HAL
open science

Identification of six open reading frames in the Salmonella enterica subsp. enterica ser. Typhi viaB locus involved in Vi antigen production

Hervé Waxin, Isabelle Payant, Suzan Kolyva, Michel y Popoff

To cite this version:

Hervé Waxin, Isabelle Payant, Suzan Kolyva, Michel y Popoff. Identification of six open reading frames in the Salmonella enterica subsp. enterica ser. Typhi viaB locus involved in Vi antigen production. Research in Microbiology, 1993, 144 (5), pp.363-371. 10.1016/0923-2508(93)90193-6 . hal-01600906

HAL Id: hal-01600906

https://hal.science/hal-01600906

Submitted on 1 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

© INSTITUT PASTEUR/ELSEVIER
Paris 1993

Res. Microbiol.
1993, 144, 363-371

Identification of six open reading frames in the *Salmonella enterica* subsp. *enterica* ser. Typhi *viaB* locus involved in Vi antigen production

H. Waxin, I. Virlogeux, S. Kolyva and M.Y. Popoff (*)

Unité des Entérobactéries, Unité INSERM 199, Institut Pasteur, 75724 Paris Cedex 15

The Vi antigen of *Salmonella enterica* subsp. *enterica* ser. Typhi (hereafter referred to as Typhi) is a capsular polysaccharide consisting of a homopolymer of α-1,4 2-deoxy-2*N*-acetyl galacturonic acid (Felix and Pitt, 1934). Determinants of Vi antigen occupy two widely separated chromosomal loci, designated *viaA* and *viaB*. The *viaB* locus, specific to Vi-expressing strains, maps at 92 min on the chromosome of Typhi (Johnson *et al.*, 1965). Cloning and molecular analysis of this chromosomal region were reported previously (Hashimoto *et al.*, 1991; Kolyva *et al.*, 1992). Here we report on the nucleotide sequence of a 9.3-kb fragment located within the Typhi strain Ty2*viaB* locus.

The dideoxy chain-termination method, using modified T7 DNA polymerase ("Sequenase"; USB Corp.) and universal forward and synthetic primers, was employed after subcloning of appropriate restriction fragments into M13 derivatives (Messing and Vieira, 1982). All the ends of restriction fragments used overlapped one another. DNA sequencing was performed at least twice on both strands. Nucleotide sequence data were analysed using the Lipman-Pearson program (Lipman and Pearson, 1985). The sequence data have been submitted to the EMBL data library and were assigned accession no. X67785.

As shown in figure 1, examination of the DNA sequence revealed six open reading frames (ORF), termed: *tviA (*Typhi *Vi), tviB, tviC, tviD, tviE* and *tviF*. All were transcribed in the same orientation. Their description is reported in table I. Significant homology was detected between TviB protein and AlgD (Deretic *et al.*, 1987), a GDP-mannose dehydrogenase of *Pseudomonas aeruginosa* (26 % identity and 71 % similarity in 288 amino acids overlap) and between TviC protein and StrE (Piscowotzki *et al.*, 1991), a TDP-glucose dehydratase of *Streptomyces griseus* (28 % identity and 68 % similarity in 326 amino acids overlap). A lipoprotein signal sequence (Hussain *et al.*, 1982) with a potential cleavage site for signal peptidase II (see fig. 1) was found in the TviF sequence. TviF protein shared significant homology to BexD (Kroll *et al.*, 1990) and CtrA (Frosch *et al.*, 1991) proteins (27 % identity and 72 % similarity in 280 amino acids overlap) which are involved in capsule export by *Haemophilus influenzae* and *Neisseria meningitidis*, respectively.

Studies designed to determine whether TviB and TviC proteins are involved in the biosynthesis, and TviF in the biogenesis, of the Vi antigen are presently under way in our laboratory.

Key-words: Vi antigen, ORF, *Salmonella* ser. Typhi; *viaB* locus, Nucleotides, Sequencing; Note.

Submitted May 7, 1993, accepted May 14, 1993.

(*) Corresponding author.

Table I. Description of the six open reading frames identified within the 9.3-kb sequenced segment of the *viaB* locus.

Open reading frames	Location (nucleotide) from:	to:	Expected protein aa	mm
tviA	770	1306	179	21099
tviB	1554	2828	425	47612
tviC	2834	3877	348	39478
tviD	4285	6396	704	80402
tviE	6399	8132	578	64902
tviF	8180	9244	355	38877

aa = number of amino acids; mm = predicted molecular mass in dalton.

Identification de six phases ouvertes de lecture dans le locus *viaB* impliqué dans la production de l'antigène Vi chez *Salmonella enterica* subsp. *enterica* ser. Typhi

L'antigène Vi de *Salmonella enterica* subsp. *enterica* ser. Typhi est un polyoside capsulaire dont l'expression est contrôlée par les loci chromosomiques *viaA* et *viaB*. La séquence d'un fragment (9,3 kb) du locus *viaB*, qui est spécifique des souches produisant l'antigène Vi, a été déterminée. Six phases ouvertes de lecture (*tviA*, *tviB*, *tviC*, *tviD*, *tviE* et *tviF*) codant respectivement pour des polypeptides de 21.099, 47.612, 39.478, 80.402, 64.902 et 38.877 daltons ont été mises en évidence. Sur la base des résultats de comparaison de séquence, les protéines TviB, TviC et TviF pourraient être impliquées dans la biosynthèse et la biogénèse de l'antigène Vi.

Mots-clés: Antigène Vi, ORF, *Salmonella* ser. Typhi; Locus *viaB*, Nucléotides, Séquençage; Note.

References

Deretic, V., Gill, J.F. & Chakrabarty, A.M. (1987), *Pseudomonas aeruginosa* infections in cystic fibrosis: nucleotide sequence and transcriptional regulation of the *algD* gene. *Nucl. Acids Res.*, 15, 4567-4581.

Felix, A. & Pitt, R.M. (1934), A new antigen of *B. typhosus*. *Lancet*, 227, 186-191.

Frosch, M., Edwards, U., Krausse, B. & Weisgerber, C. (1991), Evidence for a common molecular origin of the capsule gene loci in Gram-negative bacteria expressing group II capsular polysaccharide. *Mol. Microbiol.*, 5, 1251-1263.

Hashimoto, Y., Ezaki, T., Li, N. & Yamamoto, H. (1991), Molecular cloning of the *viaB* region of *Salmonella typhi*. *FEMS Microbiol. Letters*, 90, 53-56.

Hussain, M., Ichihara, S. & Mizushima, S. (1982), Mechanism of signal peptide cleavage in the biosynthesis of the major lipoprotein of the *Escherichia coli* outer membrane. *J. Biol. Chem.*, 257, 5177-5182.

