGEGRVVFR-----------------  Mycobacterium lepraeLQ.pro
896   ---------LVPLKSG---EGRYVAPEKVLDELEGEGRVVFR-----------------  Mycobacterium tuberculosisLQ.pro
900   ---------FVPAKHG---EGRFQAAPETIDKLEGEGQVVFR-----------------  Corynebacterium ammoniagenesLQ.pro
744   ---------RIPIAHG---FGRYVK--------IDDVNV--------------------  Thermotoga maritimaLQ.pro
773   ----------DLN---VG---------------KFFDFDF------DAEAELA----  Methanococcus jannaschiiLQ.pro
845   ---------RMPIAHA---EGNYYV------DDPSRIRIVFQYS-----DENGNLT----  Pyrococcus abyssiLQ.pro
847   ---------RMPIAHA---EGRYYVDDI----DYAKTHMVLQYC-----DENGNIS----  Sulfolobus solfataricusL-Q.pro
937   FPVAHAEGKVVFPSGK---------EDEYLERLTSNDQIVFRYVDEKGDYA----G----  Archaeoglobus fulgidusLQ.pro
886   ---------RIPIAHH---DGRYYVPEEELRKMEENGQILFRYC-----DEQGEVK----  aquifexLQ.pro
1155  PR-----FVRNTSDRFEARFSLVEVTQS------PSLLLQGMVGSQMPIAVSHGEGRVEV  Escherichia coliL-1.pro
1155  PR-----FVRNHSDRFEARFSLVEVTQS------PSLLLQGMVGSQMPIAVSHGEGRVEV  Salmonella typhimuriumL-1.pro
1158  PH-----FVRNKSERFEARVSLVKINEV------DSVWFAGMAGSHMPIAVSHGEGQVKF  Haemophilus influenzaeL1-1.pro
1182  PK-----FKRNLSEQFEARLSMVHVPKS------ASLILNEMQGSSLPVVVSHGEGRADF  Neisseria meningitidisL-1.pro
1209  PS-----FERNVSEQYEARVCMVQISQEKDNSSEESVFLNGMAGSKLPIAVAHGEGKATF  Saccharomyces cerevisiaeL-1.pro
1241  PR-----FVHNESGRFECRFTSVTIKDS------PSIMLKGMEGSTLGVWAAHGEGRAYF  Arabidopsis thalianaL-1.pro
1205  VA-----LLHNKSQRFECRWATVKIPSN------RSIMLGSMKDLVLGCWVAHGEGRFAF  Drosophila melanogasterL-1.pro
```

913	-YGSNIN---------------------------------GSVSDIAGVVNEKG------	Bacillus subtilisLQ.pro
911	-ADENPN---------------------------------GSVENIAGIINKEG------	Lactococcus lactisLQ.pro
1070	--------------------QNAIIKAG---SRAKIL------------------PTEPS	Campylobacter jejuniLQ.pro
500	-PNGSYD---------------------------------DIAGIVNEKGNVCG------	Staphylococcus aureusLQ.pro
1349	-QEKATGDFIRAMIQDGMLCAVHDISDG---GLAVALAEMALAGNIGATVEAHDKAIAEH	Zymomonas mobilisLQ.pro
964	-YADNPN---------------------------------GSLNDIAGIVNERG------	Deinococcus radioduransLQ.pro
926	-YHDNIN---------------------------------GSLRDIAGISSANG------	Mycobacterium lepraeLQ.pro
926	-YHDNVN---------------------------------GSLRDIAGICSANG------	Mycobacterium tuberculosisLQ.pro
930	-YTDNFN---------------------------------GSLNDIAGITNETG------	Corynebacterium ammoniagenesLQ.pro
763	-----------------VLRYVKDV--------------NGSDERIAGVLNESG------	Thermotoga maritimaLQ.pro
