High degree of conservation between ribosomal proteins of *Yersinia pseudotuberculosis* and *Escherichia coli*

Uwe Gross, Jin-Hai Chen, Dwight H.Kono¹, Juan G.Lobo and David T.Y.Yu*

University of California, Department of Medicine, Rehabilitation Center, 32-47, 1000 Veteran Avenue, Los Angeles, CA 90024 and ¹California Institute of Technology, Division of Biology, Pasadena, CA 91125, USA
Submitted February 28, 1989 EMBL accession no. X14363

The S10 operon of *Escherichia coli* consists of a gene cluster encoding eleven ribosomal proteins (1). In an attempt to clone a HLA-B27 cross-reactive outer membrane protein of *Yersinia pseudotuberculosis*, we have fortuitously isolated a 2.4 Kb DNA fragment. This contained five open reading frames which were homologous to the *E. coli* genes encoding the ribosomal proteins L2, L23, L4 and parts of S19 as well as L3 (Figure 1).

The degree of similarities of the genes between these two bacteria ranged from 85.9% to 93.2%. They also shared a codon bias which was distinctly different from those encoding the outer membrane proteins. An unusual feature of the *E. coli* S10 operon is that there is an overlap between the stop codon for the gene encoding L4 with the start codon for the gene encoding L23. Even this feature was conserved in *Y. pseudotuberculosis*.

The amino acid sequences of these *Y. pseudotuberculosis* proteins did not have any homology to HLA-B27 but were more than 90% homologous to the equivalent proteins in *E. coli*.

Considerable similarities in ribosomal proteins have been reported among bacteria, plants and man (2,3,4). Since the DNA of *E. coli* and *Y. pseudotuberculosis* are only 10-20% related, our findings verified the high degree of conservation of these genes in evolution.

SpaH1	
rplC	
GAA TTC CGC CTT CCA GAA GGT CAA GAG TTT ACT GCT GGT CAA GAA ATT AGC GTC GAA ATT TTT GCA GAC GTT AAG AAA GTC GAC GTT ACA	90
Glu Phe Arg Leu Pro Glu Gly Gln Glu Phe Thr Ala Gly Gln Glu Ile Ser Val Glu Ile Phe Ala Asp Val Lys Lys Val Asp Val Thr	
L3	
GGT ACG TCT AAA GGT AAA GGT TTT GCC GGT ACT GTT AAG CGC TGG AAT TTC CGC ACC CAA GAT GCT ACC CAT GGT AAC TCC TTG TCT CAC	180
Gly Thr Ser Lys Gly Lys Gly Thr Val Lys Arg Trp Asn Phe Arg Thr Gln Asp Thr His Gly Asn Ser Leu Ser His	
CGT GTT CCG GGT TCT ATC GGT CAA AAC CAG ACT CGG GGC AAA CTG TTT AAA GGC AAG AAA ATG GCA GGC CAT ATG GGT GAT GAA CGT GTA	270
Arg Val Pro Gly Ser Ile Gly Gln Asn Gln Thr Pro Gly Lys Val Phe Lys Gly Lys Lys Met Ala Gly His Met Gly Asp Glu Arg Val	
ACT GTT CAG AGC CTG GAT GTA GTA CGT GTT GAC GCT GAG CGC AAC CTA CTG CTG GTC AAG GGC GCT GTT CCG GGA GCT ACC GGT GGC AAC	360
Thr Val Gln Ser Leu Asp Val Val Arg Val Asp Ala Glu Arg Asn Leu Leu Leu Val Lys Gly Ala Val Pro Gly Ala Thr Gly Gly Asn	
rplD	
CTG ATC GTA AAA CCG GCT GTC AAG GCG TAACGTCGAGGATAGGA ATG GAA TTG GTA ATG AAA GAC GCG CCA GGC GCG CTG ACT GTT TCC GAA	454
Leu Ile Val Lys Pro Ala Val Lys Ala *** Met Glu Leu Val Met Lys Asp Ala Pro Gly Ala Leu Thr Val Ser Glu	
L4	
ACT ACC TTC GGG CGT GAT TAT AAT GAA GCG CTT GTA CAT CAG GGT GTT GTT GCT TAC GCA GCA GGT GCC CGT CAA GGT ACT CGT GCT CAG	544
Thr Thr Phe Gly Arg Asp Phe Asn Glu Ala Leu Val His Gln Val Val Val Ala Tyr Ala Ala Gly Ala Arg Gln Gln Gly Thr Arg Ala Gln	
AAG ACC CGT GCT GAA GTA ACA GGT TCC GGT AAA AAA CGG TGG CPT CAG AAA GGT ACC GGC CPT GCG CPT GCA GGT TCT GTA AAG AGC CCG	634
Lys Thr Arg Ala Glu Val Thr Gly Ser Gly Lys Lys Pro Trp Arg Gln Lys Gly Thr Gly Arg Ala Arg Ala Gly Ser Val Lys Ser Pro	
ATC TGG CGT TCA GGT GGT GTC ACC TTT GCT GCA AAG CCT CAG GAC CAG AGT CAA AAA GTA AAT AAA AAG ATG TAC CCG GCG GCG CTG AAA	724
Ile Trp Arg Met Lys Thr Phe Ala Lys Thr Gly Ser Glu Lys Lys Pro Trp Arg His Ser Gln Lys Lys Val Asn Phe Lys Met Tyr Lys Leu Lys	
AGC ATT CTG TCC GAA TTG GTA CTA GAT GAT GGT CTG ATC ATT GTC GAA AAG TTC TCT GGT GAA GCA CCT AAA ACT AAG TTG CTG GCG CAG	814
Ser Ile Leu Ser Glu Leu Val Arg Gln Asp Arg Leu Ile Ile Val Glu Lys Phe Ser Val Glu Ala Pro Lys Thr Lys Leu Leu Ala Gln	
AAA CTG AAA GAT ATG GCT GCT GAA GAT GTG CTG ATC GTA ACG GGT GAA CTG GAC GAG AAC TTG TTC TTG GCG GCT CCG AAC CTG TAC AAG	904
Lys Leu Lys Asp Met Ala Leu Glu Asp Val Phe Ala Ala Leu Ile Val Thr Gly Glu Leu Asn Glu Asn Lys Leu Thr Tyr Lys	
GTT GAT GTT CCG GAT GTA GCC GGT ATT GAT CCA GTT AGC CTG ATC GGC TTC GAC AAA GTG GTT ATG ACT GCT GAT GCT GTG AAG CAA GTT	994
Val Asp Val Arg Asp Val Ala Gly Ile Asp Pro Val Ser Leu Ile Ala Phe Asn Lys Val Val Met Thr Ala Asp Ala Val Lys Gln Val	
rplW	
GAG GAG ATG CTG GCA TG ACT GGT GAA GAA CPT CTG CTG AAA GTA CTG CCG CCG CAT GTA TCT GAA AAA GCG TCC GCT GCG ATG GAA	1083
Glu Glu Met Leu Ala Met Ile Arg Glu Glu Arg Leu Leu Lys Lys Val Leu Arg Ser Pro His Val Ser Glu Lys Ala Ser Ala Ala Met Glu	
L23	