Johnson, E.M., Krauskopf, B. & Baron, L.S. (1965), Genetic mapping of Vi and somatic antigenic determinants in *Salmonella*. *J. Bact.*, 90, 302-308.

Kolyva, S., Waxin, H. & Popoff, M.Y. (1992), The Vi antigen of *Salmonella typhi*: molecular analysis of the *viaB* locus. *J. gen. Microbiol.*, 138, 297-304.

Kroll, J.S., Loynds, B., Brophy, L.N. & Moxon, E.R. (1990), The *bex* locus in encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsule polysaccharide export. *Mol. Microbiol.*, 4, 1853-1862.

Lipman, D.J. & Pearson, W.R. (1985), Rapid and sensitive protein similarity searches. *Science*, 227, 1435-1441.

Messing, J. & Vieira, J. (1982), A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene*, 19, 269-276.

Pissowotzki, K., Mansouri, K. & Piepersberg, W. (1991), Genetics of streptomycin production in *Streptomyces griseus*: molecular structure and putative function of genes *strELMB2N*. *Mol. gen. Genet.*, 231, 113-123.

Fig. 1. Nucleotide sequence of a 9.3-kb fragment of the *Salmonella* ser. Typhi Ty2*viaB* locus.

The putative Shine-Dalgarno sequences are underlined. The cleavage site of the lipoprotein signal peptide at amino acid position 17 of TviF is indicated with a star. These sequence data were assigned accession no. X67785 in the EMBL data library.

```
                                        30                                        60
GAA TTC TGC AAA CCA GCC CTG TAC CAT CAA GTT CAT ATC CTC AGC GGA GGA AAT GCG TGA
                                        90                                       120
AAT CCA GCG CAA CAC ACG GAG TAT CAC CCA CTA CCC AGG CAT CAT GAT TCT CAG AAA CCT
                                       150                                       180
CAA ATG CTT CGC CGG CAT GAA TAA GGT ATT CAT TTC CGA AGC AGT CAC GCA CCA TCA TTT
                                       210                                       240
CAC CGG AAA TTA TAA AAC CAG TAT GAC GTT CTG ACG GTT TTC CGC TCA TTA TTA ATG GGT
                                       270                                       300
GAC TCT AAT GCA AAC ATC TAG CGA GAA AAT ATT TTG TAG ACA TCG TCA ACG GAA AAG TTA
                                       330                                       360
ACT AAC AAT AGT CAT TAT CCA TAT CTT TAA TTT GAA ATG GAA ACA GGG CAA AGA GTA ATT
                                       390                                       420
AAA ATA CAG CCT GTG ATT CTG TAC AGT TAA GGC TTA ACC CCA AAG GTG CAA ACA ATT AAA
                                       450                                       480
TTT GTA ACA GAT TAT TTC AAA TAC GAT TAG GAA TAT TCT TAT TTT GAG GAA TGG CTT AGC
                                       510                                       540
ATT TTT TAT GGT CGC CAG GAT TAA AAA ATA CGC CCT ATC CAA ACA ATT TAG CAT AAT TTA
                                       570                                       600
TCA TTC GAT TTT CTA GAC TAA ATA AGA TTT TTT GAT AGG TAC AAA CAA TGA ATT GTG CAG
                                       630                                       660
GTT TGA CAT CAA TTC TCT GTT GTG TAA AAA TCC CGT TTA GGC CGT TAG TAC TAT TAA AAT
                                       690                                       720
TAG GGT AAT AAT TTT ATT GTT AGT TAA TTG TTA ACA GGA GCA AAG AAT TAG ATA TTG CTT
                                       750                      tviA             781
GTA CCT TTT CTA TGG ATG AAA TTA GGT TAT TTC AGC ATA AGGAG ACTTC ATG AGG TTT CAT
                                                                Met Arg Phe His
                                       811                                       841
CAT TTC TGG CCT CCG AAT GAT ATC TAT TTC GGG GTT GGA GCT GCT GGC ATT ATT GAA GAA
His Phe Trp Pro Pro Asn Asp Ile Tyr Phe Gly Val Gly Ala Ala Gly Ile Ile Glu Glu
                                       871                                       901
GTG TCA CTG ATA ACA AAT GAC AGA AAT TAT TTG TTT GTG AAC CTA AAT CGC TAC AGC CTG
Val Ser Leu Ile Thr Asn Asp Arg Asn Tyr Leu Phe Val Asn Leu Asn Arg Tyr Ser Leu
                                       931                                       961
TTA AAT GCC CTG AAT TTT TTC ACG CGA ATG AGT GAT ATT AAT AAA ATA ATC GTT ATC ATT
Leu Asn Ala Leu Asn Phe Phe Thr Arg Met Ser Asp Ile Asn Lys Ile Ile Val Ile Ile
                                       991                                      1021
TCA AGT TCG CGA CTA ATG CCC CTT GCA CGT TTT TGG TTG ACA GAG TGC AAA AAT GTT ATT
Ser Ser Ser Arg Leu Met Pro Leu Ala Arg Phe Trp Leu Thr Glu Cys Lys Asn Val Ile
                                      1051                                      1081
GCT GTT TTC GAT GCG GCA ACA TCA GTC CAG GAT ATT ATC AGA AAT GTC AGT CAA CAC CAA
Ala Val Phe Asp Ala Ala Thr Ser Val Gln Asp Ile Ile Arg Asn Val Ser Gln His Gln
                                      1111                                      1141
AGT GGT GAA AAG ATC TTG ACG GAG CAG AGA GAT TAT CGT TTC AGA ATT AAC CGT AAG GAT
Ser Gly Glu Lys Ile Leu Thr Glu Gln Arg Asp Tyr Arg Phe Arg Ile Asn Arg Lys Asp
                                      1171                                      1201
ATA GTA AAG ATG AAA TAT TTC CTT TCG GAA AGT GGT ATG GAA GAG CTT CAG GAT AGA TTT
Ile Val Lys Met Lys Tyr Phe Leu Ser Glu Ser Gly Met Glu Glu Leu Gln Asp Arg Phe
                                      1231                                      1261
ATG AAC TCA TCA TCG ACT ATG TAT CGC TGG AGA AAA GAA TTG GCA GTA AAA TTT GGA GTA
Met Asn Ser Ser Ser Thr Met Tyr Arg Trp Arg Lys Glu Leu Ala Val Lys Phe Gly Val
                                      1291                                      1320
CGT GAG CCG CGC TAT CTG TTA TTG CCG GAT TCA GTT ACT TTA CTG TAA TGC GGT AAT  TT
Arg Glu Pro Arg Tyr Leu Leu Leu Pro Asp Ser Val Thr Leu Leu
                                      1350                                      1380
TTA TTG AGT AAA ACA CGG ACA AGT ATT TCG TTT CAG CAC AAA ATT ATT TTC GTT ACT CAT
                                      1410                                      1440
TGG CGT TAA TAC ATA TAT TCT CAG CGA CTT CTG TTC TAT TCA AGT AAG AAA GGG GTA CGG
                                      1470                                      1500
TTA TAC GTT TTC ATT AAC CAT ACT GGC TGC TAC GGC CAG GGG CGG TAG CGT ATC TGA ATA
                                      1530                            tviB      1562
AAC ACC TAG AAT TAA CTT TGT AAA TAT AAA ATT TTA GTA AAGGA TTAA TAAGA GTG TTC GGT
                                                                     Val Phe Gly
                                      1592                                      1622
ATA GAC GAG GTA AAA ATC GCG ATT ATT GGG CTG GGA TAT GTT GGG CTT CCT CTG GCA GTT
Ile Asp Glu Val Lys Ile Ala Ile Ile Gly Leu Gly Tyr Val Gly Leu Pro Leu Ala Val
```