792	-KE------KATEIANE---	Methanococcus jannaschiiLQ.pro
878	-EEANPN---------------------------------GSVMNIAGVSNEGG------	Pyrococcus abyssiLQ.pro
882	-EDVNPN---------------------------------GSLLNIASIANEEG------	Sulfolobus solfataricusL-Q.pro
980	-YPWNPN---------------------------------GSFYNIAGICNATH------	Archaeoglobus fulgidusLQ.pro
925	-EEVNPN---------------------------------GSVSNIAGVMNKEG------	aquifexLQ.pro
1204	-RDAAHLAALESKGLVALRYVDNFGKVTETYPANPNGSPNGITAVTTESGRVTI------	Escherichia coliL-1.pro
1204	-RDDAHLAALESKGLVALRYVDNFGKVTETYPANPNGSPNGITAVTTENGRVTI------	Salmonella typhimuriumL-1.pro
1207	-KSVEQFAGLKAQGIIAAQYIDNNGSPTELYPANPNGSSEGITAITNLDGRVAI------	Haemophilus influenzaeL1-1.pro
1231	---ALHGGNISADLGIALQYVDGQNQITQTYPLNPNGSPQGIAGVTNADGRVTI------	Neisseria meningitidisL-1.pro
1264	SKSAEQLEKFEKDGLCCIRYVDNYGNVTERFPFNPNASTNGIAGIKSPNGRVLA------	Saccharomyces cerevisiaeL-1.pro
1290	-PDEGVLDHMLHSDLAPLRYCDDDGNVTEAYPFNLNGSPLGIAAICSPDGRHLA------	Arabidopsis thalianaL-1.pro
1254	-RDEKLISHLQSEQLVTLQYVDDVGKPTELYPLNPNGSPQGIAGLCSSDGRHLA------	Drosophila melanogasterL-1.pro

```
933   ------NVLGMMPHPERAVD-ELLGSADGL----------------KLFQSIVKNWR  Bacillus subtilisLQ.pro
931   ------NVLGMMPHPERAME-ELLGGADGR----------------KMFASLLKN-F  Lactococcus lactisLQ.pro
1089  T-------YFYTANADEKVEVLGKRQNYIKVLFSD---------GKIG-WVNK-------  Campylobacter jejuniLQ.pro
520   ------MMPHPERALETLLG-TDSGVKLFE----------------AMVKSWREQHV  Staphylococcus aureusLQ.pro
1405  AYYFGEDQGRYLVSSTNAVALVSAAEKAGIPVFRLGVTGGDAVVLNSQ-SVSLEKLRKSH  Zymomonas mobilisLQ.pro
984   ------NVLGMMPHPERAVE-LLLGSEDGK----------------GV-------FE  Deinococcus radioduransLQ.pro
946   ------RVVGMMPHPEHAIEVLTGPSDDGL----------------GLF----YSAL  Mycobacterium lepraeLQ.pro
946   ------RVVGLMPHPEHAIEALTGPSDDGL----------------GLF----YSAL  Mycobacterium tuberculosisLQ.pro
950   ------RIVGLMPHPEHAIETLTGPSTDGL----------------GLF----VSAI  Corynebacterium ammoniagenesLQ.pro
786   ------NVFGLMPHPERAVE-ELIGGEDG----------------------KKVF  Thermotoga maritimaLQ.pro
802   --------LLANPNMETYKVEILTEQS----------------------------  Methanococcus jannaschiiLQ.pro
898   ------NVLGMMPHPERASD-AFLGSEDGL----------------KVF------RS  Pyrococcus abyssiLQ.pro
902   ------NVIGMMPHP----------ERASFKLTSIDGT----VDG-LILLRR-----  Sulfolobus solfataricusL-Q.pro
1000  ------TVFGLMPH---PER--------AFFGYQVGRREGY----GDG-YCIFRSVV---  Archaeoglobus fulgidusLQ.pro