AAG AAT AAC ACC ATC GTT CTC AAA GTT GGC AAA GAC GCG ACC AAA GCA GAA ATT AAA	Tr1	GCA GTC CAG AAA CTG TTT GAA GTC GAA GTC	1173
Lys Asn Asn Thr Ile Val Leu Lys Val Ala Lys Asp Ala Thr Lys Ala Glu Ile Lys Ala		Met Glu Val Glu Val	
GAA GAC GTT AAC ACC TTG CTG GTT AAA GGC AAG AGT AAG COT CAC GGT CAG COT GTT GGT		COT COT AGC GAC TCG AAA AAA GCT TAC GTC	1263
Glu Asp Val Asn Thr Leu Leu Val Lys Gly Lys Ser Lys Arg His Gly Gln Arg Val Gly Arg		Arg Ser Asp Trp Lys Lys Ala Tyr Val	
ACC CTG AAA GAA GGC CAG AAT CTG GAC TTC ATC GGC GGC GCA GAG	rps8	TRAGTCTGGGAGGATTAACA ATG CCA ATT GTT AAA TGT AAA OCT ACG TCT	1357
Thr Leu Lys Glu Gly Gln Asn Leu Asp Phe Ile Gly Gly Ala Glu ***		Met Ala Ile Val Lys Cys Lys	
CGG GGT COT CAC GAC GTT GTT AAA CTG GTT AAG COT GAG CTG GAC CAG CAG CAG	L2	ANG GGT AAG COT TAT TAT GGC CCG TTG CTT GAG AAA TTA AGC AAA AGC	1447
Pro Gly Arg Arg His Val Val Lys Val Val Asn Pro Gln Leu His Lys Gly Lys Pro Tyr Ala		Pro Leu Leu Glu Lys Leu Ser Lys Ser	
GCT GGC COT AAC AAC AAT GGC COT ATC ACT ACC COT CAT ATC GGT GGT GGC CAC AAG		GAA CAT TAT COT CTG GTT GAC TTC AAA CCG AAC	1537
Gly Gly Arg Asn Asn Asn Gly Arg Ile Thr Thr Arg His Ile Gly Gly Gly His Lys Gln		His Tyr Arg Leu Val Asp Phe Lys Arg Asn	
AAA GAT GGT ATC CCT GCT GTG GTT GAG COT CAG TAC GAT CCG AAC COT TCC GCA AAT		ATT CCG TTG GTT CTG TAC AAA GAC GGC GAA	1627
Lys Asp Arg Gly Ile Pro Ala Val Val Lys Thr Asp Pro Asn Arg Ser Ala Asn Ile Ala		Leu Val Leu Tyr Lys Asp Gly Glu	
CGC COT TAT ATC CTG GCG CCG AAA GGC CTG AAG COT GGT GAC CAG ATT CAA TCT GGC		GGT GAT GCT GCA ATT AAA GCA GGT AAC ACC CTG	1717
Arg Arg Tyr Ile Leu Ala Pro Lys Gly Leu Lys Ala Gly Asp Gln Ile Gln Ser Gly Val		Asp Ala Ala Ile Lys Ala Gly Asn Thr Leu	
CCT ATG COT AAC ATC CCA GTT GGT TCA ACG GTT CAT AAC GTA GAA ATG AAA CCA GGT		AAA GGC GGC CAG TTG GCT COT TCT GCT GGT GCA	1807
Thr Leu Gly Ile Pro Val Gly Ser Val His Asn Val Glu Met Lys Pro Gly Lys Ala Gly		Ala Ser Arg Trp Arg Gly Ile Arg Pro Thr Val	
TAC GTT CAG ATC GTT GCT COT GAC GGT TCC TAC TGT ACT CTG COT CTG CCG TCC GGC		GAA ATG CCG AAC GGT CAA GCT GAT TGC CCG GCC	1897
Tyr Val Gln Ile Val Ala Asp Asp Gly Ser Tyr Val Thr Leu Arg Leu Arg Ser Gly Glu		Met Arg Lys Val Gln Ala Asp Cys Arg Ala	
ACC TTA GGT GAA GTT GGT AAT GCT GAA CAC ATG CTG COT GTC CTG GGT AAA GCA GGT		GCT AGT COT TGG COT GGT ATT COT COT ACC GTT	1987
Thr Leu Gly Glu Val Gly Asn Val Glu His Met Leu Arg Val Leu Glu Lys Ala Gly Ala		Ser Arg Trp Arg Gly Ile Arg Pro Thr Val	
CGC GGT ACG GCG ATG AAC CCA GTA GAT CAC CCG CAC GGT GGT GAA GGT COT AAC TTT		GGT GAT AAC CAG CCG GTA ACC CCG TGG GGC GTT	2077
Arg Gly Thr Ala Met Asn Pro Val Asp His Pro His Gly Gly Gly Glu Gly Arg Asn		Phe Gly Lys His Pro Val Thr Trp Gly Val	
CAG ACC AAA GGT AAG AAG ACC COT AGC AAC AAG COT ACT GAT AAG TTC ATC GTA COT	rps5	CGC COT AGT AAA AAATAATAAGGATTAAGCCATG CCA	2172
Gln Thr Lys Gly Lys Lys Thr Arg Ser Asn Lys Arg Thr Asp Lys Phe Ile Val Arg Arg		Ser Lys Lys*** Met Pro S19	
COT TCT CTC AAG AAA GGT CCC TTC ATT GAC CTG CAC TTG CTG AAG AAG GTA GAG AAA		GCG GTG GAA AGC GGA GAC AAG AAG CCT ATT CGG	2262
Arg Ser Leu Lys Lys Gly Pro Phe Ile Asp Leu His Leu Leu Lys Lys Val Glu Lys Ala		Val Glu Ser Gly Asp Lys Lys Pro Ile Arg	
ACT TGG TCC COT COT TCA ACG GTC TTT CCA AAT ATG ATC GGT TTG ACC ATC GCT GTC		CAT AAT GGT COT CAG CAC GTT CCA GTA TTC GTT	2352
Thr Trp Ser Arg Arg Ser Thr Val Phe Pro Asn Met Ile Gly Leu Thr Ile Ala Val		His Asn Gly Arg Gln His Val Pro Val Phe Val	
TCC GAT GAA ATG GTC GGT CAC AAA CTG GGT GAA TTC	SpoRI		2388
Ser Asp Glu Met Val Gly His Lys Leu Gly Glu Phe			