1652 1682
GAA TTT GGC AAA TCT CGT CAG GTT GTT GGC TTC GAC GTT AAT AAA AAG CGT ATT CTT GAA
Glu Phe Gly Lys Ser Arg Gln Val Val Gly Phe Asp Val Asn Lys Lys Arg Ile Leu Glu
1712 1742
TTA AAG AAT GGG GTG GAT GTC AAT CTG GAA ACC ACT GAA GAA GAA TTA CGT GAG GCT CGT
Leu Lys Asn Gly Val Asp Val Asn Leu Glu Thr Thr Glu Glu Glu Leu Arg Glu Ala Arg
1772 1802
TAT CTG AAA TTT ACT TCC GAG ATT GAG AAG ATC AAA GAA TGT AAT TTT TAC ATC ATC ACC
Tyr Leu Lys Phe Thr Ser Glu Ile Glu Lys Ile Lys Glu Cys Asn Phe Tyr Ile Ile Thr
1832 1862
GTC CCG ACG CCG ATA AAT ACC TAC AGG CAA CCA GAC CTC ACC CCA CTA ATC AAG GCC AGT
Val Pro Thr Pro Ile Asn Thr Tyr Lys Gln Pro Asp Leu Thr Pro Leu Ile Lys Ala Ser
1892 1922
GAA ACC GTT GGT ACA GTG CTG AAT CGG GGA GAT ATT GTG GTA TAT GAA TCT ACG GTA TAT
Glu Thr Val Gly Thr Val Leu Asn Arg Gly Asp Ile Val Val Tyr Glu Ser Thr Val Tyr
1952 1982
CCG GGA TGT ACC GAA GAA GAA TGC GTG CCG ATC CTT GCT CGT ATG TCC GGA ATG ACT TTC
Pro Gly Cys Thr Glu Glu Glu Cys Val Pro Ile Leu Ala Arg Met Ser Gly Met Thr Phe
2012 2042
AAC CAG GAT TTC TAT GTC GGT TAT AGC CCG GAA AGG ATC AAT CCC GGT GAT AAA AAG CAC
Asn Gln Asp Phe Tyr Val Gly Tyr Ser Pro Glu Arg Ile Asn Pro Gly Asp Lys Lys His
2072 2102
CGT TTA ACC AAC ATC AAG AAA ATC ACC TCC GGT TCA ACC GCA CAG ATC GCC GAA CTT ATC
Arg Leu Thr Asn Ile Lys Lys Ile Thr Ser Gly Ser Thr Ala Gln Ile Ala Glu Leu Ile
2132 2162
GAT GAA GTA TAT CAG CAG ATC ATC AGC GCA GGT ACA TAT AAA GCA GAG AGC ATC AAA GTT
Asp Glu Val Tyr Gln Gln Ile Ile Ser Ala Gly Thr Tyr Lys Ala Glu Ser Ile Lys Val
2192 2222
GCT GAG GCA GCG AAG GTG ATT GAA AAT ACG CAA CGC GAT CTG AAT ATT GCC TTG GTC AAT
Ala Glu Ala Ala Lys Val Ile Glu Asn Thr Gln Arg Asp Leu Asn Ile Ala Leu Val Asn
2252 2282
GAG CTG GCG ATT ATT TTT AAT CGT TTA AAT ATC GAT ACT GAA GCC GTG CTA CGT GCC GCT
Glu Leu Ala Ile Ile Phe Asn Arg Leu Asn Ile Asp Thr Glu Ala Val Leu Arg Ala Ala
2312 2342
GGC AGC AAA TGG AAT TTC CTG CCA TTC CGT CCG GGA CTG GTC GGT GGT CAC TGT ATT GGC
Gly Ser Lys Trp Asn Phe Leu Pro Phe Arg Pro Gly Leu Val Gly Gly His Cys Ile Gly
2372 2402
GTA GAT CCC TAT TAT CTG ACA CAT AAA TCT CAG GGC ATT GGC TAT TAT CCA GAA ATC ATA
Val Asp Pro Tyr Tyr Leu Thr His Lys Ser Gln Gly Ile Gly Tyr Tyr Pro Glu Ile Ile
2432 2462
CTT GCA GGA CGC CGC CTG AAC GAC AAC ATG GGC AAC TAT GTC TCC GAG CAG TTG ATC AAA
Leu Ala Gly Arg Arg Leu Asn Asp Asn Met Gly Asn Tyr Val Ser Glu Gln Leu Ile Lys
2492 2522
GCA ATG ATC AAA AAA GGA ATT AAC GTT GAG GGT TCC AGC GTG CTG ATT CTC GGC TTT ACC
Ala Met Ile Lys Lys Gly Ile Asn Val Glu Gly Ser Ser Val Leu Ile Leu Gly Phe Thr
2552 2582
TTT AAA GAA AAC TGT CCG GAC ATC AGA AAT ACA CGC ATT ATT GAT GTG GTA AAG GAA CTC
Phe Lys Glu Asn Cys Pro Asp Ile Arg Asn Thr Arg Ile Ile Asp Val Val Lys Glu Leu
2612 2642
GGT AAA TAT AGT TGT AAA GTG GAT ATT TTT GAT CCA TGG GTG GAT GCC GAA GAG GTA AGA
Gly Lys Tyr Ser Cys Lys Val Asp Ile Phe Asp Pro Trp Val Asp Ala Glu Glu Val Arg
2672 2702
CGA GAG TAT GGC ATT ATC CCG GTA TCG GAA GTC AAA TCA AGC CAC TAC GAT GCG ATC ATT
Arg Glu Tyr Gly Ile Ile Pro Val Ser Glu Val Lys Ser Ser His Tyr Asp Ala Ile Ile
2732 2762
GTT GCA GTA GGA CAT CAG CAA TTT AAA CAG ATG GGA AGT GAG GAT ATT CGC GGC TTC GGA
Val Ala Val Gly His Gln Gln Phe Lys Gln Met Gly Ser Glu Asp Ile Arg Gly Phe Gly
2792 2822
AAA GAT AAA CAT GTA CTT TAT GAT TTG AAG TAT GTT CTT CCG GCT GAG CAG TCA GAT GTG
Lys Asp Lys His Val Leu Tyr Asp Leu Lys Tyr Val Leu Pro Ala Glu Gln Ser Asp Val
tvic 2851 2881
AGA TTG TAA TC ATG ACG GCT TAC GAA GAA CTA CGG ACC AAA CTG GTT CTG GCA CCA AAG
Arg Leu Met Thr Ala Tyr Glu Glu Leu Arg Thr Lys Leu Val Leu Ala Pro Lys
2911 2941
CGC TGG CTG ATC ACT GGC GTA GCA GGC TTT ATT GGC TCC GGC TTA TTA GAA GAA TTA CTC
Arg Trp Leu Ile Thr Gly Val Ala Gly Phe Ile Gly Ser Gly Leu Leu Glu Glu Leu Leu