945   ------NVFGMMPHPERASE-DILGSHDGL----------------MLWYSLLSD  aquifexLQ.pro
1257  ------MMPHPERVFRTVSNS-------WHPE-----NWGEDGPWMRIFRNAR----KQL  Escherichia coliL-1.pro
1257  ------MMPHPERVFRTVANS-------WHPE-----NWGEDSPWMRIFRNAR----KQL  Salmonella typhimuriumL-1.pro
1260  ------MMPHPERVFRAVSNS-------WHPE-----NWTEDGAWMRLFRNAR----MVF  Haemophilus influenzaeL1-1.pro
1282  ------MMPHPERVYRAAQMS-------WKPE-----DWTELSGWYRLFAGAR----KAL  Neisseria meningitidisL-1.pro
1318  ------MMPHPERVCRLEANS-------WYPEG-KYEEWGGYGPWIRLFRSAR----RWV  Saccharomyces cerevisiaeL-1.pro
1343  ------MMPHPERCFLMWQFP-------WYPTSWDVEKAGP-SPWLKMFQNARDW--LES  Arabidopsis thalianaL-1.pro
1307  ------LMPHPERCSSMYQWP-------YVPSSFEVSPTQSESPWQIMFNNAYNWCVKSN  Drosophila melanogasterL-1.pro
```

967	ET- - HVTTA	Bacillus subtilisLQ.pro
964	LVTVKN.	Lactococcus lactisLQ.pro
1125	- - - DDLQKN	Campylobacter jejuniLQ.pro
554	.	Staphylococcus aureusLQ.pro
1464	EAFLPELMQ	Zymomonas mobilisLQ.pro
1011	SL- - KTVKK	Deinococcus radioduransLQ.pro
977	DSVLAS	Mycobacterium lepraeLQ.pro
977	DAVLTG	Mycobacterium tuberculosisLQ.pro
981	NA- - ITATV	Corynebacterium ammoniagenesLQ.pro
812	QSILNYLKR	Thermotoga maritimaLQ.pro
821	- - - - EGVKK	Methanococcus jannaschiiLQ.pro
926	MV- - EYAKR	Pyrococcus abyssiLQ.pro
933	- - - - LGEWA	Sulfolobus solfataricusL-Q.pro
1035	- - - DYLEKL	Archaeoglobus fulgidusLQ.pro
976		aquifexLQ.pro
1295	G	Escherichia coliL-1.pro
1295	G	Salmonella typhimuriumL-1.pro
1297		Haemophilus influenzaeL1-1.pro
1320	G	Neisseria meningitidisL-1.pro
1360	G	Saccharomyces cerevisiaeL-1.pro
1387	C	Arabidopsis thalianaL-1.pro
1354	Q	Drosophila melanogasterL-1.pro

Decoration 'Decoration #1': Shade (with bright yellow at 90% fill) residues that match the Consensus exactly.


Pyrococcus abyssiY-1.pro
Archaeoglobus fulgidusY.pro
Campylobacter jejuniY-1.pro
Aquifex aeolicusY-1.pro
Deinococcus radioduransY-1.pro
Thermotoga maritimaY-1.pro
Bacillus subtilisY-1.pro
Corynebacterium ammoniagenesY-1.pro
mycobacterium lepraeY.pro
Lactococcus lactisY-1.pro
Staphylococcus aureusY.pro
149.8
140
120
100
80
60
40
20
0

```
                MA--KVRVYVMLKEGVLDPQGQAVERALHSLGYNGVSDVRVGKVFELEVDA---EVAAEE  Majority
                         10        20        30        40        50        60
1               MR-WKVRVLVRLKEGLNDPEGRVIGKALKNLGYK-VEELKVPKCFEFVLES---EKPEEE  Pyrococcus abyssiY-1.pro