Fig. 1: Nucleotide sequence of the cloned *EcoRI*-fragment of *Y. pseudotuberculosis* and predicted amino acid sequences of the gene products. Designations follow the ones given for corresponding genes and proteins of the *S10* operon of *E. coli*. L3, L4, L23, L2, and S19 correspond to ORF 1-5, respectively. Numbering of nucleotides starts from the 5'-*EcoRI* restriction site. Regions similar to proposed ribosomal binding sites are underlined. This sequence has been assigned the accession number x14363 in the EMBL-databank.

ACKNOWLEDGMENTS This project was supported by the Nora Eccles Treadwell Foundation and USPHS AR36834. U. Gross was a recipient of a scholarship from the Deutsche Forschungsgemeinschaft.

*To whom correspondence should be addressed

REFERENCES

1. Jaskunas, S., R., Fallon, A. M. and Nomura, M. (1977) *J. Biol. Chem.*, 252, 7323-7336.
2. Chen, I.-T., Dixit, A., Rhoads, R. R. and Roufa, D. J. (1986) *Proc. Natl. Acad. Sci. USA*, 83, 6907-6911.
3. Tanaka, M., Wakasugi, T., Sugita, M., Shinozaki, K. and Sugiura, M. (1986) *Proc. Natl. Acad. Sci. USA*, 83, 6030-6034.
4. Yagushi, M., Matheson, A. T., and Visentin, L. P. (1974) *FEBS Lett.*, 46, 296-300.