2971 3001
TTT CTC AAC [illegible] ACT GTC ATT GGA CTG GAT AAC TTT TCC ACC GGT TAT CAG [illegible]T AAT CTA
Phe Leu Asn Gln Thr Val Ile Gly Leu Asp Asn Phe Ser Thr Gly Tyr Gln His Asn Leu
3031 3061
GAC GAC GTT CGC ACG TCC GTC AGT GAG GAG CAA TGG TCG CGA TTT ATT TTT ATT CAG GGT
Asp Asp Val Arg Thr Ser Val Ser Glu Glu Gln Trp Ser Arg Phe Ile Phe Ile Gln Gly
3091 3121
GAC ATC AGG AAA TTT ACT GAC TGT CAG AAA GCG TGT AAG AAC GTT GAC TAT GTT CTC CAC
Asp Ile Arg Lys Phe Thr Asp Cys Gln Lys Ala Cys Lys Asn Val Asp Tyr Val Leu His
3151 3181
CAA GCC GCG CTA GG[illegible] AGC GTG CCA CGT TCC CTA AAG GAT CCC ATC GCG ACT AAT AGC GCC
Gln Ala Ala Leu Gl[illegible] Ser Val Pro Arg Ser Leu Lys Asp Pro Ile Ala Thr Asn Ser Ala
3211 3241
AAT ATT GAT GGT TTT TTA AAT ATG TTG ACG GCG GCG AGA GAT GCT CAT GTC TCT AGT TTC
Asn Ile Asp Gly Phe Leu Asn Met Leu Thr Ala Ala Arg Asp Ala His Val Ser Ser Phe
3271 3301
ACC TAC GCC GCA AGC AGT AGC ACC TAT GGA GAC CAT CCC GAT TTA CCT AAA ATT GAG GAA
Thr Tyr Ala Ala Ser Ser Ser Thr Tyr Gly Asp His Pro Asp Leu Pro Lys Ile Glu Glu
3331 3361
CGG ATC GGT CGA CCA CTC AGC CCG TAT GCG GTA ACA AAA TAC GTC AAT GAA TTG TAC GCT
Arg Ile Gly Arg Pro Leu Ser Pro Tyr Ala Val Thr Lys Tyr Val Asn Glu Leu Tyr Ala
3391 3421
GAT GTG TTT GCA CGT AGC TAT GAA TTT AAC GCT AAT GGC CTA CGC TAC TTT AAT GTC TTT
Asp Val Phe Ala Arg [illegible] Tyr Glu Phe Asn Ala Ile Gly Leu Arg Tyr Phe Asn Val Phe
3451 3481
GGT CGC CGC CAA A[illegible] AAC GGA GCG TAC TCG GCA GTT ATT CCT CGC TGG ATA CTA TCG
Gly Arg Arg Gln A[illegible] Asn Gly Ala Tyr Ser Ala Val Ile Pro Arg Trp Ile Leu Ser
3511 3541
CTT CTT AAA GAT GAA CCA ATT TAT ATC AAT GGC GAT GGC TCA ACA AGC AGG GAT TTT TGC
Leu Leu Lys Asp Glu Pro Ile Tyr Ile Asn Gly Asp Gly Ser Thr Ser Arg Asp Phe Cys
3571 3601
TAT ATA GAG AAT GTG ATT CAG GCC AAT CTA TTA TCA GCA ACA ACT AAT GAT TTA GCC TCT
Tyr Ile Glu Asn Val Ile Gln Ala Asn Leu Leu Ser Ala Thr Thr Asn Asp Leu Ala Ser
3631 3661
AAA AAT AAG GTC TAT AAT GTG GCA GTT GGA GAT AGA ACT TCG TTA AAT GAG CTT TAT TAT
Lys Asn Lys Val Tyr Asn Val Ala Val Gly Asp Arg Thr Ser Leu Asn Glu Leu Tyr Tyr
3691 3721
CTA ATT CGC GAT GGG CTT AAT TTA TGG CGG AAC GAA CAA AGT AGA GCT GAA CCA ATT TAT
Leu Ile Arg Asp Gly Leu Asn Leu Trp Arg Asn Glu Gln Ser Arg Ala Glu Pro Ile Tyr
3751 3781
AAA GAT TTT CGT GAC GGT GAC GTT AAG CAT AGC CAG GCA GAT ATT ACC AAA ATA AAA ACA
Lys Asp Phe Arg Asp Gly Asp Val Lys His Ser Gln Ala Asp Ile Thr Lys Ile Lys Thr
3811 3841
TTT CTT TCA TAT GAG CCT GAA TTT GAT ATC AAA GAA GGA CTT AAG CAG ACT CTA AAA TGG
Phe Leu Ser Tyr Glu Pro Glu Phe Asp Ile Lys Glu Gly Leu Lys Gln Thr Leu Lys Trp
3871 3901
TAT ATC GAT AAA CAT [illegible] ACT TTG TAT TCC TCG GTA TAA CTA CTC ACT TTC CTT TCA CGT
Tyr Ile Asp Lys His Ser Thr Leu Tyr Ser Ser Val
3931 3961
GGA TGA ATT TAA TGA AAT CGT CAG GGA TGT TTA CGC TTA CAG CCA TTG GCA GTT GCC GTA
3991 4021
TTG TGA GTC CGG TAA A[illegible]AG CTC AGC CTT ATT TTA ATT TTC AGG CAA ACT TTA AGA GAA
4051 4081
TAT ATG GTT TTA CGC [illegible]TGA GCA GCG AGG CAT TAC AGC AGA TTC GGT TTA TTT TGG GCT
4111 4141
TGA TAG ACA TTC CTC [illegible]AGG TCC GAC CAT TTA TTT TTA GAC CTA ATG TAA ACT ATT CAA
4171 4201
ACA CCG ACG TAC ATA GTC GTT CTG ATT TCT ATA TCA TTG AGA TTT CAA GTC AGA AGA AGA
4231 4261
TTA TGG CCT ATG GGT TCT GCT TAC AAA TAA ATT ATT TAA CTC GTC ATT TCT ATG AAT TTT
tviD 4290 4320
TT AGC CAA AC AGAG CGG GCG TGC ATG TAC TGG TCG TTG GCC ACG CAA GGA AAT CGA CAC
Met Tyr Trp Ser Leu Ala Thr Gln Gly Asn Arg His
4350 4380
AAA CTG CTG GCC TAT CTC AAA GAC GAT CCC TGT TTT GCC GGA ATG TCG GAA GAC GAT CGT
Lys Leu Leu Ala Tyr Leu Lys Asp Asp Pro Cys Phe Ala Gly Met Ser Glu Asp Asp Arg