1               VY-------IELKEGVADPEGEATLKALRLLGFKRVKKVSTVKVFRIDIEARSREEAERE  Archaeoglobus fulgidusY.pro
1               M---EVIVNISLKNGVLDPQGKAVEKALHSLNFNSVKEVKIAKQIKISLDEKDEKLAKEQ  Campylobacter jejuniY-1.pro
1               MK--LVRVIIMPKEGLLDPQGRAVEEMLKENGFN-VSNVRVGKVVELEVSE------DTD  Aquifex aeolicusY-1.pro
1               MPHYQAKIYVTLRPSILDPQGRTVERALSHLNHDNVGSVRIGKYIELNLSGEKADVEAQL  Deinococcus radioduransY-1.pro
1               MPLFKFAIDVQYRSNVRDPRGETIERVLREEKGLPVKKLRLGKSIHLEVEAENKEKAYEI  Thermotoga maritimaY-1.pro
1               MY--KVKVYVSLKESVLDPQGSAVQHALHSMTYNEVQDVRIGKYMELTIEKSDRDLDV-L  Bacillus subtilisY-1.pro
1               MA--RVVVNVMPKAEILDPQGQAVVRALGRLGVNGVSDVRQGKRFEIEVDDS---VSAED  Corynebacterium ammoniagenesY-1.pro
1               MA--RAVVHVMLRAEILDPQGQAIAGALGRLGHTGISDVRQGKRFELEIDDT---VDDSE  mycobacterium lepraeY.pro
1               MT--KVRVYVAYKASILDPQAQAIKAATHKMGYQEVSDLNVGKFFDFDFDA-EAELAKEK  Lactococcus lactisY-1.pro
1               MK--TIELHITLQPQVLDTQGQTLTRAVHDLGYAQVNDIRVGKVLYMTVDEVSDEKVHNI  Staphylococcus aureusY.pro

                VKEMCEKLLANPVIEDYEYEIEEVE----K  Majority
                         70        80        90
56              VEEMCRRLLANPLIHTWEYTIEPV------K  Pyrococcus abyssiY-1.pro
54              IAEMCEKLLANPVIQKY                Archaeoglobus fulgidusY.pro
58              VKKMCEELLVNSVIEDYELIIEK-------E  Campylobacter jejuniY-1.pro
52              LRKLVEKYLINPLIEDYE--IEELSQ----K  Aquifex aeolicusY-1.pro
61              -KDIVENVLSNPIMEDARWELEEVQ        Deinococcus radioduransY-1.pro
61              VKKACEELLVNPVVEE--YEVREL         Thermotoga maritimaY-1.pro
58              VKEMCEKLLANTVIEDYRYEVEEVVA----Q  Bacillus subtilisY-1.pro
56              LDKVAASLLANTVIEDYE--VVGLEV----K  Corynebacterium ammoniagenesY-1.pro
56              LAMIAESLLANTVIEDWT--ITR-ES----Q  mycobacterium lepraeY.pro
58              ATEIANELLANPNMETYKVEILTEQSEGVKK  Lactococcus lactisY-1.pro
59              ITTLSEKLFANTVIEEYSYKVLDDEKENA.   Staphylococcus aureusY.pro
```

Decoration 'Decoration #1': Shade (with bright yellow at 90% fill) residues that match the Consensus exactly.

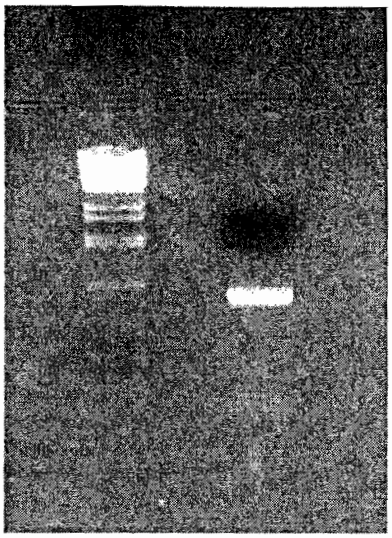


Figure 9. *PCR gel.* The lane on the left is a lambda BstEII marker. The lane on the right shows the size of the PCR product obtained with Staphylococcus aureus genomic DNA and the Sa*purQ* start NdeI and stop EcoRV primers.

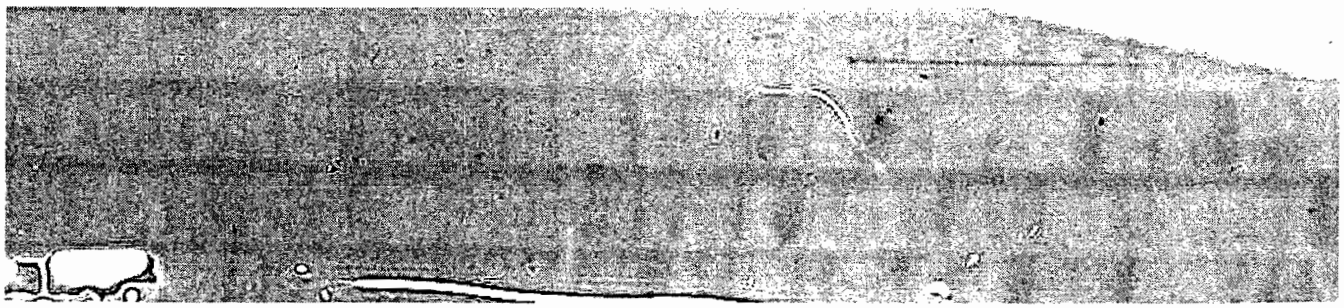


Figure 10. *Sequencing gel.* A sequencing reaction was done on the plasmid preparation of colony number 53. The sequence AAACGTCTTCAGC can be read this is identical to bases 213-227 of the *purQ* gene

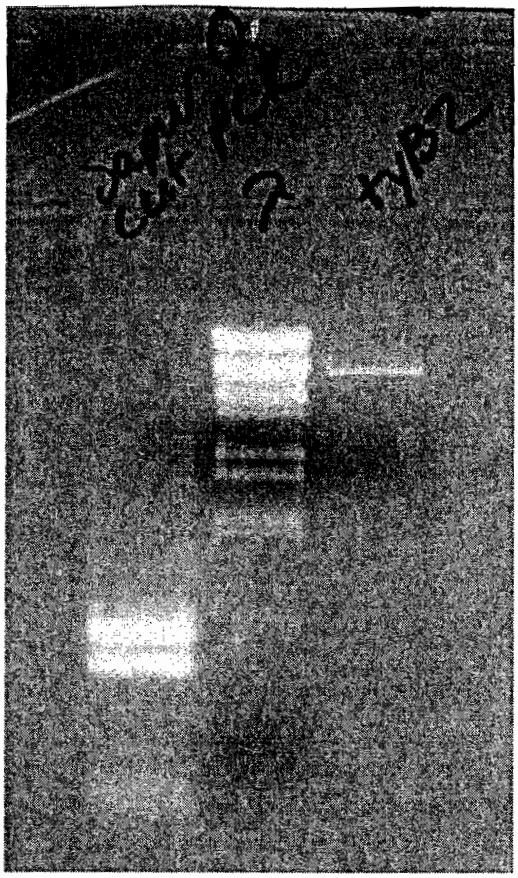


Figure 11. *Enzyme analysis on PCR.* The first lane is PCR product of genomic *S. aureus* DNA. It was cut with NdeI and EcoRV.

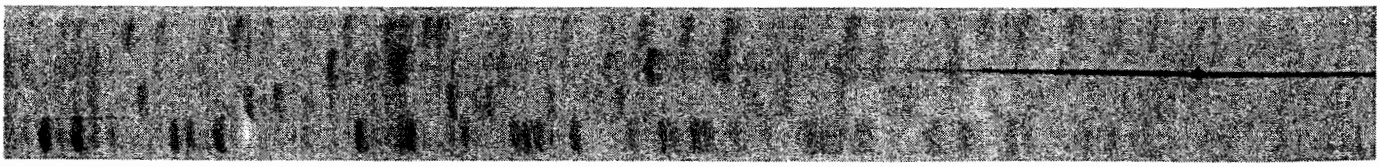


Figure 12. *Sequencing gel.* Sequencing reactions were run on PCR reactions to determine where the NdeI site is in the *purQ* gene.

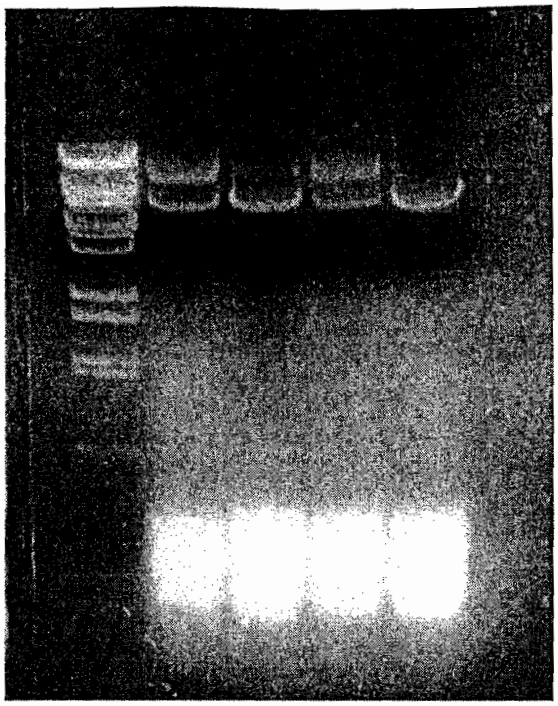


Figure 13. *Digest gel.* Phenol- Chloroform purified plasmid preparations of colonies 1, 21, 23, 27 were digested with XbaI and HindIII. This gel shows no cutting, which was probably due to the amount of RNA on the gel shown by the bright bands on the bottom.