4410 4440
GCC TTA TTA AGC AAT ATC AAT GTC GAG CAG ATG GAT GAG CAT GCT ATC GAA CAG GAT ATG
Ala Leu Leu Ser Asn Ile Asn Val Glu Gln Met Asp Glu His Ala Ile Glu Gln Asp Met
4470 4500
ATG GAA ATC GTT CAG CTT CTT GGT CGC GAT CGC GTT ATG TTT ATG ACA CAT GTT GAT GCC
Met Glu Ile Val Gln Leu Leu Gly Arg Asp Arg Val Met Phe Met Thr His Val Asp Ala
4530 4560
GTG ACT CGT GCT GGA ACC GTC ATT CTA TCC CGT AGT CGG TTG ATT AAA AAT GTC GAC ACC
Val Thr Arg Ala Gly Thr Val Ile Leu Ser Arg Ser Arg Leu Ile Lys Asn Val Asp Thr
4590 4620
ATC GCC GCC AGG ATG GAT ATT CCC TGC GTT AAC CCG ACA AAT TTG ATG GAA AAG TGG GGG
Ile Ala Ala Arg Met Asp Ile Pro Cys Val Asn Pro Thr Asn Leu Met Glu Lys Trp Gly
4650 4680
CAG AAA CGA GCC CTG GAA AAA AAT GGC GAC GAT CTT ACT CAT TAT ACC GAT ATG TTT GGT
Gln Lys Arg Ala Leu Glu Lys Asn Gly Asp Asp Leu Thr His Tyr Thr Asp Met Phe Gly
4710 4740
GAC GCG ATC GTT GCG GCT ATT TTT AAG GGA GTG ATC AAT AAT ACT AAT CAT CAT CTT GAT
Asp Ala Ile Val Ala Ala Ile Phe Lys Gly Val Ile Asn Asn Thr Asn His His Leu Asp
4770 4800
GAG GGG CGA CAA GAG AAA CAG GAC CAA ATC AGT GAG ATT ACC TTA TCG ATC ACT AAG CAG
Glu Gly Arg Gln Glu Lys Gln Asp Gln Ile Ser Glu Ile Thr Leu Ser Ile Thr Lys Gln
4830 4860
CTT GCA GAT GGC GAC ATT ATT GCT GCA TCA CAA CAA CTT TTT GCC GCA TTA AGA AAT CAG
Leu Ala Asp Gly Asp Ile Ile Ala Ala Ser Gln Gln Leu Phe Ala Ala Leu Arg Asn Gln
4890 4920
CAG CAA GAT CCC GTT CTA ATC CAA CTT CGG TCC GTA ATC TTC AGC CAT TTA GGT TAT TAT
Gln Gln Asp Pro Val Leu Ile Gln Leu Arg Ser Val Ile Phe Ser His Leu Gly Tyr Tyr
4950 4980
GAA CAG GCT TAT CAG GAT ATT AGT GAT GTT GAG AAA ATT ATC GGT ACG ACT GAC AGT ACA
Glu Gln Ala Tyr Gln Asp Ile Ser Asp Val Glu Lys Ile Ile Gly Thr Thr Asp Ser Thr
5010 5040
TTA CGT TGT CGG CTG AGG TCT CTA CAT GGA TTA GCG CGT TGG CGG GAA GCC TTA TCG ACG
Leu Arg Cys Arg Leu Arg Ser Leu His Gly Leu Ala Arg Trp Arg Glu Ala Leu Ser Thr
5070 5100
GCA GAG ATG ATG CTT TCC AAT GAA ATT GAA GAT GAA GAA GTC CTT ACC GTT GCC GCC GGC
Ala Glu Met Met Leu Ser Asn Glu Ile Glu Asp Glu Glu Val Leu Thr Val Ala Ala Gly
5130 5160
TCA GCC GAT GCT TTA CAG CTG TTT GAT AAG TCA TAT CAT TAT TGG AAA CGT GTA CTA TTA
Ser Ala Asp Ala Leu Gln Leu Phe Asp Lys Ser Tyr His Tyr Trp Lys Arg Val Leu Leu
5190 5220
TTG AAT CCT GAA ACT CAA AGC GGA TGG GTT AAT TTC CTG AGC AGC ACG CAA TAT TTC AAT
Leu Asn Pro Glu Thr Gln Ser Gly Trp Val Asn Phe Leu Ser Ser Thr Gln Tyr Phe Asn
5250 5280
GAT GGC AAC GCA TTC TCT GAA GCT TTC CAT GCC GGC ATT CAA TCG CAG CGC CTA AAT GAT
Asp Gly Asn Ala Phe Ser Glu Ala Phe His Ala Gly Ile Gln Ser Gln Arg Leu Asn Asp
5310 5340
ACG TTT ATG GAA ACG GCG TTA TCT TTG GCA ATC AAA TTC AGT GAT GAA TTG ATT TTC ATG
Thr Phe Met Glu Thr Ala Leu Ser Leu Ala Ile Lys Phe Ser Asp Glu Leu Ile Phe Met
5370 5400
CAT GCG CTC GAG CAG CTA CTC CGC CAT GAG TCA GAA TTT GCG CTG ACG GTA TTG TCG ACG
His Ala Leu Glu Gln Leu Leu Arg His Glu Ser Glu Phe Ala Leu Thr Val Leu Ser Thr
5430 5460
ATT CAT GAT ACC GGT CTC GTT ATC CGC ACA GCT TTC TGC ATC AAG AAT ATG AGC TAT CAT
Ile His Asp Thr Gly Leu Val Ile Arg Thr Ala Phe Cys Ile Lys Asn Met Ser Tyr His
5490 5520
CAA GCG CTT CGC ACC TCG TAT AAA GAT AAA ATC CAC GAC GTT TTT GAG GCA TGG AAC AAT
Gln Ala Leu Arg Thr Ser Tyr Lys Asp Lys Ile His Asp Val Phe Glu Ala Trp Asn Asn
5550 5580
ACC GCG CTG TCG CTA CAT TCG GTT GAT GAT TTT GTC TCA CTG AGT ACT TCG CTA GCC TAT
Thr Ala Leu Ser Leu His Ser Val Asp Asp Phe Val Ser Leu Ser Thr Ser Leu Ala Tyr
5610 5640
AGC TAC TCT GCA TTT ATG GTT TAT CCC CAT TCA CGT ATT TCT CGC TTT AAT AAT GAA GTT
Ser Tyr Ser Ala Phe Met Val Tyr Pro His Ser Arg Ile Ser Arg Phe Asn Asn Glu Val
5670 5700
AAA ATG GCA TGG CGC GAT AAA TTA AGA GAA ATG TAT GAG CGT GAG GAT TAT GAA AAT ATC
Lys Met Ala Trp Arg Asp Lys Leu Arg Glu Met Tyr Glu Arg Glu Asp Tyr Glu Asn Ile

```
                                        5730                                        5760
CTG GCA GGG GCG AAA ATA GTG TGG CCA CTT CTG AAG TTT GAT CCC GTT GGC ACC GTA TAT
Leu Ala Gly Ala Lys Ile Val Trp Pro Leu Leu Lys Phe Asp Pro Val Gly Thr Val Tyr
                                        5790                                        5820
TGT GCA AGA ACG CTG GTG AAT CTT GGT GCC TGG AAA GAC GCG TGC ACG TTG GCC CAC ATG
Cys Ala Arg Thr Leu Val Asn Leu Gly Ala Trp Lys Asp Ala Cys Thr Leu Ala His Met
                                        5850                                        5880
ACC TTG ATT CGT AAC TCG AAC ATT ACC AGC CTG CAG TCG ATT ATG TTA CGC AGC ATA CGT
Thr Leu Ile Arg Asn Ser Asn Ile Thr Ser Leu Gln Ser Ile Met Leu Arg Ser Ile Arg
                                        5910                                        5940
CAT ATT AAC AAC ATT CCG TTC CTC ATT GAT TTG ATT GCT AAC GTC ATG AGC ATT ACT CTA
His Ile Asn Asn Ile Pro Phe Leu Ile Asp Leu Ile Ala Asn Val Met Ser Ile Thr Leu
                                        5970                                        6000
TCA TTC CAG AAT GCC TCA ATG AAC AAG TTG TTT GAG AAA GAG TGT CGC AAT GTT GCA ACC