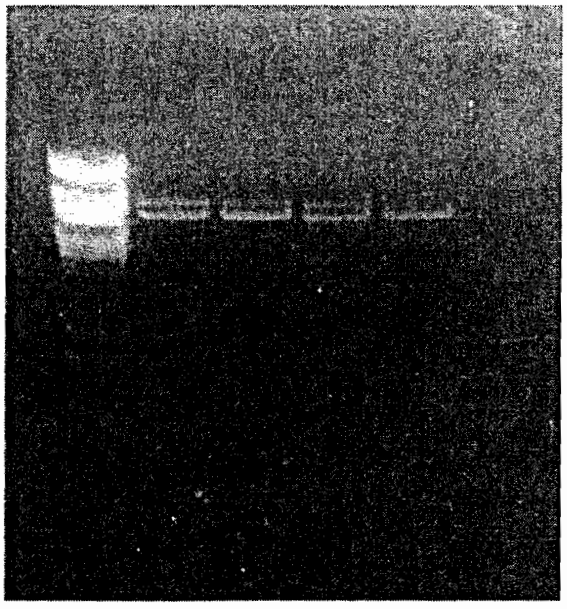


Figure 14. *Digest gel.* The samples run on the previous gel were treated with RNase and recut.

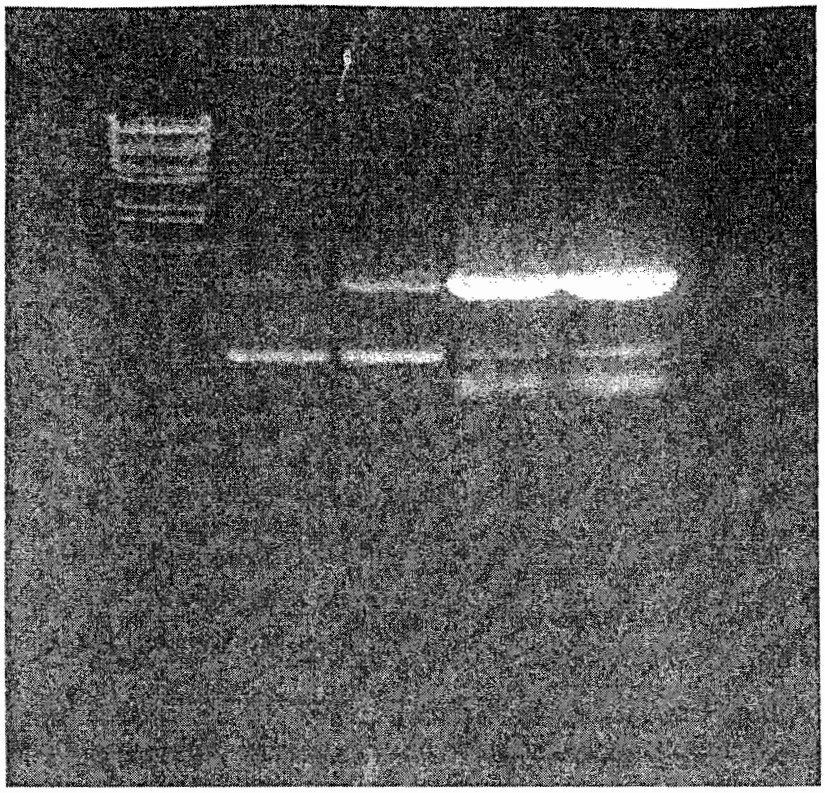


Figure 15. *PCR gel.* PCR reactions were run using the TYB2 and intien reverse primers of plasmid preparations 1, 21, 23, and 27. Bands of 1100bp can be seen along with bands of about 450bp.

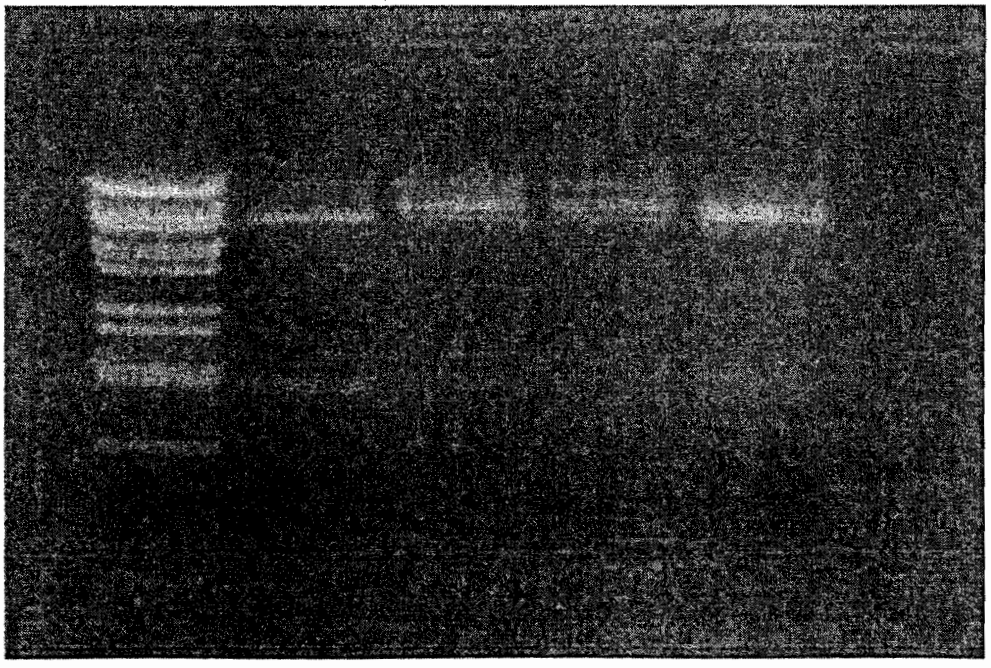


Figure 16. *Wizard™ plasmid preperation.* Plasmid preparations using the Wizard SV minipreps system. Each prep was cut with XbaI and HindIII. The lane order from left to right is Lambda marker then preps of colonies 1, 21, 23, and 27.

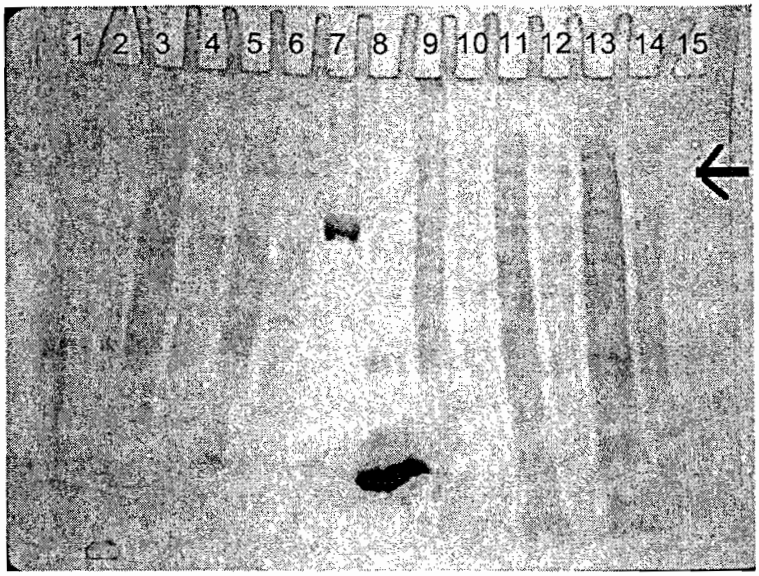


Figure 17. *SDS Page gel.* Samples from each expression temperature were run. Lane order is: Clarified 30 degrees 2:1, dilution, Clarified 25 degrees 2:1, dilution, Clarified 15 degrees 2:1, dilution, BSA, RNase, crude 30 degrees 2:1, dilution, crude 25 degrees 2:1, dilution, crude 15 degrees 2:1, dilution, and a blank sample.