Ser Phe Gln Asn Ala Ser Met Asn Lys Leu Phe Glu Lys Glu Cys Arg Asn Val Ala Thr
                                        6030                                        6060
AGA GCC CTT AAA TAT GTA CGC CAG AAG AAA ACT GAA GGG CGT CTG GAT GAA GCA TTG TCT
Arg Ala Leu Lys Tyr Val Arg Gln Lys Lys Thr Glu Gly Arg Leu Asp Glu Ala Leu Ser
                                        6090                                        6120
GTA TTG ATT AGC CTG AAA CGA ATT GAG CCT GAT GTT TCT CGT CTG ATG CGT GAA TAT AAG
Val Leu Ile Ser Leu Lys Arg Ile Glu Pro Asp Val Ser Arg Leu Met Arg Glu Tyr Lys
                                        6150                                        6180
CAA ATT ATC AGA TTA TTT AAT GAG TCA CGG AAG GAT GGC GGT AGC ACT ATC ACG TCT TAT
Gln Ile Ile Arg Leu Phe Asn Glu Ser Arg Lys Asp Gly Gly Ser Thr Ile Thr Ser Tyr
                                        6210                                        6240
GAA CAT CTA GAC TAT GCG AAA AAA TTA CTC GTT TTT GAT AGC GAA AAT GCC TAT GCC TTG
Glu His Leu Asp Tyr Ala Lys Lys Leu Leu Val Phe Asp Ser Glu Asn Ala Tyr Ala Leu
                                        6270                                        6300
AAA TAT GCC GCA TTA AAT GCA ATG CAT TTA CGC GAC TAC ACG CAG GCT TTG CAG TAT TGG
Lys Tyr Ala Ala Leu Asn Ala Met His Leu Arg Asp Tyr Thr Gln Ala Leu Gln Tyr Trp
                                        6330                                        6360
CAG CGA CTG GAG AAA GTG AAT GGA CCA ACG GAG CCG GTG ACA AGG CAG ATC TCG ACC TGC
Gln Arg Leu Glu Lys Val Asn Gly Pro Thr Glu Pro Val Thr Arg Gln Ile Ser Thr Cys
                                        6390          tviE                          6419
ATA ACC GCA TTA CAA AAA AAT ACA TCA GGG AAG TCG TA ATG ATT ACG CAG GAA GAA AAG
Ile Thr Ala Leu Gln Lys Asn Thr Ser Gly Lys Ser    Met Ile Thr Gln Glu Glu Lys
                                        6449                                        6479
TTA GCT GCA CTA GGA AAA ACG TGT TTA ACA TTA AAA CAA GAG AAG AAG CTT GCG CAA GCT
Leu Ala Ala Leu Gly Lys Thr Cys Leu Thr Leu Lys Gln Glu Lys Lys Leu Ala Gln Ala
                                        6509                                        6539
GTT GCG TTA ATT GAC AGT GAA TTA CCG ACT GAG GCT TTA ACT TCA TTA GCG ATG CTA AAA
Val Ala Leu Ile Asp Ser Glu Leu Pro Thr Glu Ala Leu Thr Ser Leu Ala Met Leu Lys
                                        6569                                        6599
AAA GCA GAG TTT CTT CAT GAT GTC AAT GAA ACG GAG CGT GCA TAC GCG CTC TAC GAA ACG
Lys Ala Glu Phe Leu His Asp Val Asn Glu Thr Glu Arg Ala Tyr Ala Leu Tyr Glu Thr
                                        6629                                        6659
CTG ATT GCA CAA AAC AAT GAT GAA GCA CGT TAT GAG TAT GCA CGT CGT TTA TAT AAT ACG
Leu Ile Ala Gln Asn Asn Asp Glu Ala Arg Tyr Glu Tyr Ala Arg Arg Leu Tyr Asn Thr
                                        6689                                        6719
GGG CTA GCC AAA GAT GCT CAG CTA ATT CTT AAA AAG GTT AGC AAT GGT GTG CAG AAA AAA
Gly Leu Ala Lys Asp Ala Gln Leu Ile Leu Lys Lys Val Ser Asn Gly Val Gln Lys Lys
                                        6749                                        6779
TAT AAC AAT TAT TTA GGC AAA ATA AAT AAG ATC TGT GAT TTG CTT GAA CGC CTT GAA GGG
Tyr Asn Asn Tyr Leu Gly Lys Ile Asn Lys Ile Cys Asp Leu Leu Glu Arg Leu Glu Gly
                                        6809                                        6839
AAA GCG ATC CCT GTG GGG ACC AAC ACC TGT ATT ATT GCA ATG AAG CAT GCC ATC TTG TTC
Lys Ala Ile Pro Val Gly Thr Asn Thr Cys Ile Ile Ala Met Lys His Ala Ile Leu Phe
                                        6869                                        6899
TAT AGA AAT CGT CAA CCC AGG CAG CTT CCC GTC GGG TCT TTC GGT CGT CTT GCG CTC TGT
Tyr Arg Asn Arg Gln Pro Arg Gln Leu Pro Val Gly Ser Phe Gly Arg Leu Ala Leu Cys
                                        6929                                        6959
ACT GGC TCG CTA GGT AGC GGT GGT GCA GAG CGT CAG ATT TCC AGG CTG GCT ATC GAA ATC
Thr Gly Ser Leu Gly Ser Gly Gly Ala Glu Arg Gln Ile Ser Arg Leu Ala Ile Glu Ile
                                        6989                                        7019
GCC AGA AAA TAT CGG CAA AAG GGG AAA ATT GGC GGC CTG AAA GTA GAA GAA CCG GTA C A
Ala Arg Lys Tyr Arg Gln Lys Gly Lys Ile Gly Gly Leu Lys Val Glu Glu Pro Val
```

```
                                        7049                                        7079
CTA ATT ATT CGC TCC CTG ACA CCG GAA CTC AGG CAA GAC TTT TTC CTG AAA GAA GTG CTG
Leu Ile Ile Arg Ser Leu Thr Pro Glu Leu Arg Gln Asp Phe Phe Leu Lys Glu Val Leu
                                        7109                                        7139
GAA GAA CAG GTC GAG GTT CTT GAA ATC GCG AAG ATT ACC GGA AAC TTG TTT GAC GAT GCG
Glu Glu Gln Val Glu Val Leu Glu Ile Ala Lys Ile Thr Gly Asn Leu Phe Asp Asp Ala
                                        7169                                        7199
ACA ATA GAA TCT CCA GAG TTG CGC TTA TTG CTA TCG CAT CTA CCG CCG GTG TGT AAA TAC
Thr Ile Glu Ser Pro Glu Leu Arg Leu Leu Leu Ser His Leu Pro Pro Val Cys Lys Tyr
                                        7229                                        7259
GGC ATC AAG CAT CTG GTC CCC CAT TTA TGC GAG CGC AAG CTG GAT TAT CTC TCC GTT TGG
Gly Ile Lys His Leu Val Pro His Leu Cys Glu Arg Lys Leu Asp Tyr Leu Ser Val Trp
                                        7289                                        7319
CAG GAT GGC GCT TGT CTG ATG ATT GCG CTT GCA GCA TTG ATT GCT GGC GTG CCC AGA ATT
Gln Asp Gly Ala Cys Leu Met Ile Ala Leu Ala Ala Leu Ile Ala Gly Val Pro Arg Ile
                                        7349                                        7379
CAA CTG GGA TTA CGT GGG TTA CCG CCG GTG GTT AGA AAG CGT CTG TTC AAG CCG GAA TAT
Gln Leu Gly Leu Arg Gly Leu Pro Pro Val Val Arg Lys Arg Leu Phe Lys Pro Glu Tyr
                                        7409                                        7439
GAG CCT CTC TAC CAG GCG CTG GCG GTC GTG CCT GGC GTT GAT TTT ATG AGT AAC AAC CAT
Glu Pro Leu Tyr Gln Ala Leu Ala Val Val Pro Gly Val Asp Phe Met Ser Asn Asn His
                                        7479                                        7499
TGT GTG ACT CGC CAT TAT GCC GAC TGG CTG AAG TTG GAG GCG AAC GAC TTC CAG GTT GTA
Cys Val Thr Arg His Tyr Ala Asp Trp Leu Lys Leu Glu Ala Asn Asp Phe Gln Val Val
                                        7529                                        7559
TAT AAC GGC GTC TTA CCG CCA TCT ACT GAA CCC TCT TCT GAG GTG CCA CAT AAA ATC TGG
Tyr Asn Gly Val Leu Pro Pro Ser Thr Glu Pro Ser Ser Glu Val Pro His Lys Ile Trp
                                        7589                                        7619
CAG CAG TTT ACG CAA AAA ACC CAG GAT GCG GAC ACG ACT ATT GGT GGC GTT TTC CGC TTT
Gln Gln Phe Thr Gln Lys Thr Gln Asp Ala Asp Thr Thr Ile Gly Gly Val Phe Arg Phe
                                        7649                                        7679
GTA GGC GAT AAG AAC CCT TTT GCA TGG ATT GAT TTT GCA GCA CGC TAT TTA CAA CAC CAC
Val Gly Asp Lys Asn Pro Phe Ala Trp Ile Asp Phe Ala Ala Arg Tyr Leu Gln His His
                                        7709                                        7739
CCC GCC ACG CGC TTT GTG CTG GTA GGC GAT GGT GAT TTA CGC GCT GAA GCG CAG AAA CGC
Pro Ala Thr Arg Phe Val Leu Val Gly Asp Gly Asp Leu Arg Ala Glu Ala Gln Lys Arg
                                        7769                                        7799
GCC GAA CAG TTA GGG ATT CTG GAG AGA ATA CTA TTC GTT GGC GCC TCG CGT GAC GTA GGG
Ala Glu Gln Leu Gly Ile Leu Glu Arg Ile Leu Phe Val Gly Ala Ser Arg Asp Val Gly
                                        7829                                        7859
TAT TGG CTG CAA AAA ATG AAT GTA TTC ATT TTG TTT TCG CGT TAT GAA GGG CTA CCT AAT
Tyr Trp Leu Gln Lys Met Asn Val Phe Ile Leu Phe Ser Arg Tyr Glu Gly Leu Pro Asn
                                        7889                                        7919
GTG CTT ATT GAA GCA CAA ATG GTC GGG GTG CCG GTG ATT TCA ACC CCT GCA GGT GGA TCG
Val Leu Ile Glu Ala Gly Met Val Gly Val Pro Val Ile Ser Thr Pro Ala Gly Gly Ser
                                        7949                                        7979
GCA GAA TGC TTT ATT GAG GGT GTT TCG GGT TTC ATT CTT GAT GAT GCA CAG ACG GTC AAT
Ala Glu Cys Phe Ile Glu Gly Val Ser Gly Phe Ile Leu Asp Asp Ala Gln Thr Val Asn
                                        8009                                        8039
CTT GAC CAG GCT TGC CGC TAT GCA GAA AAG TTG GTC AAT TTA TGG CGC AGC AGA ACC GGT
Leu Asp Gln Ala Cys Arg Tyr Ala Glu Lys Leu Val Asn Leu Trp Arg Ser Arg Thr Gly
                                        8069                                        8099
ATT TGC CAA CAG ACG CAG TCA TTT TTA CAA GAA CGC TTC ACC GTG GAA CAT ATG GTG GGA
Ile Cys Gln Gln Thr Gln Ser Phe Leu Gln Glu Arg Phe Thr Val Glu His Met Val Gly
                                        8129                                        8160
ACG TTT GTA AAA ACC ATT GCC TCT CAG CCT CGT TAA TTA ATG GGC ATC ATT TTT CAG CTAT
Thr Phe Val Lys Thr Ile Ala Ser Gln Pro Arg
                        tviF        8191                                        8221
TTCA TTT ATA AAA TAA GTT ATG AAA AAA ATC ATC ATA TTA CTA ACG ACA TTT TTC CTG CTT
                            Met Lys Lys Ile Ile Ile Leu Leu Thr Thr Phe Phe Leu Leu
      *                                 8251                                        8281
TCG GGA TGC ACT ATT CCC AGG GCG GTA TTT AAA TCC AGC CTT ATT AAT CAG GAC GAT CCT
Ser Gly Cys Thr Ile Pro Arg Ala Val Phe Lys Ser Ser Leu Ile Asn Gln Asp Asp Pro
                                        8311                                        8341
CGT TAT AAT CTG GTC GAA GTC ACG CCG ACA TTA AAA CTA AGC GCT CCC GAT ACT GTG CCG
Arg Tyr Asn Leu Val Glu Val Thr Pro Thr Leu Lys Leu Ser Ala Pro Asp Thr Val Pro
```