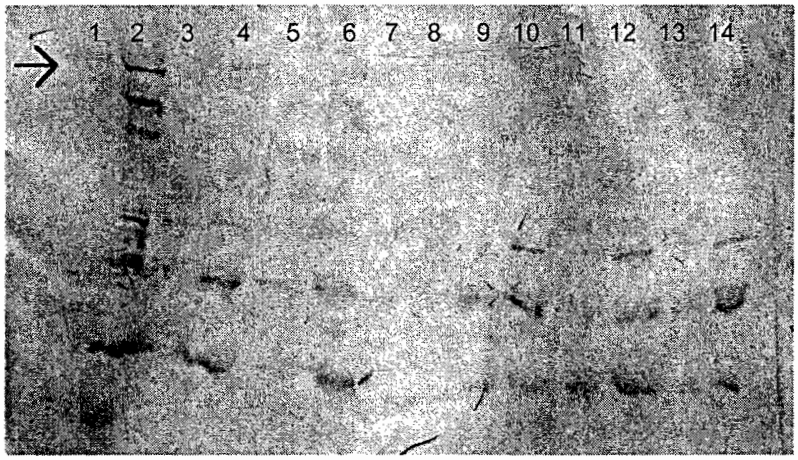


Figure 18. *Western Blot.* Lane order is the same as SDS page gel order. The arrow is pointing to the band showing expression of the target protein.

Abbreviations:

FGAR- 5'- phosphoribosyl N- formyglycinamide

FGAM- 5'- phosphoribosyl N- formyglycinamidine

BLAST- Basic local alignment search tool

PCR- polymerase chain reaction

dNTP/ ddNTP- deoxy and dideoxy nuleotides

CTAB- cetyltrimethylammonium bromide

IPTG- isopropyl- 1- thio- beta –D- galactoside

SDS- sodium dodecyl sulfate

PVDF- polyvinyl difluoride

PBS- phosphate buffered saline

NTB- nitro blue tetrazolium

BCIP- 5- bromo-4-chloro-3-indolyl phosphate

Work Cited

Bio- Rad Instruction manual. Prep- A – Gene purification systems. Catalog number 732- 6010, 732- 6011.

Del Sal, G., Man fioletti, G., Schneider, C. (1989). Biotechniques 7: 514- 520

Levy, S.B. (1998). The challenge of antibiotic resistance. Scientific American 398.

IMPACT T7 one- step protein purification system. Instruction manual. New Englang bioLabs, Inc.

Rosamond, J. and Allsop, A. (2000). Harnessing the power of the genome in the search for new antibiotics. Science287: 1973- 1976.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning A laboratory manual second edition. Cold Spring Harbor Laboratory Press. 3: 18.6918.74.

Sampei, G. and Mizobuchi, K. (1989). The organization of the *purL* gene encoding 5'-phosphoribosylglycinamide amidotransferase of *Escherichia coli*. J. Biological Chemistry 264: 21230- 21238.

Schendel, F.J., Mueller, E., Stubbe, J., Shiau, A. and Smith, J.M. (1989). Formylglycinamide ribonucleotide synthetase from *Escherichia coli.:* Cloning, sequencing, overproduction, isolation, and characterization. Biochemistry 28: 2459- 2471.

Zalkin, H. (1985). Glutamine amidotransferases. Meth. Enzymol. 113: 263- 264.

Zalkin, H. and Dixon, J.E. (1992). De novo purine nucleotide biosyntheis. Prog. Nucl. Acid Res. And Mol. Biol. 42: 259- 287.

All methods were taken from standard protocol manual in the lab or from instruction booklets sent with the product.