8371 8401
AAA ACT ATT GTC GAT CCG GTT TTT GCC GCA AAT AAC TGG CAC TGG ACA TCT TTG GCT AAA
Lys Thr Ile Val Asp Pro Val Phe Ala Ala Asn Asn Trp His Trp Thr Ser Leu Ala Lys
8431 8461
GGC GAT GTG CTG CAT ATC ACT ATT TTA TCC TCG GGC GGG GCT GGA TAT TTA TCC AAT AAC
Gly Asp Val Leu His Ile Thr Ile Leu Ser Ser Gly Gly Ala Gly Tyr Leu Ser Asn Asn
8491 8521
GCG AGC GGC GAC CGT GCG GAT TTT GAA AAT ATT CTT GTG ACT GAC AGT AAT ACC GTT CAG
Ala Ser Gly Asp Arg Ala Asp Phe Glu Asn Ile Leu Val Thr Asp Ser Asn Thr Val Gln
8551 8581
GTG CCT TAT GCC GGG ACA ATC CCG GTT TCT GGA TTG GAT GTG ACG CAA CTG GCT GAT GAG
Val Pro Tyr Ala Gly Thr Ile Pro Val Ser Gly Leu Asp Val Thr Gln Leu Ala Asp Glu
8611 8641
ATC AAA AAG CGA CTT TCG CGC GTT GTC CTG AAT CCT CAG GTG ATT GTG ACA CTT ACC GCC
Ile Lys Lys Arg Leu Ser Arg Val Val Leu Asn Pro Gln Val Ile Val Thr Leu Thr Ala
8671 8701
CGC ACC GGA GCC ATG GTG ACG GTC GAG GGC AGC GGG AAA ACG GGT CGA TAC CCT CTC GAA
Arg Thr Gly Ala Met Val Thr Val Glu Gly Ser Gly Lys Thr Gly Arg Tyr Pro Leu Glu
8731 8761
CAG AGT ATG AAT CGT CTG AGT CAT CTT TTG GCA ACG GCA GTG GCA GTA GAA AAT ACC AGC
Gln Ser Met Asn Arg Leu Ser His Leu Leu Ala Thr Ala Val Ala Val Glu Asn Thr Ser
8791 8821
ACA GAT ATG ATG GAA GTT CAC GTG ACG CGC CAG CAG CAT TAT TTC ACT GCG CGA CTG TCA
Thr Asp Met Met Glu Val His Val Thr Arg Gln Gln His Tyr Phe Thr Ala Arg Leu Ser
8851 8881
GAT ATT TAT CAG TAT CCC GGA TTG GAT ATT GCC TTA CAG CCG GAT GAC CGT ATC ACT CTG
Asp Ile Tyr Gln Tyr Pro Gly Leu Asp Ile Ala Leu Gln Pro Asp Asp Arg Ile Thr Leu
8911 8941
CGT CAA GTG ACC GAG TAT GTT AAT GTG CTC GGC GCA GCA GGT GTT CAG GGG AAA CAT GCT
Arg Gln Val Thr Glu Tyr Val Asn Val Leu Gly Ala Ala Gly Val Gln Gly Lys His Ala
8971 9001
CTG GTT CAG CGT CAT TCC AGC GTC GTG GAT GCT TTG GCG CTC GCG AAA GGA TTA AAT GAC
Leu Val Gln Arg His Ser Ser Val Val Asp Ala Leu Ala Leu Ala Lys Gly Leu Asn Asp
9031 9061
AAT CTT GCC GAT CCG CAA GCG ATT TTT CTG TAC AAG CAT AAC GAA GCG GAA CAG GCG AAA
Asn Leu Ala Asp Pro Gln Ala Ile Phe Leu Tyr Lys His Asn Glu Ala Glu Gln Ala Lys
9091 9121
CAG CAG ATG CGT AAG CTG AAT ATC TAT CAT GTC GAT ATG AGC CAA CCT AAC TCG GTG TTT
Gln Gln Met Arg Lys Leu Asn Ile Tyr His Val Asp Met Ser Gln Pro Asn Ser Val Phe
9151 9181
TTA GCG CAG GCC ATA CGA GTG GAC AAC GGC GAT GTT ATC TAT ATC TCG AAC GCC TCT TTG
Leu Ala Gln Ala Ile Arg Val Asp Asn Gly Asp Val Ile Tyr Ile Ser Asn Ala Ser Leu
9211 9241
ACT GAT TTC GCT AAG GTG AAA GCC GCA TTC GAT AGC TTT TTG ACC CGC GGC ACT AAT TCT
Thr Asp Phe Ala Lys Val Lys Ala Ala Phe Asp Ser Phe Leu Thr Arg Gly Thr Asn Ser
9271 9301
TTC TAA TGT CGC CAT ATG AAT ATT TTG AAA AAT AAT TCT TAT TAT TTT ATG AAA CTC ATT
Phe